CITY, UNIVERSITY OF LONDON

DOCTORAL THESIS

Essays on Factor Models in Asset Pricing

Author:
Alberto Ciampini

Supervisor:
Prof. Giovanni Urga
Prof. Nikos Nomikos

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

City, University of London
Bayes Business School
Faculty of Finance

July 2023

To my parents, Nazzareno and Silvana

Contents

List of Tables

List of Figures

List of Abbreviations

APT	**A**sset **P**ricing **T**heory
AR	**A**uto-**R**egressive
AMP13	Asness, Moskowitz, and Pedersen (2013)
BAB	**B**etting **A**gainst **B**eta
CAPM	**C**apital **A**sset **P**ricing **M**odel
CMA	**C**onservative **M**inus **A**ggressive
COND	**C**onditional
ESDC	**E**uropean **S**overeign **D**ebt **C**risis
EU	**E**uropean **U**nion
FF	**F**ama-**F**rench
FF3	Fama and French (2012)
FF5	Fama and French (2017)
FP14	Frazzini and Pedersen (2014)
GFC	**G**reat **F**inancial **C**risis
GDP	**G**ross **D**omestic **P**roduct
HML	**H**igh **M**inus **L**ow
IMF	**I**nternational **M**onetary **F**und
IS	**I**n-**S**ample
KF	**K**alman **F**ilter
LN	Lewellen and Nagel (2006)
MAE	**M**ean **A**bsolute **E**rror
MEA	**M**iddle **E**ast & **A**frica
MKT	**M**arket
ML(E)	**M**aximum **L**ikelihood (**E**stimation)
MSE	**M**ean **S**quare **E**rror
MSFE	**M**ean **S**quare **F**orecasting **E**rror
OLS	**O**rdinary **L**east **S**quares
OOS	**O**ut-**O**f-**S**ample
PC(A)	**P**rincipal **C**omponent (**A**nalysis)
RMW	**R**obust **M**inus **W**eak
PRED	**P**redictive
RW	**R**olling **W**indow
SMB	**S**mall **M**inus **B**ig
TTM	**T**ime-**T**o-**M**aturity
UK	the **U**nited **K**ingdom
US	the **U**nited **S**tates of America
VAR	**V**ector **A**uto-**R**egressive

Acknowledgements

I am deeply grateful to my supervisors Prof. Giovanni Urga and Prof. Nikos Nomikos for their constant support and advise throughout my PhD.
I highly appreciate the support by the whole Finance faculty during the last peculiar years for our society, always delivering at the highest standards of research and teaching. A particular mention goes to Prof. Giacinta Cestone, Malla Pratt, and Abdul Momin, a sincere thank you.
I also want to thank Lorenzo Testa for his helpful comments and research ideas.
My parents, Nazzareno and Silvana, and my siblings, Alessandro and Sara, deserve my deepest appreciation for their constant support and affection. Thank you for everything.
A scholarship by City, University of London provided the funds for this research.

Declaration

I, Alberto Ciampini, declare that this thesis titled, "Essays on Factor Models in Asset Pricing" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Alberto Ciampini

Date: July 25, 2023

Abstract

This thesis comprises three essays on multi-factor asset pricing models in finance.

In the first essay, I expand on the work of Borghi et al. (2018) by comparing the in-sample performance of their proposed ML estimator of time-varying sensitivities, which draws from Mikkelsen, Hillebrand, and Urga (2019), against the alternative rolling least square estimator. The main finding of my analysis is that the rolling OLS estimator is characterised by a great degree of instability, which is driven by the interplay between window size and sampling frequency. If one selects the window size arbitrarily the instability in the estimates can be pronounced, and the benefits of fitting a dynamic model as opposed to an equivalent static-loadings representation become slim. Overall, the ML estimator of Borghi et al. (2018) dominates the rolling OLS under many aspects.

In the second essay, I expand on the work presented in the first essay in two ways. I compare the explanatory power of the three-factor model of Borghi et al. (2018), that features a combination of observed and latent factors, against more traditional factors constructed from firm attributes, and I evaluate the out-of-sample predictive performance of rolling OLS betas in forecasting future return patterns, accounting explicitly for the window selection problem. The main finding of my analysis points to a dual role of the rolling least square estimator when I employ the measures in Kelly, Palhares, and Pruitt, 2021 to gauge the models performance. While the short-window approach provides the best results in a contemporaneous-equation setting, for predictive purposes including too little observations for estimation causes the betas to be noisy, which in turn results in forecasts with little predictive power. Out-of-sample, the trade-off between the length of the window and the variance of the estimator is resolved around the two-year point, and this is true across all model specifications. I find that the choice of the window length alone accounts for about $\pm 10\%$ of the factor model's out-of-sample forecasting performance. Analysing the out-of-sample relative performance of alternative model specifications that include observed factors, I find the performance of Fama French factors deteriorates significantly with respect to the in-sample analysis.

The third essay turns to the analysis of factor premia in international sovereign bonds. I identify the factors with country-style characteristic-based portfolios such as momentum, value, and low-risk, and study their performance under two dimensions simultaneously, issuer- and maturity-wise. My analysis reveals substantial variation in the factor premia across the cross-sections, which does not support the view of Asness, Moskowitz, and Pedersen (2013) and Frazzini and Pedersen (2014) on their unifying pricing ability across countries and asset classes. I find that risk-adjusted returns are decreasing in the maturity of the bonds for momentum strategies. When analysed across countries, momentum produces consistent statistically significant Sharpe ratios, however this is not true for value and low-risk. The former shows low and insignificant returns across countries, while low-risk yields statistically significant premia only for Euro Area bonds. Contrarily to what reported in previous literature, I find no supporting evidence for the existence of risk premia for characteristics-based global portfolios.

General Introduction

This introduction gives a broad overview of the thesis, which comprises three essays on multi-factor asset pricing models in finance. The order of the essays is chronological and thus reflects my development during the PhD program.

The research objective of this thesis is to investigate the performance of factor models in modelling international asset returns, in the equity and sovereign bond markets separately. The research uses a large set of asset-level data to answer questions at the intersection between asset pricing and applied econometrics. The investigation is mainly empirical, however the theoretical aspects of the models are also carefully considered due to the complexity of the econometric techniques employed.

The first research question is on the choice of the relevant pricing factors that drive the cross-sectional variation in asset returns. In the context of international equities, I compare observed[1] factors, that are identified a-priori, against latent[2] factors, that are not directly observed and thus need to be estimated. For sovereign bonds, I identify the factors with characteristic-based portfolios such as momentum, value, and low-risk, which have been shown in the literature to capture patterns of cross-sectional variation in the returns, and compare their performance across various dimensions. The second research question is on the estimation of the time-varying factor sensitivities[3] in international equities. Conditional on the choice of the factors, I compare two competing methodologies to introduce the time variation in the factor loadings, rolling least squares and ML estimation (ie. the Kalman filter). The former is the status quo in the literature and depends crucially on the choice of the window length, which is often judgemental and based on past experience. I provide an in-depth analysis on the statistical and economic properties of the rolling least square estimator for varying window length.

The first two essays, Chapter 1 and 2, study multi-factor asset pricing models in international equity returns. Both chapters contribute to the existing literature by providing an alternative perspective to analyse higher-frequency patterns in international stock returns, using asset-level data and country-style factors. I depart from the monthly benchmark commonly used in the literature, and work with weekly data on a sample spanning several years of recent history from 2006 to 2019. A further characteristics common to the chapters is that they use

[1] In this document I use the terms 'observed', or 'observable' factors to indicate as synonyms.
[2] In this document I use the terms 'statistical', 'latent', 'unobserved' or 'unobservable' factors as synonyms.
[3] In this document I use the terms 'sensitivity', 'exposure', 'loading', or 'beta' as synonyms.

individual stock returns as base assets, instead of portfolios as customary in the literature. This implies that my sample size is large in the time-series and cross-sectional dimensions, which motivates the use of factor structures developed in a high-dimensional framework. The chapters are complementary to each other, although each studies a revised version of the research questions.

In the first essay, Chapter 1, I expand on the work of Borghi et al. (2018) by comparing the in-sample performance of their proposed ML estimator of time-varying sensitivities, which draws from Mikkelsen, Hillebrand, and Urga (2019), against the alternative rolling least square estimator. I take the factor model of Borghi et al. (2018) as given and, conditional on ex-post factors, I consider different windows to estimate the factor loadings via rolling OLS, for fixed sampling frequency. From a statistical perspective, my analysis is supported by a battery of misspecification tests based on the residual model-implied component, which depends only on different estimates of time-varying sensitivities. This setup allows to isolate the effects of the instability in the beta estimates in explaining contemporaneous return patterns, conditional on the estimated time-invariant factors.

The main finding of my analysis in Chapter 1 is that the rolling OLS estimator is characterised by a great degree of instability, which is driven by the interplay between window size and sampling frequency. If one selects the window size arbitrarily the instability in the estimates can be pronounced, and the benefits of fitting a dynamic model as opposed to an equivalent static-loadings representation become slim. I also find that the rolling OLS estimator based on a large window size enjoy similar properties to the ML estimator. However, from an economic perspective the rolling OLS estimates based on a large window size are difficult to interpret, since much of the short-term variation is averaged out. The opposite is true for the rolling estimates based on a short window size, which capture much of the short-term variation in asset prices at the expenses of a possibly misspecified model. Overall, the ML estimator of Borghi et al. (2018) dominates the rolling OLS under many aspects.

In the second essay, Chapter 2, I expand on the work presented in the first essay in two ways. I compare the explanatory power of the three-factor model of Borghi et al. (2018), that features a combination of observed and latent factors, against more traditional factors constructed from firm attributes, and I evaluate the out-of-sample predictive performance of rolling OLS betas in forecasting future return patterns, accounting explicitly for the window selection problem. I review the asset pricing literature that uses rolling OLS estimation to introduce the time variation in the factor loadings and discuss the choices taken in relation to the estimation window, which I vary from as little as 26 observations (half-year) to 520 observations (ten-year window). I employ a suite of performance measures that evaluate different aspects of the model fits, and I complement my analysis with an economic interpretation of the estimates by considering a rich set of global and region-specific financial events.

The main finding of my analysis in Chapter 2 points to a dual role of the rolling least square estimator of the factor sensitivities when I employ the measures in Kelly, Palhares, and Pruitt, 2021 to gauge the models performance. While the short-window approach provides the best results in a contemporaneous-equation setting (in-sample), for predictive purposes including too little observations for estimation causes the betas to be noisy, which in turn results in forecasts with little predictive power. Out-of-sample, the trade-off between the length of the window and the variance of the estimator is resolved around the two-year point, and this is true across all model specifications considered. I find that the choice of the window length alone accounts for about $\pm 10\%$ of the factor model's out-of-sample forecasting performance. The above is true when I use the measures of Kelly, Palhares, and Pruitt, 2021, but fails to be consistent for standard MSE and MSA functions. In this case, I find that for nearly all models their explanatory and forecasting performance is increasing in the window size, with the five- and ten-year windows providing the best results.

Analysing the relative performance of alternative model specifications that include observed factors, in Chapter 2 I find that the in-sample loadings of the FF three-factor model are on average statistically different from zero across all regions, while the two additional factors in their five-factor model affect only few of the groups considered. However based on the out-of-sample analysis I find that their performance deteriorates significantly, to the point that the average beta estimates constructed for window sizes up to the ten-year mark are not statistically different from zero for the SMB and HML factors. This is also true for the CMA and RMW factors in the five-factor model. This result may be due to the fact that the FF factors are not orthogonal to each other, which implies that they fail to isolate different sources of systematic variation in the returns, thus yielding statistically insignificant betas with the addition of new factors in the regressions, e.g. the redundancy of the HML factor when the CMA is added.

The third essay, Chapter 3, turns to the analysis of international sovereign bond returns. Similarly to the first two essays, I use asset-level data and higher-frequency returns (from 2010 to 2021) to document my findings, and most importantly I depart from the analysis of stock returns as customary in the literature and consider a different asset class. I review the literature on observed factor models in the context of international bond returns and identify the factors with characteristic-based portfolios such as momentum, value, and low-risk. A key challenge in modelling returns for a panel of international bond returns is that they feature two cross-sectional dimensions, issuer (country) and maturity-wise, and often times the literature neglects one of the two in the empirical applications (e.g. constant-maturity portfolios or issuer-specific portfolios), which makes it difficult to understand cross-sectional variation in the factor premia. I discuss the challenges in the portfolio construction phase using asset-level data and examine the cross-country performance of the factors in relation to the maturity buckets along the issuer-specific curves, thereby simultaneously considering the two dimensions.

The main finding of my analysis in Chapter 3 is that factor premia on momentum, value, and low-risk exhibit substantial variation across countries and maturities, which does not support the view of Asness, Moskowitz, and Pedersen (2013) and Frazzini and Pedersen (2014) on their unifying relevance and pricing ability across countries and asset classes. I find that risk-adjusted returns are decreasing in the maturity of the bonds for momentum strategies, with the highest Sharpe ratios found for portfolios formed on short-maturity bonds (less than 5 years). For longer-dated securities, momentum does not deliver statistically significant returns. When analysed across countries, my results reveal that momentum produces consistent statistically significant Sharpe ratios, however this is not true for value and low-risk. Using value measures based on past returns as in Asness, Moskowitz, and Pedersen (2013) leads to low and insignificant risk-adjusted returns across countries. Although lower in magnitude than momentum, low-risk yields statistically significant returns only for Euro Area bonds. Contrarily to what reported in previous literature, I find no supporting evidence for momentum, value and low-risk when bonds across all countries are considered in a global portfolio.

Understanding return patterns of financial assets internationally is a complex challenge which requires sophisticated modelling frameworks to account for the variation across asset classes, cross-section (e.g. country, industry, maturity-wise), and time. Paying tribute to such complexity, my dissertation includes a suite of empirical analyses and models to enhance our understanding of factor structures in asset pricing in a large sample context, using asset-level data at higher-frequency across countries and asset classes. In Section General Conclusion, I make the concluding remarks of this thesis and summarise further research questions that expand on the limitations of my analysis.

Chapter 1

Estimating Time-Varying Betas in Multi-Factor Asset Pricing Models

Abstract

This chapter expands on the work of Borghi et al. (2018) by comparing the in-sample performance of their proposed ML estimator of time-varying sensitivities against the alternative rolling least square estimator. I take the factor model of Borghi et al. (2018) as given and, conditional on ex-post factors, I consider different windows to estimate the factor loading via rolling OLS, for fixed sampling frequency. From a statistical perspective, I find that the rolling OLS estimator is characterised by a great degree of instability which is driven by the interplay between window size and sampling frequency. If one selects the window size arbitrarily, the instability in the beta estimates can be pronounced, and the benefits of fitting a dynamic model as opposed to an equivalent static-loadings representation become slim. From an economic perspective the rolling OLS estimates based on a large window size are difficult to interpret, since much of the short-term variation is averaged out. The opposite is true for the rolling estimates based on a short window size, in which much of the short-term variation is captured at the expenses of a possibly misspecified model. Overall, the ML estimator shows an unambiguously better in-sample performance. I document my findings considering a large panel of weekly stock returns from 40 different countries in the period from January 2006 to May 2019.

1.1 Introduction

A broad range of asset pricing studies focus on the analysis of contemporaneous return variation in international equities by employing models with observable pricing factors, and time-varying sensitivities estimated via rolling least squares, see e.g. Bekaert, Hodrick, and Zhang (2009), Bekaert et al. (2014), Fama and French (2012), or Fama and French (2017). A central debate in the literature revolves around the identification of these factors, with the first two papers advocating for value-weighted country-style portfolios, and the works of Fama and French using a combination of accounting measures and past-return performance indicators for their characteristics-based portfolios. The challenge in determining the most influential pricing factors can be solved empirically via latent factor models, which summarise a large amount of economic information without a-priori knowledge of the true factor space, see e.g. Gagliardini and Ma (2019), or Barigozzi, Hallin, and Soccorsi (2019).

Borghi et al. (2018) build on this literature by analysing international stock returns using a model that features a combination of observable and latent factors, the latter consistently estimated via PCA, and time-varying sensitivities estimated via ML. Their work draws from Mikkelsen, Hillebrand, and Urga (2019) who prove consistency of the PC estimator of the unknown factors, even in the presence of time variation in the loadings, and develop a consistent ML estimator of the time-varying loadings. The latter are modeled as stationary autoregressive processes and estimated in a state-space framework using the Kalman filter. From a methodological perspective, the contribution of Borghi et al. (2018) provides a competing modelling framework to existing studies in the international asset pricing literature in relation to estimation of the relevant factors, and estimation of time-varying factor sensitivities.

The main contribution of this chapter is to expand on the work of Borghi et al. (2018) by comparing the performance of their proposed ML estimator of time-varying sensitivities against the alternative rolling least square estimator, which is the status quo in the asset pricing literature. I take the factor model of Borghi et al. (2018) as given and, conditional on ex-post factors estimated via PCA, I consider different window sizes to estimate the factor loadings via OLS for fixed sampling frequency. From a statistical perspective, my analysis is supported by a battery of misspecification tests based on the residual model-implied component, which depends only on different estimates of time-varying sensitivities. This setup allows to isolate the effects of the instability in the beta estimates in explaining contemporaneous return patterns, conditional on the estimated time-invariant factors. I also briefly discuss the implications of changing the sampling frequency on the behavior of the rolling OLS estimator, keeping fixed the window size.

The secondary contribution of this chapter is to compare alternative identification procedures for the observed factor featured in Borghi et al. (2018). Their model features an observed financial factor that has an effect on all stocks (i.e. global observed factor), a global unobserved factor, and a regional unobserved factor. For the a-priori identification of the financial factor, I consider four stock market indexes and compare the explanatory power of a model featuring each candidate index as financial factor against a model in which all three factors are unobserved. The PCA estimator of the latent factors maximises the share of explained variance by definition, and as such is the benchmark for my analysis. Among the candidate factors, I identify the financial factor with the S&P500 Financials index which provides the best results among the candidate models from a statistical perspective and also eases economic interpretability.

A further contribution of this chapter is to reconcile the model of Borghi et al. (2018), which is developed in a contemporaneous-equation setup, with the one of Inoue, Jin, and Rossi (2017) who develop an optimal window selection criterion in rolling out-of-sample forecasting. This would allow me to have a unifying modelling framework to compare the performance of different estimators of time-varying sensitivities, however in Appendix A.1 I show why the optimality of their criterion fails to hold in a contemporeaneous-equation setting, due to the properties of conditional expectations. This problem remains an open research question that I leave for future studies, see Section 1.4.1.

Additionally, I replicate the study of Borghi et al. (2018) on an extended time frame, from January 2006 to May 2019 instead of from 2002 to 2016, which provides an additional layer of robustness to their findings in an 'out-of-sample' context. Similarly to Borghi et al. (2018), I use individual stock returns as base assets, instead of stock portfolios as standard practice in the literature, which necessarily increases the number of test assets under consideration. Without grouping assets into portfolios, the number of test assets in my analysis is at least two orders of magnitude greater than what is commonly used in the literature[1], which motivates the use of large-dimensional factor models for the analysis of individual stock returns. Secondly, I depart from the monthly benchmark popularised by the works of Fama and French, and choose to consider weekly data as in Bekaert et al. (2014). The regional classification of the stock universe also follows their framework.

My analysis yields a number of results which point to a superior performance of the ML estimator of time-varying loadings in Mikkelsen, Hillebrand, and Urga (2019) with respect to the rolling OLS estimator. Keeping fixed the sampling frequency of the data to weekly, I find great variation in the statistical properties of a dynamic-loadings model with respect to an equivalent static-loadings representation when I use two windows made of the most-recent half- and five-year observations to estimate the loadings via rolling OLS. This is not the case when the ML estimator of the time-varying loadings is used in the dynamic-loadings representation, which dominates the static counterpart under many aspects. Moreover, while the five-year rolling estimator enjoys similar properties to the ML estimator, from an economic perspective the rolling

[1] I consider thousands of stocks coming from 40 different countries, as opposed to 10-50 portfolios that are often used as test assets in the literature.

OLS estimates based on a large window size are difficult to interpret, since much of the short-term variation is averaged out. I find that the opposite is true for the rolling estimates based on a short window size, in which much the short-term variation in asset prices is captured at the expenses of a possibly misspecified model.

When I replicate the analysis of Borghi et al. (2018) on an extend time sample, I expand on several of their initial findings. Firstly, when loadings are estimated via ML, I corroborate the evidence that the relative importance of the factors is time-varying. When unexpected events happen globally, stock return co-movements increase, and stocks tend to become marginally more exposed to financial and global shocks, which I am able to map to relevant macro events. This is not necessarily true when the loadings are estimated via rolling least squares. In fact, while the the estimates constructed from a short window have a great degree of variability and appear to capture changing market conditions similarly to the ML estimator, the variance of the estimates constructed from a five-year window is too low to allow for a consistent mapping of the events. Secondly, replicating the analysis of Borghi et al. (2018) on the relationship between expected returns and beta parameters, I find no supporting evidence for the existence of a premium for holding stocks with highly volatile factor exposures. See Appendix A.2 for further details.

Organisation of the chapter. The remainder of this chapter is organised as follows. Section 1.2 describes the models for the factor extraction procedure, and estimation of time-varying factor sensitivities via ML and rolling least squares. Section 1.3 presents the data and reports the bulk of results. I firstly present the results on the estimation of the latent factors via PCA, and subsequently compare the estimation of time-varying betas via the two methods. Finally, Section 1.4 makes the closing remarks and details future research developments. The chapter is accompanied by Appendix A.

1.2 Methodology

In this section I describe the models for the factor extraction procedure and estimation of time-varying factor sensitivities. As a starting point I define the benchmark model featuring unobserved global and regional risk drivers and static factor sensitivities, Section 1.2.1. I then introduce the factor extraction procedure via PCA, Section 1.2.2, and lastly in Section 1.2.3 I relax the assumption of static factor betas and present two approaches for the the estimation of time-varying factor sensitivities, ML estimation via the Kalman filter and rolling OLS estimation.

1.2.1 Baseline Model

I consider an approximate factor model for the analysis of stock returns. This means that I relax the baseline framework of classical factor models[2] and work in a framework in which the cross-sectional dimension N and the number of time periods T are both large, the idiosyncratic errors are uncorrelated across i, $i = 1, ..., N$, but the data are correlated across t, $t = 1, ..., T$,

$$N, T \to \infty \tag{1.1}$$

$$E[\boldsymbol{e}^\top \boldsymbol{e}] = diag(\psi_1, ..., \psi_N) \tag{1.2}$$

$$E[X_t^\top X_{t-l}] = \rho_{l,t}, \quad l = 1, ..., t \tag{1.3}$$

with $\boldsymbol{e}$ being the $(T \times N)$ matrix of idiosyncratic components, ψ_i the idiosyncratic variance for stock i, X_t the N-dimensional vector of t-period log-returns, and $\rho_{l,t}$ the N-dimensional vector of autocorrelation coefficients of asset returns at time t for lag l.

I assume that a 'small' number of K unobserved factors captures the systematic variation in stock returns of N assets over T time periods,

$$\underbrace{\boldsymbol{X}}_{(T\times N)} = \underbrace{\boldsymbol{F}}_{(T\times K)} \underbrace{\boldsymbol{\Lambda}^\top}_{(K\times N)} + \underbrace{\boldsymbol{e}}_{(T\times N)} \tag{1.4}$$

$$\iff X_{i,t} = F_t \, \Lambda_i^\top + e_{i,t} \tag{1.5}$$

where $X_{i,t}$ is the t-period log-return of stock i, F_t are the K factor realisations at time t, Λ_i is the vector of static factor loadings, and $e_{i,t}$ is the stock-specific t-period residual component. In matrix notation, $\boldsymbol{X}$ is the $(T \times N)$ matrix of returns, $\boldsymbol{F}$ the $(T \times K)$ matrix of factor returns, $\boldsymbol{\Lambda}$ the $(N \times K)$ matrix of factor sensitivities, and $\boldsymbol{e}$ the $(T \times N)$ matrix of idiosyncratic components.

[2] I follow the canonical definition of classical factor models given in namely iid stock-individual idiosyncratic component, cross-sectional dimension fixed, time periods going to infinity, and normally distributed factors and stock-individual residual components.

I also require a strong form of APT to hold, where residual risk has a premium of zero, $E[\boldsymbol{e}|\boldsymbol{X}] = 0_N$. The key requirement of the APT is that idiosyncratic risk can be diversified away with a sufficiently large number of assets, an assumption that is common to the asset pricing literature using portfolios as base assets. For the purpose of this chapter I maintain this assumption and take the factor model as given, being my focus on the estimation of time-varying loadings (and not on the estimation of the number of factors K). In its weak form in fact, the ATP implies that the specific number and nature of the factors is unknown and need to be estimated, see for instance the setup of Gagliardini and Ma (2019). Only upon correct specification of the factor space, residual risk premium is zero and there are no sources of exploitable opportunities. In this chapter I use individual stock returns as base assets and rely on a relatively standard set of assumptions regarding the nature of the factor space. I take the factor model of Borghi et al. (2018) as given and assume that a fixed number of K factors are uniquely identified and agreed upon.

Under the APT, investors cannot generate arbitrage profits by trading on publicly available information since stock returns are fully determined by their exposure to the common risk factors. I therefore assume no sources of exploitable opportunities conditional on the factors, which translates into having a model with zero alpha. Additionally, given that the data is de-meaned prior to the analysis I treat the data matrix as having zero mean, $\bar{X} = E[\boldsymbol{X}] = 0_N$, an assumption that is common in the factor model literature, se e.g. Connor and Korajczyk (1986), or Stock and Watson (2002). It follows that also $\bar{F} = E[F] = 0$ given that the factor extraction procedure is based on de-meaned data, see Section 1.2.2 for further details.

Under the assumption of weak exogeneity between the factors and residuals[3], the covariance matrix of stock returns can be decomposed into a systematic and idiosyncratic component

$$Var(\boldsymbol{X}) = \boldsymbol{\Lambda} Var(\boldsymbol{F})\boldsymbol{\Lambda}^\top + Var(\boldsymbol{e}). \tag{1.6}$$

where $Var(\boldsymbol{X}) = \boldsymbol{\Sigma_X} = E[X^\top X]$ is the variance-covariance matrix of returns, similarly $Var(\boldsymbol{F}) = \boldsymbol{\Sigma_F} = E[F^\top F]$ and $Var(\boldsymbol{e}) = \boldsymbol{\Sigma_e} = E[e^\top e]$ are the variance-covariance matrices of factors and idiosyncratic component respectively.

I follow the framework of Bekaert et al. (2014) and distinguish between K_{glob} global and K_{reg} regional drivers of systematic variation, $K = K_{glob} + K_{reg}$. In particular, I divide the N stocks into regions R regions, with each region containing N_r securities: $\sum_{r=1}^{R} N_r = N$. The model could easily accommodate multiple regional risk drivers, but for simplicity I assume that there is one factor for each region, $K_{reg} = R$. The total number of factors is therefore $K = K_{glob} + R$.

[3]Note that in the APT framework where $E[\boldsymbol{e}|\boldsymbol{X}] = 0_N$ and factors are correctly specified, it follows implicitly that there is zero correlation between the factors are residuals, $E[\boldsymbol{X}^\top \boldsymbol{e}] = \boldsymbol{0}_N$, i.e. weak exogeneity. By the definition of covariance, $Cov(\boldsymbol{X}, \boldsymbol{e}) = E[\boldsymbol{X}^\top \boldsymbol{e}] - E[\boldsymbol{X}]^\top E[\boldsymbol{e}]$, where $E[\boldsymbol{X}^\top \boldsymbol{e}] = E_X[E_e[\boldsymbol{X}^\top \boldsymbol{e}|\boldsymbol{X}]] = E_X[X^\top E[\boldsymbol{e}|\boldsymbol{X}]] = E_X[\boldsymbol{X}^\top 0_N] = \boldsymbol{0}_N$. Without loss of generality we can assume that also that $E[\boldsymbol{X}] = 0_N$.

The model for each stock i can be rewritten as

$$X_{i,t} = F_t^{glob}\ \Lambda_i^{glob\top} + \sum_{r=1}^{R} \lambda_i^{reg}\ F_{r,t}^{reg}\ \mathbf{1}_{\{i \in r\}} + e_{i,t} \tag{1.7}$$

where F_t^{glob} is the $(1 \times K_{glob})$ vector of systematic global risk drivers at t, $\Lambda_i^{glob\top}$ the respective $(K_{glob} \times 1)$ loadings vector, $\Lambda_i^{glob} = (\lambda_{1,i}^{glob}, ..., \lambda_{K_{glob},i}^{glob})^\top$, λ_i^{reg} is the loading on the regional risk factor for stock i belonging to region r, $F_{r,t}^{reg}$ is the realisation at time t of the r-th regional factor.

The formulation in (1.4) implies that the $(K \times N)$ loading matrix $\mathbf{\Lambda}^\top$, formed by concatenation of the N loading vectors, is sparse given the different regional characteristics of the stocks, $\mathbf{\Lambda} = (\Lambda_1, ..., \Lambda_N)^\top$. Λ_i is the K dimensional vector that contains the same number of non-zero elements across i. Consider for instance the loading vector for the first two equities which, for illustrative purposes, I assume to belong to two different regions: $\Lambda_1 = (\lambda_{1,1}^{glob}, ..., \lambda_{K_{glob},1}^{glob}, \lambda_1^{reg}, 0, ..., 0)$ and $\Lambda_2 = (\lambda_{1,2}^{glob}, ..., \lambda_{K_{glob},2}^{glob}, 0, \lambda_2^{reg}, 0, ...)$. Using this logic, the model for the N stocks grouped in R regions can be written in a more compact form as

$$X_t = \Lambda^{glob\top}\ F_t^{glob} + \Lambda^{reg\top}\ F_t^{reg} + e_t \tag{1.8}$$

where $X_t = (X_{1,t}, ..., X_{R,t})^\top$ with each $X_{r,t}$ being N_r-dimensional, $\Lambda^{glob} = (\Lambda_1^{glob}, ..., \Lambda_R^{glob})$ is the $(N \times K^{glob})$ matrix of loads for the global factor realisations F_t^{glob}, grouped by blocks of stocks in region r, each Λ_r^{glob} is $(K^{glob} \times N_r)$. $\Lambda^{reg} = (\lambda_1^{reg}, 0, ...)^\top$ refers to the $(N \times K^{reg})$ sparse matrix of factor sensitivities for the R regional drivers F_t^{reg}. Finally,

$$X_t = \mathbf{\Lambda}\ F_t + e_t \tag{1.9}$$

with $\mathbf{\Lambda} = (\Lambda^{glob}, \Lambda^{reg})^\top$ being the $(N \times K)$ (sparse) matrix of factor loadings and $F_t = (F_t^{glob}, F_t^{reg})$ the K factor realisations at time t.

1.2.2 Principal Component Estimation of the Latent Factors

In the APT, factors and loadings in equation (1.9) are unknown and have to be estimated. Under a large T and N setup, it is possible to estimate $\mathbf{\Lambda}$ and F_t simultaneously via PCA. In contrast, classical factor analysis requires the estimation of $\mathbf{\Lambda}$, under fixed N, and of the covariance matrix of idiosyncratic errors, $\mathbf{\Sigma}_e$, which is assumed to be diagonal. Given $\mathbf{\Lambda}$, F_t is then estimated at the second stage, however the estimate of F_t is not consistent under fixed N.

Estimators of the latent factors include the asymptotic PCs of Connor and Korajczyk (1986) who build on the theory of approximate factor models of Chamberlain (1983) and develop an estimator for the first K eigenvector of the $(T \times T)$ cross-product matrix (instead of the $(N \times N)$ covariance matrix of asset returns), showing that in a large N setup, the first K eigenvectors of this cross-product matrix are consistent estimates of the $(K \times T)$ matrix of factor returns. Stock and Watson (2002) extend the Connor and Korajczyk (1986) framework to a large N and T setup and to time-varying factor betas. Recently, Mikkelsen, Hillebrand, and Urga (2019) prove that the PCs uniformly converge in t when $T/N^2 \to 0$ in a high-dimensional factor model with time-varying loadings, they extend the results of Bates et al. (2013) who prove the average convergence in t of the PCs to the true factor space. Lettau and Pelger (2020b) employ the risk-premium PCA estimator to study a large cross section of stock returns, their estimator generalises PCA by including a penalty on the pricing error in expected returns, and it is developed under $N/T \to c$ setup featuring static factor loads. See Lettau and Pelger (2020b) for the theoretical results.

The PCA estimator of the latent factors attempts to minimise the residual time-series variation in the returns by solving the following optimisation problem

$$\begin{aligned} \min_{\hat{\Lambda}_{PCA}, \hat{F}_{PCA}} \quad & RSS(K) \\ \text{with} \quad & RSS(K) = \frac{1}{T}\sum_{t=1}^{T}(X_t - \mathbf{\Lambda}\, F_t)^\top (X_t - \mathbf{\Lambda}\, F_t) \end{aligned} \tag{1.10}$$

Statistical factor analysis conventionally applies PCA to the sample covariance matrix $\frac{1}{T}(X^\top X) - (\bar{X}\bar{X}^\top)$. Alternatively, asset returns are usually normalised by their standard deviation so that $diag(\mathbf{\Sigma_X}) = 1$, which is equivalent to PCA applied to the correlation matrix. In practice, the data matrix X is also demeaned before PCA is applied, so that the objective function in equation (1.10) does not depend on the means of the test assets, concordant with the assumption of $\bar{X} = 0$. In this framework the estimated factors $\hat{F}_{PCA}$ have zero mean and unitary variance by construction.

In equation (1.8), factors F_t^{glob}, F_t^{reg} and the respective sensitivities $\Lambda^{glob\top}$, $\Lambda^{reg\top}$ are unobserved. Thus, to disentangle the effect of global and regional risk drivers, I need to impose the following identifying restrictions to obtain a unique solution:

IR1 $T^{-1}\sum_{t=1}^{T}(F_{k,t}^{glob})^2 = 1$ for all $k \in K^{glob}$, and $T^{-1}\sum_{t=1}^{T}(F_{r,t}^{reg})^2 = 1$ for all $r \in R$. Normalising the estimated factors to have unit length allows me to compare them.

IR2 $T^{-1}\sum_{t=1}^{T} F_{r,t}^{reg}\ F_{k,t}^{glob\top} = 0$ for all $r \in R$ and $k \in K^{glob}$. This ensures regional factors are orthogonal to the global ones.

IR3 $\sum_{t=1}^{T} F_{r,t}^{reg}\ S_{r,t}^\top > 0$, where $S_{k,t}$ is the biggest country's stock market index return at time t in region r. This identifies the sign of the factors by imposing positive correlation with the most important stock index of the region, whicheliminates the rotation indeterminacy and allows to interpret the sign of the factor loadings. This assumption is taken from Breitung and Eickmeier (2015).

1.2.3 Time-Varying Factor Sensitivities

Conditional on the estimated factor space, I now relax the static-loadings assumption and introduce two approaches for the estimation of time-varying factor betas, rolling OLS estimation and ML estimation via the Kalman filter.

Rolling OLS Estimation

The rolling OLS estimator is the status quo in the Finance literature to address parameters instability and is based on the idea of using the most recent observations to estimate the parameters, instead of using all available observations as in standard OLS regressions. The choice of how many observations should be used to yield the best predictor is often judgmental, for instance Fama and French (2012) use a ten-year window to estimate the time-varying slopes of their three-factor model using monthly data, Bekaert, Hodrick, and Zhang (2009) rely on half-year window regressions to examine international stock return co-movements with weekly data, while Armstrong, Banerjee, and Corona (2013) employ a five-year window in their CAPM framework for US stocks with monthly data. Suppose the sample size is T and I am choosing to retain W observations. With the number of increments between successive window being one period (day, week, or month), the data set is partitioned into $T_W = T - W + 1$ subsamples, each made of W observations. The standard approach in the literature has been to roll one or more observations ahead and estimate the beta on overlapping windows.

Given the the wide-spread use in the literature of the rolling OLS estimator of factor betas, I review the modelling choices underpinning this approach, accounting explicitly for the window selection problem. To do so I consider the modelling setup of Inoue, Jin, and Rossi (2017) who develop an optimal window selection criterion in rolling out-of-sample forecasting, and treats the rolling OLS estimator as nonparametric. This is because in the OLS framework, the factor sensitivities are assumed to be locally constant[4], and as such the rolling OLS estimator of the factor loadings can be thought as a non-parametric OLS estimator where the window size plays the role of the bandwidth. The main limitation of reconciling the approach in Borghi et al. (2018) with Inoue, Jin, and Rossi (2017) lies in the fact that the former features a contemporaneous-equation setting while the latter is predictive, which makes the optimal window criterion invalid. In Appendix A I show why this reconciliation is difficult due to the properties of conditional expectations.

Rewriting model (1.7) using the Inoue, Jin, and Rossi (2017) framework yields

$$X_{i,t} = F_t^{glob}\ \Lambda_i^{glob}(t/T)^\top + \sum_{r=1}^{R} \lambda_i^{reg}(t/T)\ F_{r,t}^{reg}\ \mathbf{1}_{\{i \in r\}} + e_{i,t} \tag{1.11}$$

[4]The factor betas are assumed to be constant within a given estimation window.

where $\Lambda_i^{glob}(t/T)^\top$ is the $(K_{glob} \times 1)$ vector of unknown functions of t which defines the relationship between the returns of stock i with the vector of global factor realisations, F_t^{glob}. The same holds for $\lambda_i^{reg}(t/T)$, and the respective regional factor $F_{r,t}^{reg}$. The main assumption of this framework is that the matrices of unknown functions of t that introduce the time variation in the loadings are defined on an equally spaced grid over the support $(0, 1]$, and the grid becomes finer as $T \to \infty$. This requirement is common to the non-parametric estimation literature in which the amount of local information on which an estimator depends increases suitably as the sample size T increases. Using the most recent W observations, the rolling OLS estimator of the time-varying factor sensitivities based on the information set at time $t = T$ is given by

$$\hat{\Lambda}_{i,W}^{glob}(1) = \hat{\Lambda}_{i,W}^{glob}(T/T) = \left(\sum_{t=T_W}^{T} F_t^{glob\top} F_t^{glob} \right)^{-1} \left(\sum_{t=T_W}^{T} F_t^{glob\top} X_{i,t} \right) \tag{1.12}$$

$$\hat{\lambda}_{i,W}^{reg}(1) = \hat{\lambda}_{i,W}^{reg}(T/T) = \left(\sum_{t=T_W}^{T} (F_{r,t}^{reg})^2 \right)^{-1} \left(\sum_{t=T_W}^{T} F_{r,t}^{reg} X_{i,t} \right). \tag{1.13}$$

Similarly to the conditions required on the full sample, I assume that within each window W the factor realisations and the data matrix have zero mean.

Maximum Likelihood Estimation

Borghi et al. (2018) propose a two-level factor model with time-varying loadings to investigate the dynamics of factor betas in the cross-section of a large panel of stock returns. Their work exploits the modelling approach of Mikkelsen, Hillebrand, and Urga (2019), who develop a consistent two-step ML estimator of time-varying loadings in a high-dimensional setting. The loadings are allowed to temporarily depart from their long-run averages and evolve according to stationary autoregressive processes.

The $(N \times K)$ matrix $\mathbf{\Lambda}_t$ made of the N firm-individual vectors of time-varying factor sensitivities $\mathbf{\Lambda}_t = (\Lambda_{1,t}, ..., \Lambda_{N,t})^\top$ is composed of the K^{glob} loadings on the global factor, Λ_i^{glob}, and the R factor sensitivities for the regional factors, which for each stock i contains one non-zero entry only λ_i^{reg}. In light of equation (1.7), $\Lambda_{i,t}$ is K-dimensional and contains the same number of non-zero elements across i, due to the different regional characteristics of the stocks. In this framework, the non-zero elements of $\Lambda_{i,t}$ evolve according to

$$\Lambda_{i,t} = (1 - \Phi_i)\bar{\Lambda}_i + \Phi_i \Lambda_{i,t-1} + \eta_{i,t} \tag{1.14}$$

where $\bar{\Lambda}_i = E[\Lambda_{i,t}] = (\bar{\lambda}_{i,1}^{glob}, ..., \bar{\lambda}_{i,K_{glob}}^{glob}, \bar{\lambda}_i^{reg})^\top$ is the unconditional mean vector of factor sensitivities, $\Phi_i = diag(\phi_{i,1}^{glob}, ..., \phi_{i,K_{glob}}^{glob}, \phi_i^{reg})$ is the persistence parameter matrix, and the characteristic roots of equation (1.14) lie outside the unit circle. $Q_i = E[\eta_{i,t}\eta_{i,t}^\top] = (q_{i,1}^{glob}, ..., q_{i,K_{glob}}^{glob}, q_i^{reg})$ is the covariance matrix of the innovations $\eta_{i,t}$, which is assumed to be a Gaussian white noise process.

Under these conditions, the loadings of stock i on the factors evolve as independent AR(1) processes, around their respective unconditional means, with AR coefficient ϕ_i^f, $f \in \{1, ..., K_{glob}, 1, ..., R\}$, and stationarity condition $|\phi_i^f| < 1$ respected for all f. Stationarity of the loadings on market factors have been documented, among the many, by Andersen et al. (2006) and Armstrong, Banerjee, and Corona (2013). Borghi et al. (2018) extend this setting to the case of global and regional factors.

I now turn to the MLE of Φ_i, Λ_i, Q_i and ψ_i for $i = 1, ..., N$, following the theory developed by Mikkelsen, Hillebrand, and Urga (2019) on consistent estimation of the unknown parameters in a two-level factor model analogous to the one I employ. A key result of Mikkelsen, Hillebrand, and Urga (2019) is that the feasible likelihood function, with unobserved factors replaced by PCs, converges uniformly to the infeasible one, even in the presence of estimation error in the principal components and time-variation in the loadings. In the setup of Borghi et al. (2018) however, global and regional factors can be cross-sectional dependent, considering that the firms in the universe are partitioned into geographical areas. This feature can potentially render the ML estimator of the unknown parameters complicated, since I would have to take into account cross-sectional dependence across idiosyncratic errors in the likelihood function. To overcome this problem and simplify the estimation procedure, I exploit the fact that the ML estimator of the unknown parameters Φ_i, Λ_i, Q_i and ψ_i remains consistent in the presence of cross-sectional and temporal dependence in the errors. I refer to Mikkelsen, Hillebrand, and Urga (2019) for details on the estimation procedure.

Thus, conditional on the factors, one can treat X_i as uncorrelated across stocks and the likelihood function can be analysed separately for each i. Therefore, if X_i is T-dimensional vector of time-series observations for stock i, I can rewrite equation (1.9) for the cross-section of stock returns as

$$X_i = \hat{\boldsymbol{F}}^* \Lambda_i^* + e_i \tag{1.15}$$

where $\hat{\boldsymbol{F}}^* = diag(\hat{F}_1^\top, ..., \hat{F}_T^\top)$ is a $(T \times TK)$ block-diagonal matrix that stacks the time-series observations on the estimated factors, with diagonal elements given by the observations of each factor at time t, $\hat{F}_t = (\hat{F}_{1,t}^{glob}, ..., \hat{F}_{K_{glob},t}^{glob}, \hat{F}_{1,t}^{reg}, ..., \hat{F}_{R,t}^{reg})^\top$. $\Lambda_i^* = (\Lambda_{i,1}^\top, ..., \Lambda_{i,T}^\top)$ is the TK-dimensional vector of time-varying loadings for stock i. Under the assumption of normally distributed idiosyncratic errors, the feasible likelihood function for X_i is Gaussian and conditional on the estimated factors $\hat{\boldsymbol{F}}^*$ it can be separated for each stock i

$$\hat{L}_T(X_i|\hat{\boldsymbol{F}}^*;\theta_i) = -\frac{1}{2}log(2\pi) - \frac{1}{2T}log(|\boldsymbol{\Sigma}_i|) - \frac{1}{2T}(X_i - E[X_i])^\top \boldsymbol{\Sigma}_i^{-1}(X_i - E[X_i]) \tag{1.16}$$

with parameter vector $\theta_i = \{\Phi_i, \Lambda_i, Q_i, \psi_i\}$. $E[X_i]$ is the T-dimensional mean vector of X_i, which is zero in my setup since the data is de-meaned, and the covariance matrix of X_i is $\mathbf{\Sigma}_i = Var(X_i) = \hat{\boldsymbol{F}} Var(\Lambda_i^*) \hat{\boldsymbol{F}}^\top + \psi_i I_T$. Finally, the ML estimator of θ_i is given by the following maximisation problem

$$\hat{\theta}_i = \underset{\theta}{\operatorname{argmax}} \; \hat{L}_T(X_i | \hat{\boldsymbol{F}}^*; \theta_i). \tag{1.17}$$

In my empirical applications, I interpret equations (1.14) and (1.15) as transition and measurement equations in a linear state-space model, where the factor loadings and their parameters are unobserved stationary states, and the likelihood function is maximised via the Kalman filter.

1.3 Data and Model Fit

In this section I describe the data sources for the analysis and present the bulk of results. In Section 1.3.1 I describe the characteristics of my international stock universe, in Section 1.3.2 I identify the observed financial factor and report the results of the estimation of the unknown factors via PCA, and finally in Section 1.3.3 I compare the results on the estimation of time-varying factor sensitivities via ML and rolling OLS.

1.3.1 Data Description

Stock Universe

A total of 3294 equities entered the national stock market indexes from January 6th 2006 to May 31st 2019. Among these, I select 2873 tickers with no less than two years of data, and no more than eight consecutive missing observations. I apply linear interpolation for the missing data. Throughout the 13-year period considered, there were 1692 stocks that remained part of the indexes, while the remaining 1181 stocks were de-listed at various points in the sample period. The main data source is Bloomberg. Stock prices refer to the last transaction of the week, and balance sheet data is available at quarterly frequency. Prices are expressed in US dollars, and are ex-dividends and split-adjusted. Returns are defined as the first differences of the natural logarithm of stock prices, unless stated otherwise. Prior to my analysis, I winsorise the individual price series at 95% level to avoid the effects of possible data entry errors.

Table 1.1 reports detailed information on my universe of securities, which are listed in 40 different countries. The countries are then grouped into six geographical regions: North America, Latin America, Asia-Pacific, Western Europe, Emerging Europe and Middle-East & Africa (MEA). I take the geographical classification as given, following the framework of Bekaert et al., 2014.

[Table 1.1 about here.]

The factor extraction procedure as well as the estimation of time-varying betas using the Kalman filter are performed on the cross-section of stocks active on the full sample, 1692 equities. On the other hand, when I estimate the time-varying factor sensitivities using rolling least squares, I work with an unbalanced panel of stock returns. Figure 1.1 describes the time-varying composition of the universe. Depending on the chosen window size, I estimate the factor sensitivities on a restricted time span with respect to the full sample, January 2006 to May 2019. When I employ a window made of the most recent five-year observations, the rolling beta estimates are available for the period January 2011 to May 2019. On the other hand, in the short window approach the sub-sample starts on July 2006 and ends on May 2019. For the rolling window estimation, I select all the equities that entered the national stock indexes with at least two years of available data on the full sample, and no more than eight consecutive missing observations.

[Figure 1.1 about here.]

Summary Statistics

Table 1.2 reports the summary statistics of returns, market capitalisation, total assets and total debt of the 1692 companies active in the sample period. Following the Bloomberg 10-sector classification structure, I divide stocks in the following groups: Basic Materials, Communications, Consumer Cyclical, Consumer Non-Cyclical, Diversified, Energy, Financial, Industrial, Technology and Utilities. In panel 1.2a I report the average pair-wise Pearson correlation coefficient for the stocks belonging to each sector or region to have an estimate of the intra-group dependence of the equities. The Middle East & Africa region has the lowest average value, with a coefficient of 0.184, in line with the economic diversity of this area. The highest intra-group dependence is found in North America and Western Europe, with an average correlation of 0.351 and 0.406 respectively. The sector with the highest value is Energy, with an average linear dependence of 0.351.

In panel 1.2b I report simple averages of the balance sheet information for the retained stocks. The order statistics appear to be in line with expectations. In fact, North American and Western European stocks have the highest market capitalisation, together with Energy stocks. I also notice that Energy, Communications and Utilities companies have highest assets (Financial stocks excluded), in agreement with the considerable investments in infrastructures and operations needed for the business. The data filters that I apply can also help prevent the inclusion of ‘micro-cap’ stocks. Given the requirement for the stocks to remain listed throughout the 13-year period in the national equity indices, which usually comprise the top 20/50 names in each country, the minimum average market capitalisation across regions is in fact about 1.5 Billion USD for MEA and Emerging Europe.

[Table 1.2 about here.]

Data Frequency

The frequency of my data is weekly. The choice of data frequency and window size are intrinsically related and depend on the assumptions I am prepared to make on the behavior of the factor betas. Consider for instance the standard monthly benchmark commonly used in the empirical asset pricing literature, with an estimation window typically made of the most recent five or ten years of data. As Robertson (2018) reports, if the estimated betas vary slowly with respect to the sampling frequency then the betas in the immediate future can be approximated with current values. Under this logic, OLS on the most recent data is the candidate estimation method. The choice of the window length determines how much of the recent short term-variation in the betas is incorporated into the current-period estimates, for fixed observation frequency. If I use a relatively short window of six months, the within-window constant-loadings assumption underpinning the OLS estimator is more likely to be respected, at the expenses of estimation accuracy which calls for a sufficiently long history of data for estimation. Analogously, when I let the observation frequency vary for a fixed window length, I expect the betas constructed from data sampled at higher frequency (e.g. daily, weekly) to exhibit less short-term variation with respect to the monthly benchmark. However, if I am interested in studying beta over longer horizons, I have to take into account the possibility of some longer-run mean-reverting behavior in the beta estimates and in this case the appropriate modelling choice resolves to the stationary VAR model of Mikkelsen, Hillebrand, and Urga (2019), see equation (1.14).

[Figure 1.2 about here.]

The interplay between sampling frequency and window size determines the properties of the rolling OLS estimator. Figure 1.2 reports the results of a study that analyses the behavior of the rolling beta estimates for varying sampling frequency, keeping fixed the window size. I choose a window made of the most recent half-year observations to capture the short-term variability in the beta estimates. I then estimate the market beta for the returns on Apple Inc. over a 13 years span, considering different sampling frequencies (daily, weekly, monthly).

Panels 1.2d and 1.2e report the histogram of the estimated rolling betas, together with the summary statistics for different frequency. As Robertson (2018) reports, were beta strictly constant then I should obtain quantitatively similar estimates from daily, weekly and monthly data. I notice however that the variability of the beta estimates decreases with the sampling frequency. Over very short horizons (e.g. daily) the estimated betas have a standard deviation of 0.157, lowering the frequency to weekly shows that the variability of the market factor sensitivity increases to 0.185, and finally with data sampled every month the variability reaches 0.313, which is almost twice as large as in the daily case. From the summary statistics of the estimated betas, I find that on average the rolling estimator is centered around a long-run mean for varying sampling frequency, however from the higher-moment statistics I observe a great degree of instability, with the distribution of estimated betas becoming progressively skewed towards negative values and showing increasing evidence of fat tails for lower frequencies.

All things considered, panel 1.2c seems to suggest that a great portion of the high-frequency variation in the betas is averaged out when considering a monthly benchmark. This might explain why I see substantial instability in the monthly series of beta estimates, where the addition of a single month can generate significant variation in the estimated factor sensitivities. To tackle this problem I depart from the standard monthly benchmark, and opt for the weekly holding-period returns which allow me to provide a deeper analysis on the time-varying behavior of the factor sensitivities, without compromising on the bias induced by market microstructure at daily frequency[5], and on the pronounced instability of the beta estimates at

[5]The issue is relevant especially in the analysis of a large international panel of stock returns from different geographical areas due to the asynchronicity on the opening times of the various national stock indexes.

monthly frequency.

1.3.2 Observed and Unobserved Factors

In this section I identify the observed financial factor with the S&P 500 Financials, and estimate the latent global and regional factors.

Identification of Financial Factor

In Section 1.2 I considered a framework in which the K factors are unknown, as in APT. I now proceed to the identification of the global financial factor, which I assume is a source of risk common to all stocks in my universe. I do so by studying the information content of the remaining $K-1$ (latent) factors when I identify the financial factor with different stock market indexes.

The number of factors is assumed to be $K = 3$. My setup closely follows Bekaert et al. (2014) who show that the market factor plays an important role in explaining the cross-section of returns together with global and one region-specific (latent) risk drivers. In this study I do not focus on determining the number of systematic risk factors, thus I take K as given. Several papers document the superiority of a three-factor model as opposed to a model featuring four or more factors, e.g. Lettau and Pelger (2020b) shows that for 270 individual stocks PCA with $K = 3$ dominates other specifications with $K = 5$ in terms of unexplained residual variance, both in-sample and out-of-sample. Gagliardini and Ma (2019) propose a model in which the number of systematic factors is time-varying and they estimate it via machine learning methods. Their results indicate that the number of factors fluctuates between two and six from 1975 to 2015, however their method depends on a number of nuisance parameters, e.g. bandwidth, penalisation function, which makes the results vulnerable to alternative modelling choices. The authors suggest that while the number of systematic factors can be considered time-varying, when results are smoothed the time-variation is limited and the number of conditional factors ranges between 1 and 2 in most periods. They consider US stock returns at monthly frequency.

Table 1.3 reports cross-sectional averages of the OLS estimates on the full sample of the factor betas, for various model specifications. I firstly consider a model featuring three unobserved factors with $K_{glob} = 2$ and $K_{reg} = R$, one for each region. I then proceed to the identification of the first unobserved global factor considering four different candidate stock market indexes, and study the information content of the remaining two (unobserved) factors.

[Table 1.3 about here.]

I find that the estimation of the second global factor F_2^{glob} is problematic, especially when considering cross-sectional industry averages. This may be due to the fact that standard PCA may fail to detect weak factors as Lettau and Pelger (2020b) reports. According to the authors' definition, 'strong' factors affect a large number of assets, e.g. the market portfolio, in contrast 'weak' ones affect only a small subset of assets. To address the problem estimating weak factors, the authors develop an alternative PCA estimator, RP-PCA, which uses information contained in the first moment of the test assets. Their estimator imposes a penalty on the residual cross-sectional pricing error of the model and is showed to provide a higher 'signal-to-noise' ratio (in terms of difference between the eigenvalue of the weakest factor and largest idiosyncratic eigenvalue) which eases the extraction of these factors.

When I identify the financial factor with the four stock market indexes $F_1^{glob} = F^{fin}$, the estimation of the (uncorrelated) global factor gives more consistent results, with respect to the full unobserved case. The differences in the relative importance of the global factors in these instances is driven by the information content of the observed factor. Consider the case of the S&P 500 Financials, which captures the dynamics of those companies included in the S&P 500 that are classified as members of the GICS financials sector. I notice that the loadings on the global (unobserved) factor are consistently higher for all groups with respect to the case in which the financial factor is the S&P 500. This is because the latter index also include co-movements which are driven by non-financial stocks. When the observed factors capture a wide variety of shocks (the S&P 500 for American stocks, and the MSCI World for global stocks), the loadings on the global factor are consistently lower with respect to the instances in which the observed factors strictly relate to movements in financial stocks (S&P 500 Financials and MSCI World Financials). On the other hand, the loadings on the regional factor do not change significantly for varying model specifications, indicating that they represent uncorrelated region-specific risk drivers.

Considering all models, the one featuring the S&P 500 Financials as financial factor gives stronger results in terms of relative importance of the factors. The global factor loadings are consistently higher in this case with respect to the alternative models, at the expenses of the financial factor loadings which remains the key driver in regions such as North America and Western Europe. Excluding the regional factor, the global factor is the predominant source of systematic risk in regions such as MEA, Emerging Europe and Asia-Pacific - in line with the economic diversity of the companies operating in these domains.

[Table 1.4 about here.]

Table 1.4 reports the average R^2 of the model fit. I expect the model with three unobserved factor to provide the best fit, since I extract the factors via PCA which by definition maximizes the share of total variance explained by the factors. The results show that the models featuring the observed financial factor perform even better than the full unobserved model in some cases. When it under-performs, the difference is not substantial. This is true especially for the North America, and Emerging Europe regions in which the models featuring the S&P indexes show the best fit. In Latin America, the models with a higher degree of explanatory power are the ones with the MSCI indexes. For the Asia-Pacific and Western Europe regions, together with most of the sectors, the full unobserved model still provides the best fit.

The identification of the financial factor with the S&P 500 Financials index is motivated by the greater information content in the estimation of the global unobserved factor, supported by the explanatory power of the model which remains strong, if not stronger, with respect to the full unobserved framework. Upon identification of the financial factor, the unobserved factors (global and regional) are estimated from the residual stock returns after orthogonalising against the S&P 500 Financials index. This ensures orthogonality between the observed and unobserved global and regional factors.

Global and Regional Factors

Figure 1.3 plots the estimated global factor and regional factors for Asia-Pacific, Emerging Europe, Latin America, Middle-East and Africa, North America, and Western Europe. The factors are extracted via PCA from model (1.10) and are rotated to ensure that they are positively correlated with the stock market index of the biggest country in the region.

[Figure 1.3 about here.]

Table 1.5 reports the correlation between my set of factors and exogenous variables, such as the S&P 500 and its financials-only counterpart.

[Table 1.6 about here.]

In panel 1.5b I report the correlation among the six regional factors. As anticipated, the sparsity assumption on the loadings matrix $\mathbf{\Lambda}$ allows for cross-sectional dependence between the regional factors, although they do not interact in the model (since each stock belongs to a specific geographical region). The factors are on average only mildly correlated, with the highest correlation being the one between Western Europe and North America, 0.23. The second highest is the one between Latin America and Emerging Europe, 0.13. As Borghi et al. (2018) report, since the former region includes Brazil and the latter Russia, the connection may be due to the presence of large oil companies in the stock market indexes of these countries. I report near-zero (if not negative) correlations for the MEA factor with the other set of factors, Western Europe and North America excluded. The correlation between MEA and North America is negative, -0.15, with a similar coefficient but with opposite sign for the Western Europe region, 0.14.

In panel 1.5a I report the results of a further robustness check in which I calculate the correlation coefficients between the first principal component of the six regional portfolios and the global PC, together with two exogenous variables. As expected, the first PC extracted from the stocks in the North American region correlates almost perfectly with the S&P 500 index with a coefficient of 0.95, and to a lesser extent with the financials-only counterpart, with a coefficient of 0.85. The firms in the Western European region also closely mimic the pattern of the US stock markets, with correlations up to 0.83 with the S&P 500 index. All factors, except the MEA, are highly correlated with the global factor. Given the variety in the relationships between the PCs estimated from the regional stock portfolios and the global factor, this pattern further corroborates the distinction of global and regional shocks in my model.

1.3.3 Time-Varying Factor Sensitivities

In this section, I compare the results on the estimation of time-varying factor betas via rolling OLS regressions and ML estimation using the Kalman filter. Throughout this chapter, I refer to the first approach as RW, and to the latter as KF.

Model Fit and Misspecification Tests

In this section I compare the residuals obtained from a multi-level factor model with static loading, equation (1.9), with those obtained from the model with time-varying loadings, equations (1.14)-(1.15) and (1.11). Conditional on the estimated factor space, the time-varying betas are estimated via MLE, and rolling OLS regressions. For the latter, I distinguish between a 'long' estimation window made of the most recent five-year observations, as in Armstrong, Banerjee, and Corona (2013), and a 'short' window of half-year observations, following Bekaert, Hodrick, and Zhang (2009). To justify the use of time-varying loadings and compare the MLE and rolling OLS estimation procedures, I extract the firm-individual series of beta estimates for the $K = 3$ factors and compute the systematic (model-implied) component of returns for the three different approaches (ML, and rolling least squares with two different estimation windows).

The factors are estimated via PCA on the full sample, while factor sensitivities are estimated on a real time basis. This procedures ensures that the rolling estimates obtained from different windows are comparable, considering the same set of factors. Most importantly, this setup allows me to disentangle the effects of the instability in the factor estimation procedure and in the loadings estimation. To assess the benefits of fitting a dynamic-loadings representation with respect to a static-loadings equivalent, I also compute the common component for the model with loadings estimated via OLS on the full-sample and perform a battery of misspecification tests to compare against.

[Table 1.7 about here.]

In panel 1.6a, I report the statistics on the goodness of fit for model (1.9) with loadings estimated via OLS on the full-sample, versus its equivalent dynamic-loadings representation (with $\mathbf{\Lambda}_t$ instead of $\mathbf{\Lambda}$) when I use a long estimation window for the RW approach, a short window, and the KF. In panel 1.6b I report the average improvement of using time-varying betas with respect to the static OLS case. For the rolling OLS estimator, I consider T_W-averages of the time-series of R^2 from the window-specific regressions. I then average the stock-specific mean R^2s across regions or sectors, and report the results in panel 1.6a. For the static approach I am given a single figure for the whole sample and in these cases I report the N-average statistics for the stocks belonging to a particular group. The R^2 is calculated as the squared correlation between the model-implied $(T \times N)$ common component and the matrix of returns. Conditional on the estimated factors, the common component is constructed as the linear combination of the factors and the vector (static case), or matrix (time-varying case) of beta estimates.

Across all groups considered, I find that the goodness of fit for the models featuring rolling betas vary with the window size. With a long window, the figures show a consistently lower explanatory power of the dynamic model representation with respect to the static benchmark. The discrepancy is especially relevant for the Emerging Europe region and Energy sector for which I record a -4% in average explanatory power. With a short window made of the most recent half-year observations, the figures show a consistently higher explanatory power with respect to the static benchmark model, although substantially lower than the KF approach. On average, I find that the latter has a 19% higher goodness of fit with respect to the static case. The rolling window scheme with a long window yields a 6% average increase, whilst the short window counterpart results in a 3% loss in average explanatory power, across all groups. The biggest improvement in explanatory power is found for the Asia-Pacific region, this is true for the KF, and RW approach with a short window. The stocks belonging to Asia-Pacific are primarily driven by global and region-specific shocks, rather than financial shocks. This indicates that the two methodologies are able to better capture the share of co-movements implied by the unobserved factors with respect to the benchmark. On the other hand, the groups where the static OLS estimator gives the best results correspond to the regions which are more integrated (Western Europe and North America). This is in line with the average pair-wise Pearson correlation reported in panel 1.2a.

Overall, the explanatory power of the model featuring time-varying betas estimated via the Kalman filter is substantially higher with respect to alternative specifications, and the baseline model. Comparing the rolling window approaches, I find that using a shorter window for estimation results in a substantial deterioration in the explained portion of systematic variation of stock returns, with respect to the a long window of five years.

When I study the explanatory power of the models featuring rolling least squares estimation, I take T_W-averages of the window-specific R^2s. In figure 1.4 I report the time-series of R^2 from the rolling regressions, and I compare it against the R^2 calculated using the common component from the static model, and from the dynamic model with time-varying betas estimated via ML.

[Figure 1.4 about here.]

Considering a five-year estimation window, I find that the benefits of using a static model representation are lower with respect to rolling scheme featuring time-varying betas, and this is true especially in the first three or five years in the sample (up to 2014 or 2016). From 2014 onwards, I notice a sharp decay in the average explanatory power for most of the groups considered, which makes the dynamic representation inferior to the static case. In the North America and European regions, from mid 2016 onwards, there are substantially lower benefits of fitting a dynamic model with a long estimation window as opposed to a static model. For the Asia-Pacific and Latin-America regions this behavior is evident from the end of 2015, and for the MEA region from the beginning of 2014.

This is also true, although to a lesser extent, for a short estimation window. Across all groups, I notice a sharp increase in the average goodness of fit of the RW estimator during the periods early 2009 to the end-of 2011, before rising again from mid-2014 up to the beginning of 2016. In period 2016-2018, average goodness of fit decreases substantially across all stocks, before increasing from the end of 2018 up to May 2019. For some regions, the average explanatory power of the RW approach is even higher than the share of total variation captured with the KF approach during the period 2011-2012. This is true especially for the American regions and Western Europe. The years 2009-2011 can be associated mainly to the global financial crisis, and from 2011 to 2013 fiscal policy in the Eurozone turned progressively more restrictive in response to the sovereign-debt crisis.

To justify the use of time-varying factor loadings from a statistical perspective, I perform a suite of misspecification tests based on the residual component of stock returns. Firstly, I implement a White-type test under the null of homoscedasticity and estimate the following auxiliary regression

$$\hat{e}_{i,t} = \alpha_i + \hat{F}^2_{i,t}\Gamma_i^\top + u_{i,t} \qquad u_{i,t} \sim N(0, \sigma^2_{u,i})$$

for $i = 1, ..., N$ stocks, where $\hat{F}_{i,t}$ contains the three factors specific to stock i, estimated by PCA, and $\Gamma_i = (\gamma_{i,1}, ..., \gamma_{i,K})^\top$ is the K-dimensional vector of factor sensitivities for stock i. The test statistic is equal to the R^2 times the sample size T, and is distributed as a χ^2 with degrees of freedom equal to the number of factors in $\hat{F}_{i,t}$. In this setting, Mikkelsen (2017) proves that testing for constant variance is equivalent to testing for constant loadings. I refer to his work for details of the properties of the test.

Secondly, I conduct two tests on the null hypothesis of no serial correlation of the error term $\hat{e_{i,t}}$ using the Breusch-Godfrey test up to two and five lags. The test statistic is equal to R^2 times

T and is distributed as a χ^2 with $T - p$ degrees of freedom, where p is the maximum lag up to which the test is conducted (two and five in my case).

[Table 1.8 about here.]

Table 1.7 reports the results. For the heteroscedasticity and serial correlation tests, in table 1.7 I report the percentage of stocks for which I reject the null at 95% confidence. I find that for roughly half of the 1692 equities considered the static-loadings model fail to capture the volatility pattern of returns. When I allow for time-varying factor loadings, the number of stocks showing evidence of residual heteroscedasticity decreases. The reduction is substantial for the rolling OLS estimator with a long window, in which only 270 equities show evidence of (residual) volatility clustering, and moderate for the remaining two estimation criteria. For the KF, this percentage of stocks decreases by 16%, from 49% to 33%, and similarly for the rolling OLS approach with a short window from 49% to 36%.

The results on the serial correlation tests provide an indication on the magnitude of the induced auto-correlation in the beta estimates, which are estimated on overlapping windows in the RW approach. Considering a short window, the autocorrelation in the overlapping beta estimates is passed over to the common component of stock results, which then invalidates standard residual-based misspecification tests. I find that for the static-loadings model the vast majority of stocks show evidence of residual autocorrelation, 67% for lags 1-2 and 85% for lags 1-5. Across all model specifications, the RW model with a short window shows the lowest decrease in the percentage of stocks showing evidence of residual serial correlation. Surprisingly, I find that the figure based on the test considering lags up to 5 increase for the latter model, up to 87%. This is a further indication that the structural instability in the beta estimates estimated using a short window has a considerable impact on standard misspecifications tests. I find that when I estimate the beta dynamics via the KF, the share of equities showing evidence of residual autocorrelation is similar to the case in which I employ a long rolling window.

Overall, I find that the model with time-varying loadings has an unambiguously higher performance with respect to the static-loadings specification. When I estimate the time-varying betas via rolling OLS regressions, the results are highly influenced by the choice of the window length. Using fewer observations for estimation, e.g. half-year windows, increases the instability in the beta estimates, however the assumption of constant beta may be more appropriate. When I consider five years of observations for the estimation window, much of the short-term variability in the beta estimates is averaged out, and the assumption of constant beta underpinning the rolling OLS estimator may be difficult to defend. Standard residual-based misspecificaton tests suggest that the trade off between estimation accuracy and beta instability is mitigated when I consider a long estimation window. However, the improvement in the average explanatory power with respect to a static-loadings model is negative, implying that the dynamic model specifications is inferior to the benchmark. Choosing between the two methods for the estimation of time-varying betas via rolling least squares, I find that a longer window provides the best results from a statistical perspective. Altogether, the ML estimator of the time-varying factor sensitivities remains undoubtedly the best-performing from a statistical viewpoint.

Static and Dynamic Variance Decomposition

In this section I use variance decomposition methods to study the model-implied co-movements in the panel of stock returns. This approach is employed extensively in the literature to analyse the results of different model specifications, as in the static model of Breitung and Eickmeier (2015), or the dynamic model of Bekaert, Hodrick, and Zhang (2009). The intuition is that the higher the (average) share of variance explained by the common factors, compared to the idiosyncratic variance, the higher the level of co-movements. For static factor models, this method provides one number for the whole dataset, however for dynamic specifications, the conditional variance decomposition can be performed at each time t. In what follows, I compare the static and dynamic variance decompositions, and for the latter I analyse the differences for each estimation method.

Table 1.8 reports the static variance decomposition calculated on the full sample via OLS. In line with the preliminary correlation analysis of table 1.2, I find that the firms listed in North America and in Western Europe have the highest level of commonality. On the other hand, the idiosyncratic component dominates the returns on the equities listed in MEA. For these stocks, the regional factor appears to describe a much larger portion of total variance compared with stocks in other regions or sectors.

[Table 1.9 about here.]

With time-varying factor sensitivities, I can calculate the share of variance explained by each factor at each point in time, which allows to capture shifts in the relative importance of the factors and map them to macro events. Conditional on the estimated factor loadings, the variance of stock returns can be decomposed into a systematic and idiosyncratic component

$$Var(\boldsymbol{X})_t = \hat{\boldsymbol{\Lambda}}_t Var(\boldsymbol{F}) \hat{\boldsymbol{\Lambda}}_t^\top + Var(\boldsymbol{e}). \tag{1.18}$$

If the factor model is true, the common factors should explain as much as possible of the variation in $\boldsymbol{X}$ and the residual variance component should be zero. In small sample this may not necessarily be the case even if the factor model is correctly specified. Under a strong form of APT however, the residual variance should tend to zero asymptotically, see Chamberlain (1983). The variance of the returns on stock i at time t, conditional on the estimated loadings, can be expressed as

$$Var(X_i|\hat{\boldsymbol{\Lambda}})_t = (\hat{\lambda}_{i,t}^{fin})^2 \; Var(F_t^{fin}) + (\hat{\lambda}_{i,t}^{glob})^2 \; Var(F_t^{glob}) + (\hat{\lambda}_{i,t}^{reg})^2 \; Var(F_{r,t}^{reg}) \; \mathbf{1}_{\{i \in r\}} \tag{1.19}$$

for $i = 1, ..., N$ and $t = 1, ..., T$, assuming that the factors and the errors are orthogonal.

The factors are normalised to have unconditional variance equal to one. This is done over the full sample when I estimate the loadings via MLE, and for each window when I use rolling OLS regressions. The share of total variation explained by the factors is defined as

$$\begin{aligned} FV_{i,t} &= \frac{(\hat{\lambda}_{i,t}^{fin})^2}{Var(X_i|\hat{\mathbf{\Lambda}})_t} \qquad \text{Financial} \\ GV_{i,t} &= \frac{(\hat{\lambda}_{i,t}^{glob})^2}{Var(X_i|\hat{\mathbf{\Lambda}})_t} \qquad \text{Global} \\ RV_{i,t} &= \frac{(\hat{\lambda}_{i,t}^{reg})^2}{Var(X_i|\hat{\mathbf{\Lambda}})_t} \qquad \text{Regional} \end{aligned}$$

where $FV_{i,t}$ denotes the share of variance explained by the financial factor at time t, and similarly $GV_{i,t}$ and $RV_{i,t}$ for the global and regional factors respectively.

Figure 1.5 reports the share of variance explained by each factor during the sample period, averaged across all N stocks. I repeat the exercise considering three different estimation methods for the time-varying betas, panel 1.5a reports the results with time-varying betas estimated via MLE, panel 1.5c shows the results of rolling OLS regressions with a window of half-year, and panel 1.5b reports the analogous considering a window of five years.

[Figure 1.5 about here.]

Figure 1.5 reports a consistent increase in the exposure of stocks to financial shocks during the GFC. The estimated exposures are substantially greater in panel 1.5c with respect to the KF in panel 1.5a. With betas estimated via the KF, I find that at the outset of the crisis in the late 2009 the financial factor reaches an explained variation of more than 16%, whilst it is about 27% in the RW case. For the period from mid-2011 to end-of 2013, which agree with the dates normally considered for the ESD crisis, I find substantial similarities in the relative importance of the factors implied by panel 1.5a and 1.5c. Considering half-year rolling estimates, I find that the share of variance explained by the financial factor spikes to more than 25% at the beginning of 2012, before declining rapidly. I can reconcile the spike in the financial factor sensitivities with the EU-IMF bailout of Portugal in May 2011, and the deterioration of the credit worthiness of Greece. The Greek GDP had its worst decline in the sample period in 2011 with -6.9%. During the course of 2012, due to the successful fiscal consolidation and implementation of structural reforms in the countries at the highest risk of defaulting, financial stability in the wide Eurozone improved and contagion risk diminished. This may help explain the reduction in the financial factor sensitivities towards the end of 2012 in panel 1.5c, with respect to pre-2010 levels. When I analyse the results of the five-year rolling estimates, panel 1.5b, I find an increased volatility in the financial factor sensitivities for all stocks in mid-2012.

In contrast, the results in panel 1.5c point to a reduction in volatility for that period.

I report an increase in the financial factor sensitivities at the beginning of 2018 up to the end of the sample for the five-year estimates, however it is modest for the stocks belonging to Western Europe, and is difficult to reconcile with macro relevant events. In contrast, when I restrict the sample from the beginning of 2016 to the end of 2017 in panel 1.5a, I find evidence of increased co-movements in the period that runs from the week before the Brexit referendum to the week after. As expected, the results are stronger if I consider equities specific to the Western Europe region, rather than those in other regions. The spike in explained variance at the outset of the referendum for the reference region is of roughly 8%, compared to a modest 2% recorded on aggregate. The only region which shows increased level of co-movements is North America, even though the increment is negligible. I also find a substantial increase in the global factor sensitivities in panel 1.5c for that period, which is particularly relevant for stocks belonging to the Western Europe region, concordant with the Brexit referendum.

By partitioning the universe of securities into groups (countries, regions or sectors), I can evaluate the share of variance explained by each factor, for each group individually. This is done by taking cross-sectional averages of the quantities defined above for the equities within a particular group. Figure 1.6 plots the share of variance explained by the free factors for each region.

[Figure 1.6 about here.]

The increase in stock return comovements at the outset of the GFC varies across regions. In Western Europe and Asia-Pacific I find that some stocks start to become more sensitive to global shocks already at the end of 2007, see the panels 1.6g-1.6k and 1.6m-1.6q respectively. I also find evidence of increased co-movements during the ESD crisis at the end of 2011 in the Emerging Europe, which includes countries who were at higher risk of defaulting with respect to more stable economies in Western Europe. In Middle East Africa, I find a strong prevalence of the regional factor throughout the sample.

The results featuring time-varying betas estimated via five-year rolling OLS regressions are difficult to interpret. I find a spike in the explained variance share of the financial factor at the end of 2013 in Western Europe and Emerging Europe (the regions affected the most by the ESDC), however the share of explained variance by the fianancial factor is steadily decreasing from the beginning of the sample up to 2018. This is due to the extended time frame, which necessarily includes too distant and possibly irrelevant information for the estimation of the current-period betas. For instance, the estimation of the first (time-varying) beta at the beginning of 2011 also include data relevant to the GFC in 2008 and 2009, a period in which the financial factor explained a substantial portion of co-variations in stock returns (more than any other factor, in any other time period). I record a significant increase in the share of variance of the financial factor in mid 2016, for the Western Europe and North America region, before declining rapidly at the end of the year. This can be potentially be reconciled with relevant macro events such as the Brexit referendum or the American elections in the second half of the year, however the increase is modest and also relevant for regions other than Western Europe and North America which makes the mapping challenging.

[Figure 1.7 about here.]

Figure 1.7 plots the share of variance explained by the three factors in each sector. The effects of the oil market shocks at the beginning of 2015 on the equities belonging to the Energy, Basic Materials and Utilities sectors are captured by the model featuring time-varying loadings estimated via ML. The share of explained variance increases considerably for the stocks in the relevant sectors, this is true especially for the global factor. Given the market-wide nature of these shocks, which are not strictly related to the US financial sector, I expect an increase in explained variance of the global factor, which is supported by the data. When I consider the five-year rolling OLS estimates, I find scattered results across comparable sectors.

Model Estimates

In this section I report the aggregating results on the estimation of time-varying loadings via ML and rolling OLS regressions. Table 1.9 reports the average magnitude of the factor loadings, their persistence (AR(1) parameter) and their volatilities, aggregated by either region or sector. The loading magnitude is estimated via OLS from the static loading model, and as the sample mean of the beta estimates obtained from the rolling OLS regressions. The volatility of the loadings is equivalently estimated as the sample standard deviation of the time-varying sensitivities, which are estimated either via MLE, or rolling OLS regressions. The AR(1) parameter is estimated via MLE from the dynamic factor model. I also report the number of stocks with loadings that vary so little that they are indistinguishable from the OLS estimates on the full sample. The volatility threshold is set to 0.01, under which it is very difficult to identify the autoregressive parameter. In all these cases I treat the loadings as static. Finally, I report the number of stocks with large AR(1) coefficients (larger than 0.5), which indicate a high persistence in the loadings dynamics, hence high predictability.

[Table 1.10 about here.]

I firstly analyse the results by regions. I find that regardless of the estimation method, the stocks in North America are the ones with highest exposure to the financial factor, followed by the equities in Western Europe. On average, almost half of the total return variation for these stocks is explained by the observed S&P 500 Financials factor. As anticipated, the rolling OLS estimator appears to be centered around a long-run mean which coincides with the OLS estimate on full sample. However, when I look at the variability of the estimated betas I find substantially different results. For the five-year estimates, the sample standard deviation is on average one order of magnitude lower with respect to the Kalman filter and half-year window. Among the two, the rolling window estimator consistently yields higher variability. This may result in the financial factor loadings failing to be identified for some stocks in regions such as Asia-Pacific, and Emerging Europe. For the former region, the standard deviation of the half-year rolling betas reaches 0.215, whilst it is 0.181 for the KF, and 0.058 for the five-year rolling estimates. In Emerging Europe, the standard deviation of the KF estimates is 0.208, and 0.219 for the half-year estimates. On average, the standard deviation of the rolling estimator of the financial factor betas with a short window is 0.21, as opposed to 0.17 for the KF.

When I analyse the results for the betas on the unobserved factors, I find that the KF estimator still yields a higher degree of accuracy (i.e. less variability) with respect to the rolling estimator. In fact, across all regions, the variability of the half-year rolling least squares estimator of the global and regional factor loadings is 0.19. For comparison, the variability of the KF estimates is 0.12 and 0.14 respectively. With respect to loadings' persistence, I find that the stocks in the North America region are the ones with highest AR(1) parameter of the financial factor loadings, 0.227, followed by those in Latin America, 0.221, and finally Western Europe, 0.183. The number of stocks with high AR(1) parameter is greater in Asia Pacific with a total of 202 stocks over 484, followed by the equities in Western Europe with 177 over 518. For the global factor, stocks in Western Europe show on average the largest factor exposures with 0.367, followed by the equities in Emerging Europe. For the North America region I find that on average the AR(1) parameter on the global factor is negative and close to zero, implying a low degree of predictability of the global factor betas for the stocks in this region. Moreover, across all regions the stocks in North America have lowest sensitivity with respect to changes in the global factor. This is in line with the definition of global factor, which captures (global) sources of systematic risk uncorrelated with the US financial sectors. I record a significant level of persistence in the global factor betas for the stocks belonging to Eastern Europe, with an average AR(1) parameter of 0.227, and for those in the MEA region, which show a negative coefficient of -0.118. On average, I find that global shocks are characterised by a modest level of persistence. The regions with the highest number of stocks with autoregressive parameter on the global factor beta greater than 0.5 are still Asia-Pacific and Western Europe, with comparable numbers with respect to the financial factor. When I consider the regional factor sensitivities, stocks in the MEA region are the ones with highest loadings on average, followed by the equities in the Latin America and Western Europe regions. When I consider the AR(1) parameter, I find that the regional factor loadings are the most persistent, implying higher predictability of regional systematic risk compared to financial and global risk, this is particularly relevant for the Latin and North America regions.

Secondly, I analyse the results by sectors. As expected, when I estimate the loading magnitude via OLS on the full sample, financial firms are the most exposed to the financial factor followed by the equities in the Energy sector. For the other sectors, I do not find substantial variation in the firms' exposures to the financial factor, on average I report a loading of 0.24. Financial firms instead record a loading of 0.357 for the financial factor, similar to the Energy and Communications sectors. Basic Materials equities have the highest exposures to the estimated global factor, and similarly Industrial equities to the regional factor.

1.4 Conclusion and Further Research

In this chapter, I reviewed and compared two methodologies for the estimation of time-varying loadings in linear asset Pricing models, MLE via the Kalman filter and rolling OLS estimation. I leveraged the modelling setup of Borghi et al. (2018) and distinguished between a financial, global and region-specific risk drivers to estimate the common components in the stock returns. The financial factor is observed, and I identify it with the S&P 500 Financial index after careful consideration of alternative stock market indexes. The global and regional factors are latent and estimated via PCA. Conditional on the ex-post factors, I compare the in-sample performance of models with time-varying sensitivities estimated using the two methodologies, from a statistical and economical perspective. I document my findings for a large panel of 1692 international stocks listed in 40 countries with weekly return observations in the period January 6th 2006 to May 31st 2019.

My analysis yields two main findings. Firstly, I propose a different identification procedure for the observed financial factor, and let the data suggest which index is best suited to describe the systematic variation in stock returns due to movements in the (global) financial sector. I find that the S&P500 Financials yields the best results from a statistical, and economic standpoint. Upon estimation of the latent factors via PCA, I corroborate the evidence that the estimated factor space is very similar to the one spanned by the FF factors, which indicates that the statistical factor model of Borghi et al. (2018) captures to a large degree the risks to which firms are exposed in a more traditional factor model constructed from firm attributes.

Secondly, I find that estimating time-varying factor sensitivities via rolling OLS regressions requires careful consideration of the window length and frequency of the data. The rolling OLS estimator is the status quo in the literature to estimate time-varying loadings, however I find that the ML estimator provides considerably better results. From a statistical perspective, I find that when loadings are estimated via rolling OLS regressions with a short window size, the benefits of fitting a dynamic factor model become slim and a comparable static-loadings model performs better than the dynamic specification. This is because much of the instability in the beta estimates is passed over to the common component of stocks returns, which results in a misspecified model when I perform standard residual-based misspecification tests. On the other hand when I consider a longer window, I find that the dynamic model is correctly specified, although it is not able to capture a considerable portion of stock returns variation. When I analyse the relative importance of the factors via time-varying variance decomposition techniques, I find that the rolling OLS estimator shows a higher portion of estimated total variation in stock returns with respect to the MLE case. This is spuriously due to the higher degree of volatility in the beta estimates that results from the rolling scheme. When I estimate the loadings via ML, I corroborate the evidence that the relative importance of the factors is time-varying: when unexpected events happen globally stock return co-movements increase. In these cases I find that stocks become marginally more exposed to financial and global shocks, and I are able to map the relative changes to pertinent macro events. When I consider the rolling OLS estimates constructed on a long window, I am not able to study the evolution in the relative importance of the factors over time, despite the model being correctly specified. On the other hand, when I consider the least squares estimates constructed from a short rolling window, I find a greater degree of short-term variability which ease the economic interpretation of stock returns co-movements, despite the model suffering from severe misspecification issues.

1.4.1 Further Research

The analysis in this chapter provides a first overview of the role of time-varying factor sensitivities in explaining contemporaneous return patterns in international asset pricing, a research question that I elaborate further in Chapter 2. In particular, I depart from the contemporaneous-equation framework of Borghi et al. (2018) and forecast the $t+1$ return based on the estimated loadings and factors up to time t. Additionally, I evaluate different model specifications, such as FF3 and FF5, and compare their out-of-sample performance with the factor model in Borghi et al. (2018).

Studying how time-varying estimates via rolling OLS can improve the factor model's explanatory power (from stock-specific time-series regressions) is a further research question which I begin to analyse in Appendix A.1, using the framework of Inoue, Jin, and Rossi (2017). The theory developed in their work provide the conditions for the asymptotic optimality of their MSFE criterion, so that the error introduced by estimating time-varying betas using the most recent observations is negligible. When I try to adapt their methodology to consider a contemporaneous rather than predictive framework, I am not able to prove the unbiasedness of the rolling OLS estimator due to the properties of conditional expectations. Note that also Lettau and Pelger (2020b) use out-of-sample forecasting to test the predictive performance of their RP-PCA estimator of the unknown factors. They estimate factors the loadings based on a rolling window of 20 years of monthly data ($T = 240$), and subsequently predict the $t+1$ return and obtain the out-of-sample pricing error. The contributions in the literature that employ rolling OLS in a contemporaneous-equation setting are many, e.g. Bekaert, Hodrick, and Zhang (2009), Bekaert et al. (2014), Fama and French (2012), or Fama and French (2017), and thus far the theory that extend the validity of the optimal window selection criteria in Inoue, Jin, and Rossi (2017), or Pesaran and Timmermann (2007) to a contemporaneous-equation framework is scarce, which leaves this research question open to further investigation.

Additionally, in Appendix A.2 I analyse the relationship between magnitude, variance, and persistence of the factor loadings estimated via the KF with expected returns. I replicate the analysis of Borghi et al. (2018) for an extended time span and I fail to confirm their evidence of a premium for holding stocks with highly volatile global factor exposure in an 'out-of-sample' context. My results point to an irrelevance of the beta parameters in the cross-section of stock returns, in contrast to both Borghi et al. (2018), and Armstrong, Banerjee, and Corona (2013) who claim that firm-specific uncertainty of factor sensitivities negatively affects expected returns. A further extension to the analysis in this chapter is to study the relevance of firm-specific beta characteristics in constructing equity portfolios for asset allocation purposes, employing a consistent out-of-sample methodology. In fact, although my results are in contrast to the existing literature, there may exists sample-specific considerations (between 2016 and 2019, the difference in sample between my work and Borghi et al. (2018)) that can affect the results. I leave this research question open for future studies.

Further research on this chapter include examining the properties of the rolling least squares estimator of time-varying betas at higher frequencies. As I anticipated, if the estimated betas are believed to vary slowly with respect to the sampling frequency, then OLS on the most recent data is the appropriate modelling choice. This may not be the case when I consider monthly holding-period returns, but it may well be suited for daily data. From a methodological perspective, a further extension of this study is to change the grouping structure of the universe. In this chapter I take it as given following the geographical partition in Bekaert, Hodrick, and Zhang (2009) and Bekaert et al. (2014), however the factor extraction procedure can be adapted to consider group-specific factors such as industry drivers, regional drivers based on alternative geographical partitions, and other drivers which affect a specific portion of the whole cross-section of stocks ('local' factors). The geographical classification of the equities in my universe is based on the country in which their shares are listed, however it can be argued that a more appropriate criteria to classify stocks is based on the country in which the majority of their business operations are conducted.

TABLE 1.1: **Universe of Securities**

The table reports the countries and the respective national stock indexes that are considered for each region. *#Stocks* is the number of equities that entered the index in the period from January 6th 2006 to May 31st 2019, *#Selected* is the number of stocks with no more than eight consecutive missing observations and at least two years of data, *Avg Active* is the average number of index members at the beginning of every month in the sample period.

Ticker	Country	Region	#Stocks	#Selected	Avg Active
SPTSX60	Canada	North America	106	92	60
OEX	US	North America	179	173	101
MEXBOL	Mexico	North America	76	69	35
MERVAL	Argentina	Latin America	87	81	17
IBOV	Brazil	Latin America	139	123	65
IPSA	Chile	Latin America	86	80	37
SPBLPGPT	Peru	Latin America	90	61	34
TPXL70	Japan	Asia-Pacific	128	126	70
SSE50	China	Asia-Pacific	158	150	50
HSCEI	HongKong	Asia-Pacific	93	89	41
SENSEX	India	Asia-Pacific	85	71	30
LQ45	Indonesia	Asia-Pacific	122	116	45
KOSPI50	Korea	Asia-Pacific	87	63	50
SET50	Thailand	Asia-Pacific	107	100	50
NZSE50FG	NewZealand	Asia-Pacific	96	89	50
AS31	Australia	Asia-Pacific	94	83	50
ATX	Austria	Western Europe	41	39	20
BEL20	Belgium	Western Europe	38	37	20
KFX	Denmark	Western Europe	34	33	20
HEX25	Finland	Western Europe	35	34	25
CAC	France	Western Europe	68	66	40
DAX	Germany	Western Europe	49	45	30
ISEQ	Ireland	Western Europe	94	73	51
AEX	Netherlands	Western Europe	61	49	25
OBX	Norway	Western Europe	67	59	25
PSI20	Portugal	Western Europe	38	37	19
IBEX	Spain	Western Europe	62	58	35
OMX	Sweden	Western Europe	43	40	30
SMI	Switzerland	Western Europe	55	53	21
UKX	UK	Western Europe	208	191	101
CRO	Croatia	Emerging Europe	94	89	23
CCTX	CzechRepublic	Emerging Europe	15	14	9
TALSE	Estonia	Emerging Europe	23	22	16
BUX	Hungary	Emerging Europe	33	28	14
RIGSE	Latvia	Emerging Europe	73	45	28
MALTEX	Malta	Emerging Europe	33	19	18
VILSE	Lithuania	Emerging Europe	47	38	27
WIG20	Poland	Emerging Europe	45	44	20
ROTXEUR	Romania	Emerging Europe	25	23	11
CRTX	Russia	Emerging Europe	49	9	13
BELEX15	Serbia	Emerging Europe	26	12	13
XU030	Turkey	Emerging Europe	78	64	30
PFTS	Ukraine	Emerging Europe	45	8	17
MOSEMDX	Morocco	MEA	82	78	48
Total			3294	2873	1534

TABLE 1.2: **Summary Statistics**

The table reports the summary statistics for the 1692 companies that have been part of the 40 national stock indexes in the sample period, January 2006 to May 2019. Panel 1.2a reports cross-sectional averages of the summary statistics for the weekly log-returns. Panel 1.2b reports average market capitalisation, total assets and debt. *Min* and *Max* refer to the absolute minimum and maximum, considering all stocks within a specific group. The remaining statistics are *N*-averages of the relevant measures: *Mean*, *Med* are cross-sectional averages of mean and median, *Std*, *Skew* and *Kurt* are the average standard deviation, skewness and kurtosi, $\rho(1)$ is the OLS estimate of the first autocorrelation coefficient, and *ADF* is the Augmented Dickey-Fuller test statistics, which is run with a constant, time trend and one lag. The critical value at 95% significance is -3.41, with the null hypothesis being the presence of a unit root. Lastly, *Pearson* is the average pair-wise Pearson correlation of the stocks in the relevant group. Ex-dividend data.

(A) *Stock Returns*

	Mean	Med	Min	Max	Std	Skew	Kurt	$\rho(1)$	ADF	Pearson	N
North America	0.045	0.083	-34.779	25.051	4.758	-0.695	15.31	-0.049	-18.798	0.353	232
Latin America	0.024	-0.023	-38.657	31.297	5.896	-0.258	11.136	-0.022	-18.208	0.274	218
Asia-Pacific	0.092	0.021	-32.67	27.957	5.425	-0.238	9.773	-0.024	-18.605	0.254	484
Western Europe	-0.017	0.059	-35.167	25.921	5.232	-0.715	12.189	-0.054	-18.94	0.406	518
Emerging Europe	-0.102	-0.026	-46.169	41.612	6.57	-0.389	18.041	-0.012	-18.055	0.294	208
MEA	0.053	-0.001	-19.9	21.471	4.074	0.142	7.9	-0.069	-19.749	0.184	32
Basic Materials	0.001	-0.01	-38.224	33.055	6.374	-0.205	9.764	-0.009	-18.194	0.268	176
Communications	-0.049	0.017	-36.051	32.294	5.51	-0.32	11.689	-0.047	-18.701	0.24	124
Energy	-0.052	0.01	-38.458	30.092	5.937	-0.522	10.455	-0.032	-18.839	0.351	113
Consumer, Cyclical	0.046	0.039	-35.349	30.479	5.628	-0.279	10.101	-0.028	-18.465	0.24	212
Financial	-0.001	0.023	-39.145	30.062	5.489	-0.695	17.126	-0.047	-18.657	0.294	338
Technology	0.05	0.08	-30.681	27.182	5.244	-0.305	9.846	-0.023	-18.736	0.234	67
Industrial	-0.003	0.027	-38.28	28.619	5.583	-0.559	13.967	-0.027	-18.473	0.271	263
Consumer, Non-cyclical	0.102	0.058	-30.852	24.089	4.672	-0.457	11.246	-0.048	-18.891	0.209	282
Utilities	0.026	0.031	-31.081	24.05	4.703	-0.507	10.53	-0.054	-19.078	0.24	108
Diversified	0.014	0.003	-35.098	39.071	5.822	0.016	11.966	0.004	-17.964	0.271	9

(B) *Balance Sheet*

	Market Cap ($B)	Tot Assets ($B)	Tot Debt ($B)
North America	50.933	110.676	28.591
Latin America	5.392	17.414	5.78
Asia-Pacific	9.465	38.578	9.665
Western Europe	18.938	106.493	30.504
Emerging Europe	1.687	7.269	1.629
MEA	1.439	3.963	0.779
Basic Materials	7.893	13.646	3.478
Communications	25.611	30.007	9.36
Energy	23.201	41.433	8.377
Consumer, Cyclical	11.528	18.958	6.538
Financial	17.016	229.052	61.98
Technology	40.871	22.197	4.702
Industrial	10.09	14.947	4.478
Consumer, Non-cyclical	20.931	15.82	4.222
Utilities	9.701	27.226	9.518
Diversified	5.355	22.738	5.946

$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 1.3: **Information Content of Estimated Factors**

The table reports cross-sectional averages of the OLS estimates on the full sample of the factor sensitivities for $K = 3$ factors, with $K_{glob} = 2$ and $K_{reg} = R$, one for each region. *Unobserved Factors* reports the beta estimates for the model featuring two unobserved global factors, both estimated via PCA. The remaining columns report the factor sensitivities for the model featuring one observed financial factor, which I identify with four different stock indexes, and two unobserved factors, the global and the region-specific driver. Factors are standardised to have zero mean and unit variance. I report only the identified betas.

	Unobserved Factors			MSCI World			MSCI World Financials			S&P500			S&P500 Financials		
	F_1^{glob}	F_2^{glob}	F^{reg}	F^{fin}	F^{glob}	F^{reg}	F^{fin}	F^{glob}	F^{reg}	F^{fin}	F^{glob}	F^{reg}	F^{fin}	F^{glob}	F^{reg}
North America	0.507		0.233	0.516	0.105	0.212	0.476	0.184	0.248	0.541	0.126		0.47	0.245	0.2
Latin America	0.407	0.076	0.246	0.339	0.271	0.205	0.321	0.288	0.218	0.318	0.28	0.241	0.274	0.322	0.246
Asia-Pacific	0.362	0.227	0.07	0.301	0.239	0.169	0.28	0.261	0.177	0.263	0.283	0.172	0.218	0.314	0.182
Western Europe	0.607	-0.107		0.542	0.193	0.234	0.517	0.245	0.241	0.499	0.281	0.238	0.436	0.367	0.244
Emerging Europe	0.394		0.18	0.314	0.266	0.096	0.304	0.268	0.131	0.272	0.314	0.107	0.236	0.329	0.14
MEA	0.207		0.43	0.131	0.186	0.42	0.124	0.182	0.423	0.084	0.225	0.412	0.065	0.219	0.419
Basic Materials	0.47	0.107	0.094	0.393	0.271	0.199	0.363	0.308	0.215	0.356	0.312	0.233	0.298	0.371	0.214
Communications	0.462		0.114	0.418	0.184	0.17	0.392	0.227	0.191	0.391	0.24	0.141	0.334	0.304	0.194
Energy	0.517		0.076	0.444	0.278	0.156	0.399	0.34	0.184	0.411	0.316	0.224	0.337	0.404	0.167
Consumer, Cyclical	0.445		0.147	0.392	0.198	0.195	0.373	0.228	0.209	0.365	0.248	0.171	0.318	0.301	0.21
Ficial	0.506		0.18	0.448	0.224	0.211	0.449	0.234	0.202	0.416	0.281	0.2	0.382	0.327	0.209
Technology	0.458		0.134	0.436	0.125	0.196	0.399	0.189	0.238	0.425	0.178	0.134	0.357	0.268	0.219
Industrial	0.489		0.132	0.432	0.213	0.212	0.408	0.251	0.225	0.402	0.267	0.192	0.35	0.33	0.219
Consumer, Non-cyclical	0.406		0.078	0.36	0.179	0.175	0.328	0.224	0.205	0.334	0.229	0.129	0.274	0.287	0.209
Utilities	0.397		0.114	0.338	0.219	0.22	0.308	0.255	0.244	0.307	0.259	0.213	0.251	0.306	0.261
Diversified	0.419	0.061	0.24	0.352	0.273	0.186	0.335	0.285	0.195	0.334	0.294	0.147	0.293	0.324	0.205

$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 1.4: **Identification of Financial Factor**

The table reports the cross-sectional average R^2s for the model with $K = 3$ factors, $K_{glob} = 2$ and $K_{reg} = R$, one for each region. *Unobserved Factors* reports the average goodness of fit for the model featuring two unobserved global factors, both estimated via PCA. The remaining columns report the R^2 for the model featuring one observed financial factor, which I identify with four different stock indexes, and two unobserved factors, the global and the region-specific driver. In bold I highlight the best model(s) for each region or sector.

	Unobserved Factors (%)	MSCI World (%)	MSCI World Financials (%)	S&P 500 (%)	S&P 500 Financials (%)
North America	36	36	37	**38**	**38**
Latin America	28	**30**	**30**	29	29
Asia-Pacific	**28**	26	26	26	26
Western Europe	**42**	41	41	41	41
Emerging Europe	26	**27**	26	**27**	**27**
MEA	25	25	25	25	25
Basic Materials	35	35	35	**36**	35
Communications	**30**	29	29	29	29
Energy	36	36	36	**39**	36
Consumer, Cyclical	31	31	31	31	31
Financial	**39**	38	38	38	**39**
Technology	31	31	31	31	30
Industrial	**36**	35	35	35	35
Consumer, Non-cyclical	**27**	26	26	26	26
Utilities	29	29	29	29	29
Diversified	**33**	**33**	32	32	32

$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 1.5: **Correlation Between Factors and Exogenous Variables**

Panel 1.5a reports the correlation between the regional factors (each estimated by the first PC of a portfolio of the relevant stocks) and a global factor, the S&P500 and S&P500 Financials indexes. Panel 1.5b reports the correlation matrix of the six estimated regional factors.

(A) *PC1*

	Glob PC	S&P 500	S&P 500 Fin
North America	0.923	0.964	0.846
Latin America	0.866	0.697	0.588
Asia-Pacific	0.882	0.68	0.568
Western Europe	0.971	0.834	0.738
Emerging Europe	0.88	0.665	0.597
MEA	0.412	0.211	0.163

(B) *Correlation Between Regional Factors*

	(1)	(2)	(3)	(4)	(5)	(6)
North America (1)	1					
Latin America (2)	0.077	1				
Asia-Pacific (3)	-0.064	0.077	1			
Western Europe (4)	0.231	-0.006	-0.117	1		
Emerging Europe (5)	0.091	0.125	-0.006	0.076	1	
MEA (6)	-0.149	0.005	0.001	0.137	-0.055	1

$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 1.6: **Goodness of Fit**

The table reports the statistics on the goodness of fit of the dynamic factor model with loadings estimated via MLE, and via rolling OLS regressions, compared with a static model where the loadings are estimated using OLS on the full sample. For the rolling OLS estimator I consider the T_W-averages of the time-series of R^2 from the OLS regressions as a measure of fit, I then average these in each group and report the results below. In panel 1.6a I report the average R^2 of the four model specifications, whilst in panel 1.6b I report the improvement in the explanatory power with respect to the static OLS case. For the rolling window scheme I consider a long window made of the most recent five-year observations (*RW-5*), and a short window of six months (*RW-H*).

(A) *Goodness of Fit*

	$R^2_{KF}(\%)$	$R^2_{RW-5}(\%)$	$R^2_{RW-H}(\%)$	$R^2_{OLS}(\%)$
North America	58.02	36.11	43.87	37.94
Latin America	43.46	27.18	35.3	29.22
Asia-Pacific	51.11	23.51	33	26.25
Western Europe	60.36	37.8	45.89	41.28
Emerging Europe	48.04	22.46	32.83	26.56
MEA	39.8	19.92	29.9	24.52
Basic Materials	54.21	31.18	40.35	35.23
Communications	48.9	26.26	35.25	29.18
Energy	53.6	31.76	41.66	36.29
Consumer, Cyclical	52.08	26.81	36.79	30.63
Financial	59.79	36.53	43.73	38.6
Technology	49.22	28.64	36.33	30.48
Industrial	53.7	31.25	40.45	35.05
Consumer, Non-cyclical	49.86	24.46	33.05	26.18
Utilities	49.41	26.83	36.26	29.44
Diversified	52.97	30.33	37.17	32.13

(B) *Improvement*

	$\Delta_{KF}(\%)$	$\Delta_{RW-5}(\%)$	$\Delta_{RW-H}(\%)$
North America	20.08	-1.83	5.93
Latin America	14.24	-2.04	6.08
Asia-Pacific	24.86	-2.74	6.75
Western Europe	19.08	-3.48	4.61
Emerging Europe	21.47	-4.11	6.26
MEA	15.28	-4.6	5.38
Basic Materials	18.98	-4.05	5.12
Communications	19.72	-2.92	6.07
Energy	17.31	-4.53	5.37
Consumer, Cyclical	21.45	-3.82	6.16
Financial	21.18	-2.07	5.13
Technology	18.74	-1.83	5.85
Industrial	18.65	-3.8	5.4
Consumer, Non-cyclical	23.69	-1.72	6.87
Utilities	19.97	-2.6	6.83
Diversified	20.84	-1.8	5.04

$N = 1692$

$T = 439$ (five-year, 7th Jan 2011 - 31st May 2019)

$T = 674$ (half-year 7th Jul 2006 - 31st May 2019)

TABLE 1.7: **Misspecification Tests**

The table reports the aggregated results of the misspecification tests, firstly using the standardised returns matrix, secondly the residual matrix from a static factor model (*OLS*), and thirdly the residuals from the dynamic model specifications. The latter include the residuals from the model featuring loadings estimated via MLE (*KF*), and via rolling OLS regressions (*RW*) with two different window sizes. The tests are: *White*, the percentage of stocks for which I reject the null at 95% confidence level using the White's test, *BG 1-2* and *BG 1-5*, the percentage of stocks that exhibit residual serial correlation, using the Breusch and Godfrey test, up to lag two and five respectively.

	White (%)	BG 1-2 (%)	BG 1-5 (%)
Returns			
Static loadings (*OLS*)	49	67	85
TV loadings (*KF*)	33	51	74
TV loadings (*RW* five-year)	16	48	72
TV loadings (*RW* half-year)	36	63	87

$N = 1692$

$T = 700$ (full-sample, 6th Jan 2006 - 31st May 2019)

$T = 440$ (five-year, 7th Jan 2011 - 31st May 2019)

$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

TABLE 1.8: **Static Variance Decomposition**

The table reports the average share of variance explained by the common factors (financial, global and regional) in the sample period, grouped by region or sector.

	Fin (%)	Glob (%)	Reg (%)	Residual (%)
North America	24.42	7.94	5.56	62.08
Latin America	8.53	11.45	9.21	70.81
Asia-Pacific	6.22	11.35	8.65	73.78
Western Europe	20.33	14.16	6.76	58.75
Emerging Europe	6.72	11.88	7.94	73.46
MEA	0.56	4.99	18.94	75.51
Basic Materials	11.65	15.24	8.31	64.8
Communications	13.15	10.11	5.9	70.84
Energy	12.74	17.45	6.06	63.75
Consumer, Cyclical	12.57	10.34	7.69	69.4
Financial	18.38	12.06	8.13	61.43
Technology	15.21	7.9	7.34	69.55
Industrial	14.91	11.9	8.21	64.98
Consumer, Non-cyclical	9.26	9.52	7.38	73.85
Utilities	7.82	10.85	10.73	70.59
Diversified	9.69	11.46	10.95	67.9

$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 1.9: **Model Estimates**

The table reports cross-sectional averages of factor loadings' magnitude, their persistence (AR(1) parameter) and their volatilities, aggregated by either region or sector. The magnitude of the loadings is estimated via OLS from a static factor model on the full sample, and via rolling OLS regressions with a five-year (*RW5*) and half-year (*RWH*) window size. The volatility of the loadings is estimated as the sample standard deviation of the time-varying sensitivities estimated via MLE (*KF*), and as the sample standard deviation of the time-varying sensitivities estimated via rolling least squares. The AR(1) parameter is estimated via MLE from the dynamic factor model. The table also reports the average persistence, and the number of stocks having AR(1) parameter greater than 0.5 in absolute value, together with the number of stocks with loadings which vary so little that I consider them constant.

$$X_{i,t} = \lambda_{i,t}^{fin} \; F_t^{fin} + \lambda_{i,t}^{glob} \; F_t^{glob} + \lambda_{i,t}^{reg} \; F_{r,t}^{reg} \; \mathbf{1}_{\{i \in r\}} + e_{i,t}$$

	Financial								
	Magnitude			Volatility			Persistence		
	$\hat{\lambda}_i^{fin}$ *OLS*	$\hat{\lambda}_{i,t}^{fin}$ *RW5*	$\hat{\lambda}_{i,t}^{fin}$ *RWH*	Std $\hat{\lambda}_{i,t}^{fin}$ *KF*	Std $\hat{\lambda}_{i,t}^{fin}$ *RW5*	Std $\hat{\lambda}_{i,t}^{fin}$ *RWH*	AR(1)	#AR(1)> 0.5	#Static
North America	0.47	0.463	0.446	0.167	0.068	0.21	0.227	85	24
Latin America	0.274	0.263	0.266	0.128	0.07	0.213	0.221	122	26
Asia-Pacific	0.218	0.207	0.207	0.181	0.058	0.215	0.16	202	36
Western Europe	0.436	0.416	0.412	0.184	0.062	0.2	0.183	177	39
Emerging Europe	0.236	0.222	0.204	0.208	0.078	0.219	0.156	62	16
MEA	0.065	0.064	0.061	0.166	0.047	0.213	0.418	14	6
Basic Materials	0.298	0.29	0.286	0.172	0.06	0.208	0.175	67	15
Communications	0.334	0.321	0.316	0.17	0.066	0.209	0.074	43	14
Energy	0.337	0.342	0.326	0.149	0.057	0.211	0.251	54	14
Consumer, Cyclical	0.318	0.3	0.287	0.187	0.072	0.212	0.134	79	23
Financial	0.382	0.366	0.369	0.182	0.062	0.198	0.211	129	19
Technology	0.357	0.348	0.337	0.169	0.055	0.21	0.153	23	9
Industrial	0.35	0.334	0.324	0.166	0.065	0.209	0.152	97	30
Consumer, Non-cyclical	0.274	0.256	0.253	0.195	0.063	0.219	0.235	112	15
Utilities	0.251	0.241	0.238	0.164	0.074	0.223	0.267	54	7
Diversified	0.293	0.291	0.276	0.174	0.06	0.209	0.223	4	1

	Global								
	Magnitude			Volatility			Persistence		
	$\hat{\lambda}_i^{glob}$ *OLS*	$\hat{\lambda}_{i,t}^{glob}$ *RW5*	$\hat{\lambda}_{i,t}^{glob}$ *RWH*	Std $\hat{\lambda}_{i,t}^{glob}$ *KF*	Std $\hat{\lambda}_{i,t}^{glob}$ *RW5*	Std $\hat{\lambda}_{i,t}^{glob}$ *RWH*	AR(1)	#AR(1)> 0.5	#Static
North America	0.245	0.227	0.216	0.158	0.049	0.184	-0.017	96	16
Latin America	0.322	0.312	0.289	0.097	0.056	0.2	0.13	114	65
Asia-Pacific	0.314	0.301	0.292	0.14	0.052	0.197	0.051	216	104
Western Europe	0.367	0.344	0.336	0.113	0.049	0.183	0.121	238	136
Emerging Europe	0.329	0.304	0.289	0.114	0.066	0.198	0.227	98	57
MEA	0.219	0.215	0.205	0.133	0.051	0.204	-0.118	8	20
Basic Materials	0.371	0.348	0.341	0.115	0.055	0.19	-0.044	87	47
Communications	0.304	0.287	0.278	0.121	0.048	0.195	0.071	60	35
Energy	0.404	0.373	0.361	0.091	0.059	0.192	0.307	44	49
Consumer, Cyclical	0.301	0.282	0.27	0.135	0.052	0.195	0.031	91	42
Financial	0.327	0.313	0.301	0.145	0.047	0.182	0.13	145	70
Technology	0.268	0.249	0.24	0.135	0.048	0.193	0.02	31	12
Industrial	0.33	0.305	0.296	0.119	0.058	0.191	0.109	118	69
Consumer, Non-cyclical	0.287	0.279	0.263	0.131	0.052	0.197	0.048	134	44
Utilities	0.306	0.294	0.279	0.101	0.056	0.201	0.204	56	29
Diversified	0.324	0.317	0.281	0.106	0.048	0.203	0.545	4	1

	Regional									
	Magnitude			Volatility			Persistence			
	$\hat{\lambda}_i^{reg}$ *OLS*	$\hat{\lambda}_{i,t}^{reg}$ *RW5*	$\hat{\lambda}_{i,t}^{reg}$ *RWH*	Std $\hat{\lambda}_{i,t}^{reg}$ *KF*	Std $\hat{\lambda}_{i,t}^{reg}$ *RW5*	Std $\hat{\lambda}_{i,t}^{reg}$ *RWH*	AR(1)	#AR(1)> 0.5	#Static	N
North America	0.2	0.198	0.19	0.146	0.048	0.188	0.269	120	18	232
Latin America	0.246	0.24	0.228	0.121	0.061	0.201	0.278	104	57	218
Asia-Pacific	0.182	0.186	0.174	0.193	0.046	0.193	0.042	168	28	484
Western Europe	0.244	0.246	0.233	0.159	0.047	0.182	0.171	233	23	518
Emerging Europe	0.14	0.131	0.138	0.129	0.049	0.194	0.223	91	49	208
MEA	0.419	0.415	0.398	0.131	0.044	0.191	0.192	13	7	32
Basic Materials	0.214	0.217	0.21	0.16	0.047	0.19	0.189	79	26	176
Communications	0.194	0.198	0.187	0.156	0.048	0.192	0.047	56	11	124
Energy	0.167	0.167	0.155	0.182	0.048	0.186	0.19	42	7	113
Consumer, Cyclical	0.21	0.207	0.202	0.149	0.048	0.193	0.131	91	25	212
Financial	0.209	0.209	0.198	0.168	0.047	0.182	0.19	135	30	338
Technology	0.219	0.215	0.203	0.158	0.046	0.191	0.189	25	16	67
Industrial	0.219	0.22	0.21	0.155	0.047	0.186	0.181	120	21	263
Consumer, Non-cyclical	0.209	0.206	0.198	0.158	0.052	0.198	0.149	127	29	282
Utilities	0.261	0.259	0.242	0.148	0.054	0.201	0.199	52	14	108
Diversified	0.205	0.21	0.212	0.22	0.053	0.177	-0.003	2	3	9

$N = 1692$

$T = 700$ (full-sample, 6th Jan 2006 - 31st May 2019)

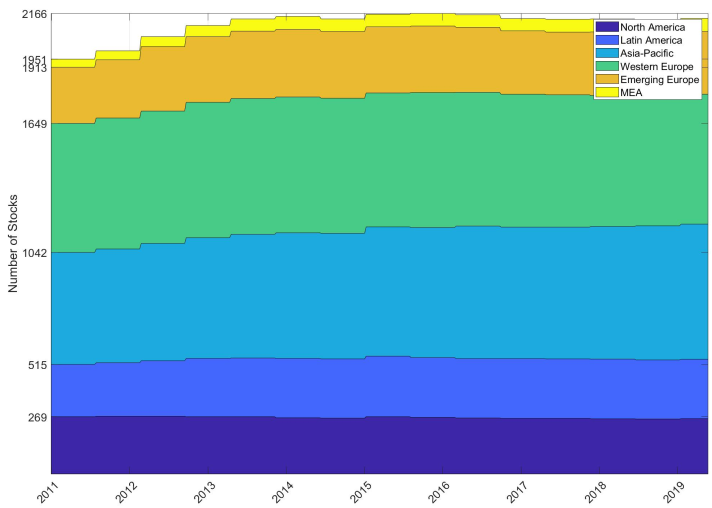
$T = 440$ (five-year, 7th Jan 2011 - 31st May 2019)

$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

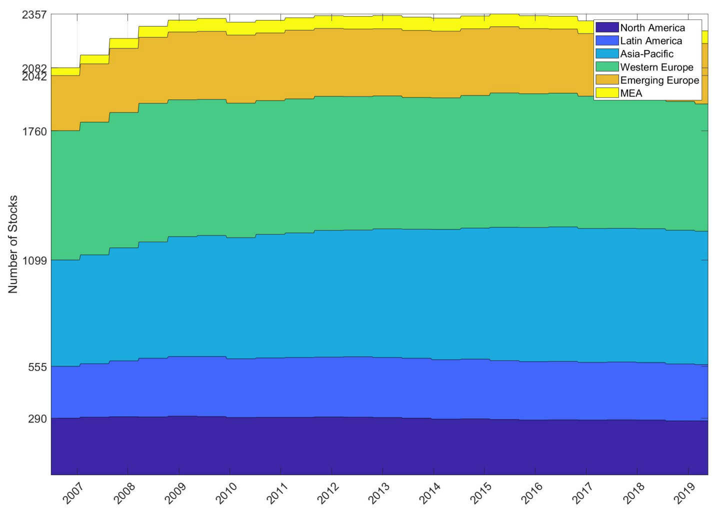
FIGURE 1.1: **Unbalanced Panel of Stock Returns**

The figure plots the number of index members at the beginning of each month that are considered for the rolling OLS scheme, aggregated by region. Panel 1.1a reports the number of stocks that entered the indexes in the period from January 2011 to May 2019, when I employ a long window made of the most recent five-year observations. Panel 1.1b reports the analogous considering a longer time span from July 2006 to May 2019, which corresponds to a shorter half-year window. I select all the equities that entered the national stock indexes with at least two years of available data, and no more than eight consecutive missing observations on the full sample.

(A) *Five-Year*



(B) *Half-Year*



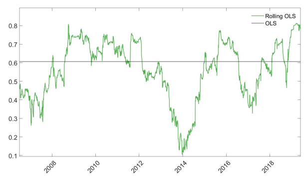
$T = 439$ (five-year, 7th Jan 2011 - 31st May 2019)

$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

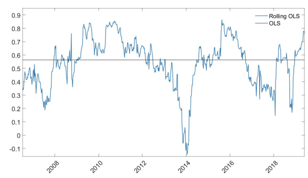
FIGURE 1.2: **Rolling OLS Betas and Data Frequency**

The figure plots the rolling OLS beta estimates on the S&P500 for the returns of Apple Inc. from July 2006 to May 2019 using a fixed window of six months, with varying sampling frequency. Panel 1.2a reports the daily OLS estimates, panel 1.2b the weekly estimates, and panel 1.2c the monthly estimates. In each panel, I also report the OLS estimate on the full sample as a black solid horizontal line. Finally, panels 1.2d and 1.2e compare the empirical distribution of the estimated betas for varying sampling frequency. Stocks and market factor returns are standardised to have zero mean and unit variance.

(A) *Daily*

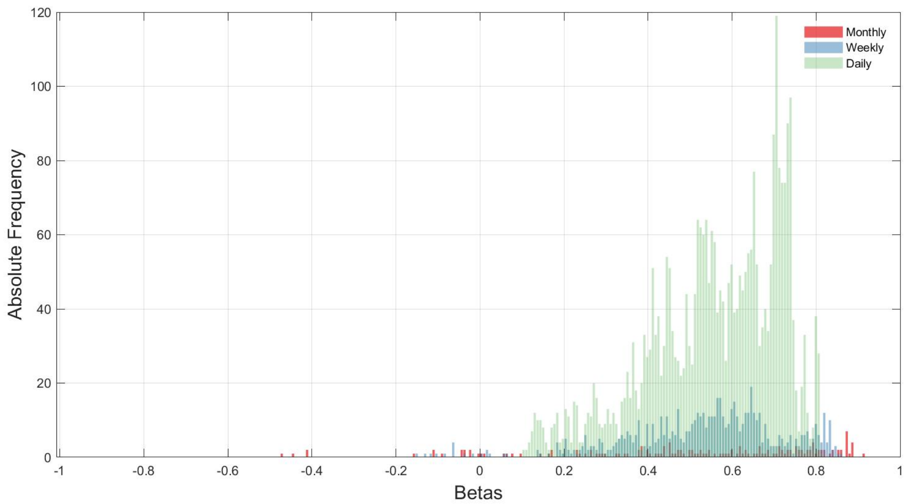


(B) *Weekly*



(C) *Monthly*

(D) *Histogram of Estimated Betas*



(E) *Summary Statistics*

Frequency	Mean	Min	Median	Max	Std	Skew	Kurt	N
Daily	0.562	-0.105	0.582	0.813	0.157	-0.706	2.879	3249
Weekly	0.538	-0.146	0.561	0.865	0.185	-0.774	3.993	674
Monthly	0.512	-0.468	0.577	0.918	0.313	-1.007	3.57	155

$T = 3249$ (daily, 7th Jul 2006 - 31st May 2019

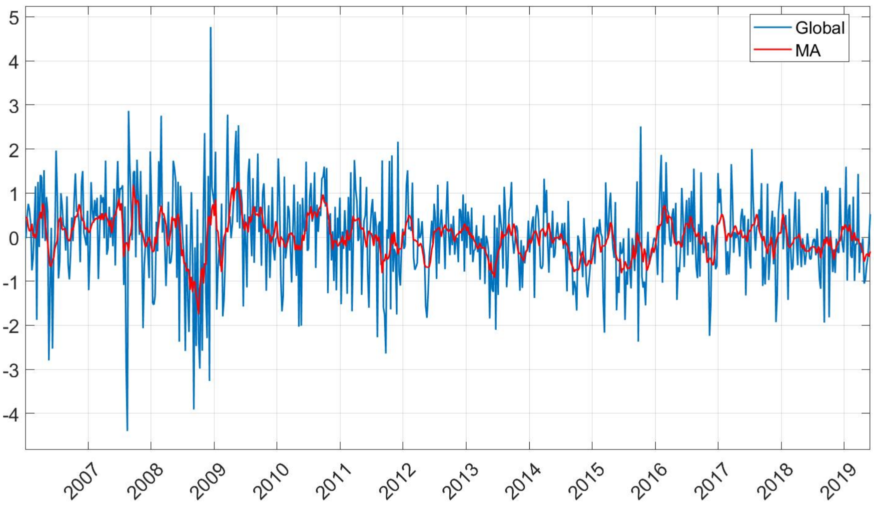
$T = 674$ (weekly, 7th Jul 2006 - 31st May 2019)

$T = 155$ (monthly, 1st Jul 2006 - 1st June 2019)

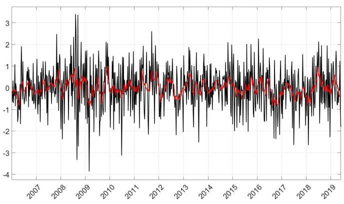
FIGURE 1.3: **Estimated Global and Regional Factors**

The figure plots the estimated global factor, and the regional factors for Asia-Pacific, Emerging Europe, Latin America, Middle-East Africa, North America and Western Europe. Together with the estimated factors, I also plot a double-sided two-month moving average. The factors are estimated by PCA from the model with static loadings in equation (1.9). The factors are rotated to ensure that they are positively correlated with the stock market index of the biggest country in the region.

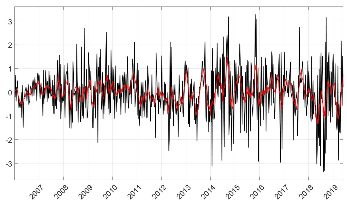
(A) *Global*



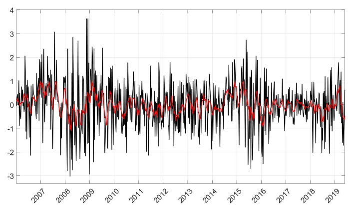
(B) *North America*



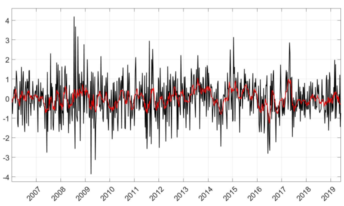
(C) *Latin America*



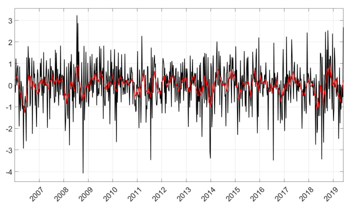
(D) *Asia-Pacific*



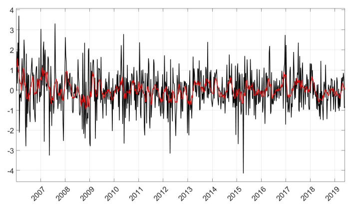
(E) *Western Europe*



(F) *Emerging Europe*



(G) *MEA*



$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

FIGURE 1.4: **Time-Varying R^2 from Rolling OLS Regressions by Region**

The figure plots, for every region, the time-series of R^2 from the rolling OLS regressions with a five-year window. I also report the R^2 calculated using the common component from the static model in red, and the R^2 from the dynamic model with time-varying betas estimated via MLE, black line.

(A) *North America - Five-Year*

(B) *Latin America - Five-Year*

(C) *North America - Half-Year*

(D) *Latin America - Half-Year*

(E) *Emerging Europe - Five-Year*

(F) *Western Europe - Five-Year*

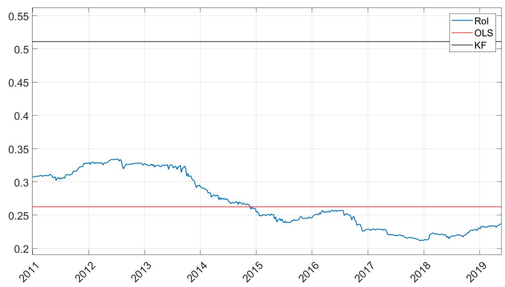
(G) *Emerging Europe - Half-Year*

(H) *Western Europe - Half-Year*

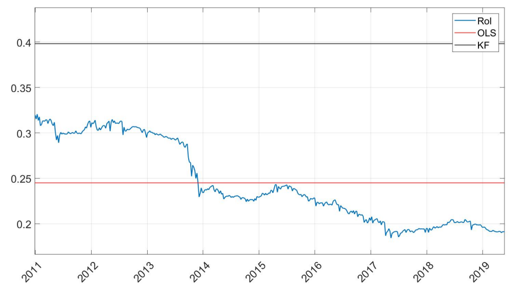
(To be continued)

(Continued)

(I) *Asia-Pacific - Five-Year*

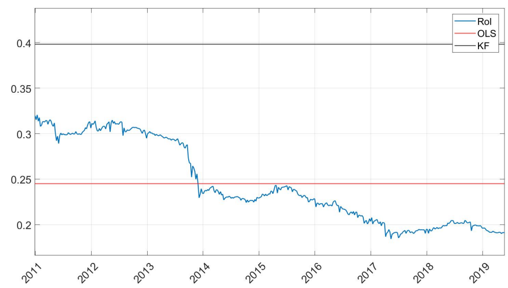


(J) *MEA - Five-Year*



(K) *Asia-Pacific - Half-Year*

(L) *MEA - Half-Year*



$N = 1692$

$T = 700$ (full-sample, 6th Jan 2006 - 31st May 2019)

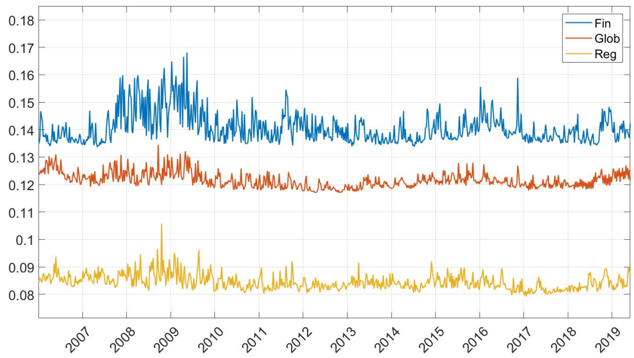
$T = 439$ (five-year, 7th Jan 2011 - 31st May 2019)

$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

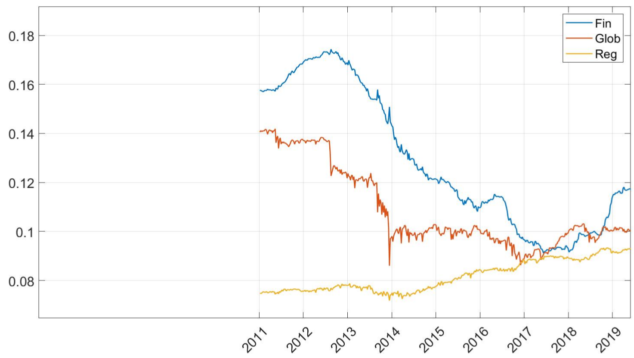
FIGURE 1.5: **Time-Varying Variance Decomposition**

The figure plots the time-varying variance decomposition based on the estimated factor betas. Panel 1.5a shows the cross-sectional average share of variance explained by each factor, with time-varying betas estimated via MLE (*KF*). Panel 1.5b reports the analogous for the time-varying betas estimated via rolling least squares with a fixed window of five years (*Five-Year*). Finally, panel 1.5c reports the rolling OLS results considering a window made of the most recent half-year observations (*Half-Year*). The blue line represents the share of variance explained by the financial factor, yellow and orange lines represent the regional and global factors, respectively.

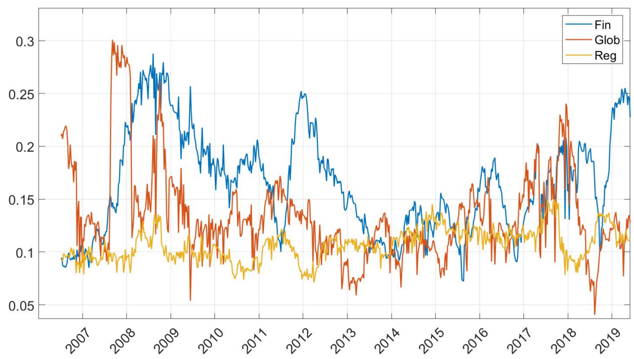
(A) *KF*



(B) *Five-Year*



(C) *Half-Year*



$N = 1692$

$T = 700$ (6th Jan 2006 - 31st May 2019)

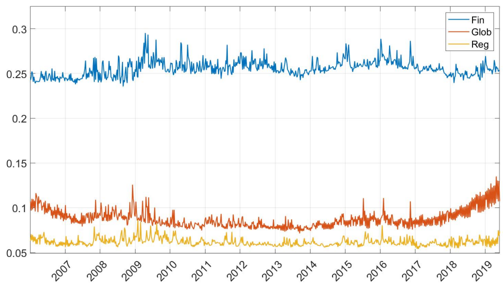
$T = 440$ (five-year, 7th Jan 2011 - 31st May 2019)

$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

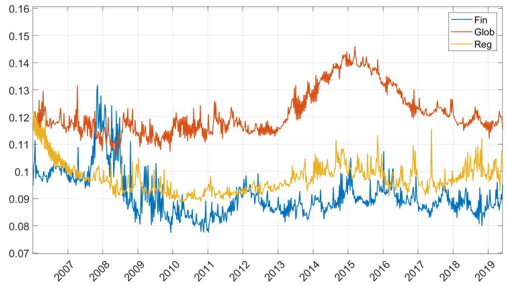
FIGURE 1.6: **Time-Varying Variance Decomposition by Region**

The figure plots, for every region, the time-varying variance decomposition based on the estimated factor betas. The panels report the cross-sectional average share of variance explained by each factor, with time-varying betas estimated via MLE (*KF*), rolling OLS with a five-year window (*Five-Year*), and rolling OLS with a half-yer window (*Half-Year*). The blue line represents the share of variance explained by the financial factor, yellow and orange lines represent the regional and global factors, respectively.

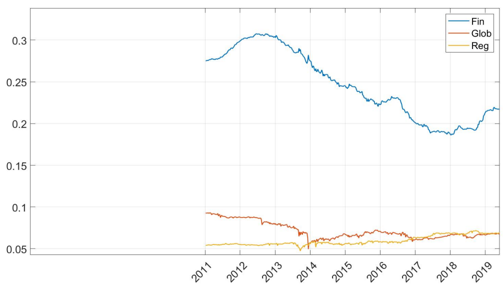
(A) *North America - KF*



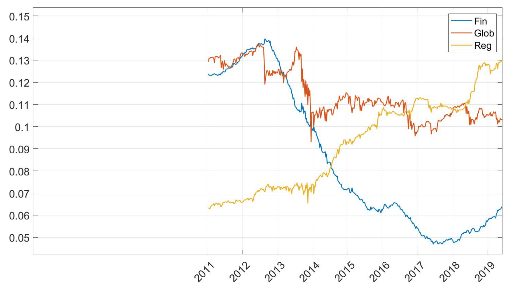
(B) *Latin America - KF*



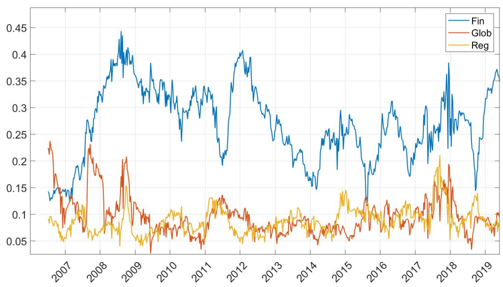
(C) *North America - Five-Year*



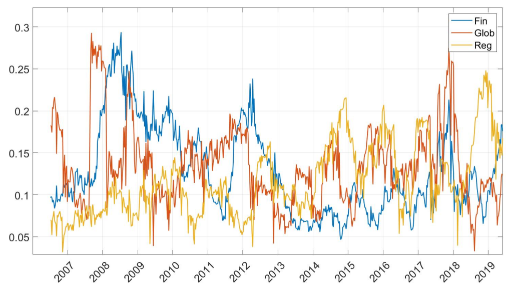
(D) *Latin America - Five-Year*



(E) *North America - Half-Year*



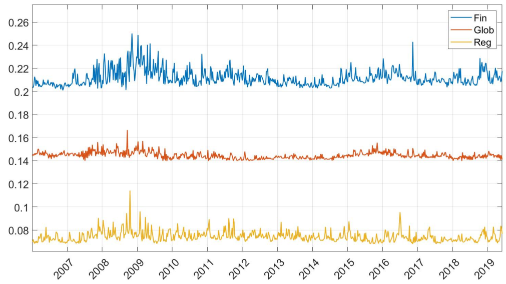
(F) *Latin America - Half-Year*



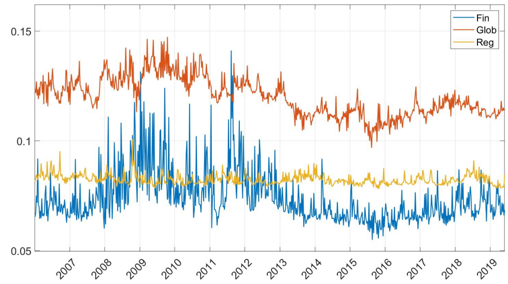
(To be continued)

(Continued)

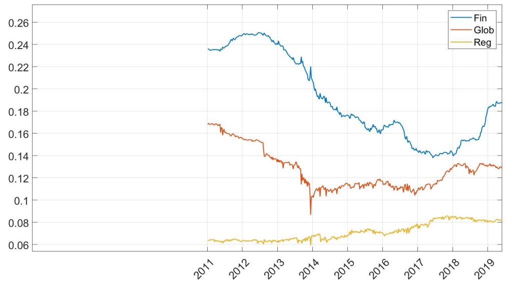
(G) *Western Europe - KF*



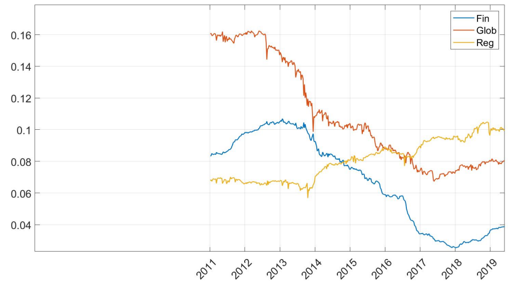
(H) *Emerging Europe - KF*



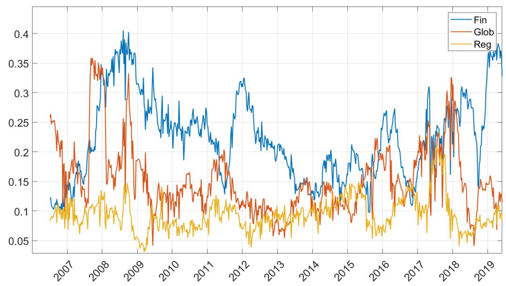
(I) *Western Europe - Five-Year*



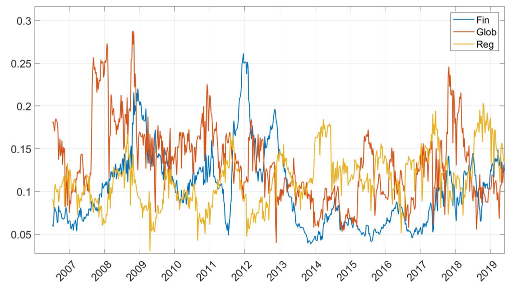
(J) *Emerging Europe - Five-Year*



(K) *Western Europe - Half-Year*



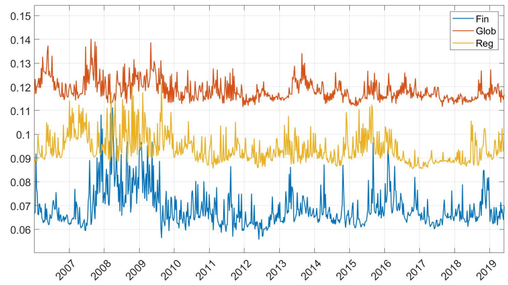
(L) *Emerging Europe - Half-Year*



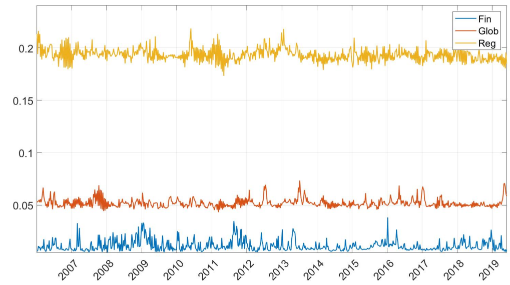
(To be continued)

(Continued)

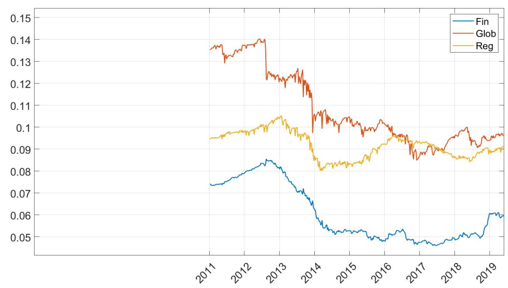
(M) *Asia-Pacific - KF*



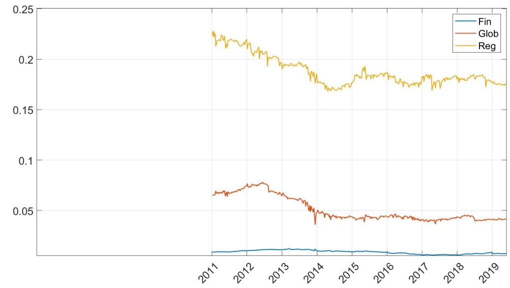
(N) *MEA - KF*



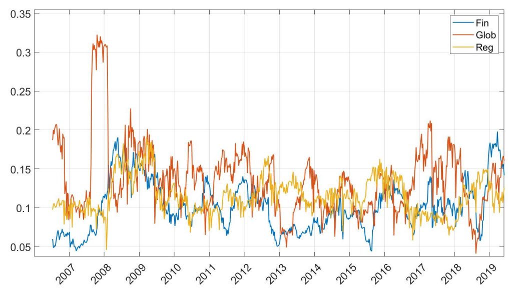
(O) *Asia-Pacific - Five-Year*



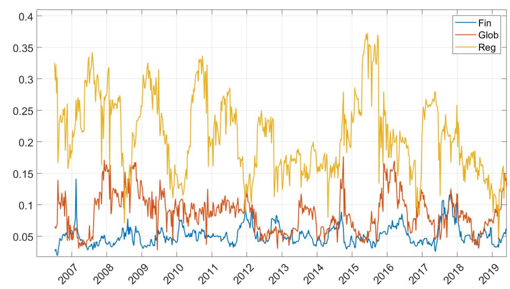
(P) *MEA - Five-Year*



(Q) *Asia-Pacific - Half-Year*



(R) *MEA - Half-Year*

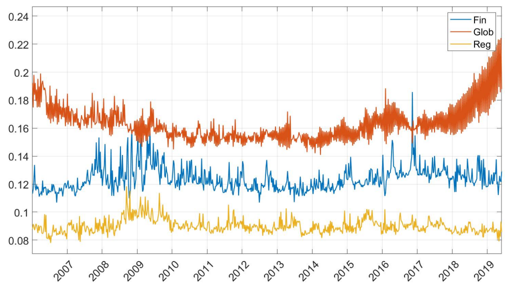


$N = 1692$
$T = 700$ (full-sample, 6th Jan 2006 - 31st May 2019)
$T = 440$ (five-year, 7th Jan 2011 - 31st May 2019)
$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

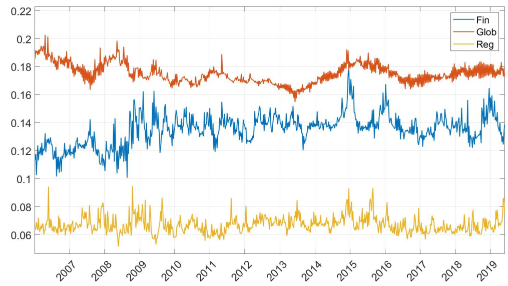
FIGURE 1.7: **Time-Varying Variance Decomposition by Sector**

The figure plots, for every sector, the time-varying variance decomposition based on the estimated factor betas. The panels report the cross-sectional average share of variance explained by each factor, with time-varying betas estimated via MLE (*KF*), rolling OLS with a five-year window (*Five-Year*), and rolling OLS with a half-yer window (*Half-Year*). The blue line represents the share of variance explained by the financial factor, yellow and orange lines represent the regional and global factors, respectively.

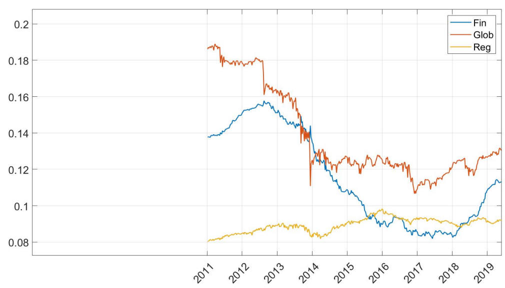
(A) *Basic Materials - KF*



(B) *Energy - KF*

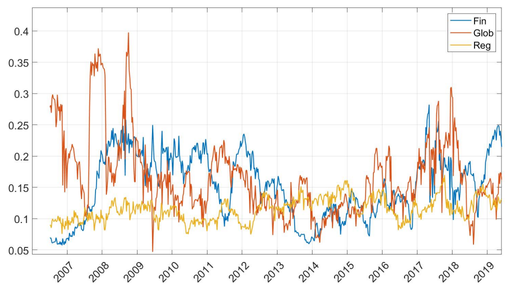


(C) *Basic Materials - Five-Year*

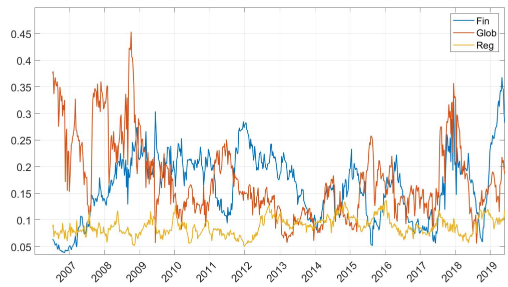


(D) *Energy - Five-Year*

(E) *Basic Materials - Half-Year*



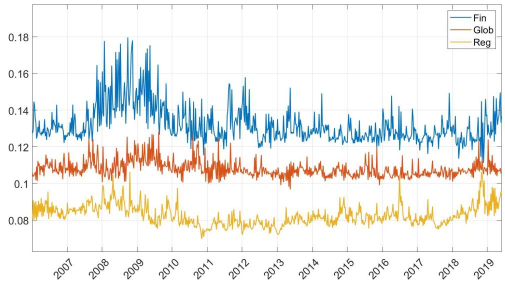
(F) *Energy - Half-Year*



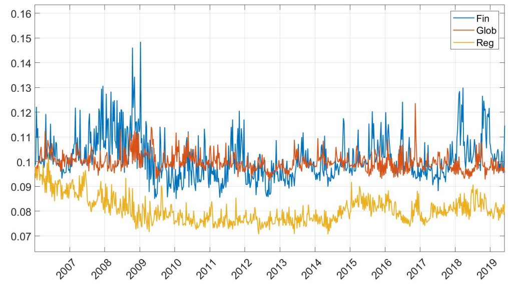
(To be continued)

(Continued)

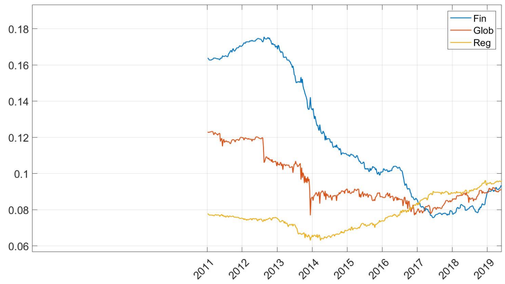
(G) *Consumer, Cyclical - KF*



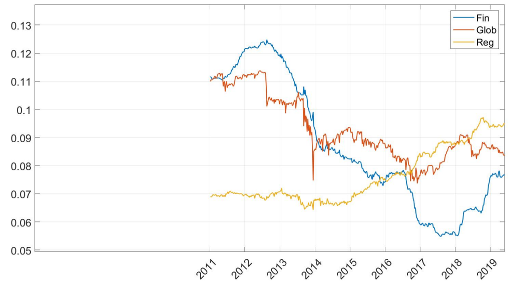
(H) *Consumer, Non-cyclical - KF*



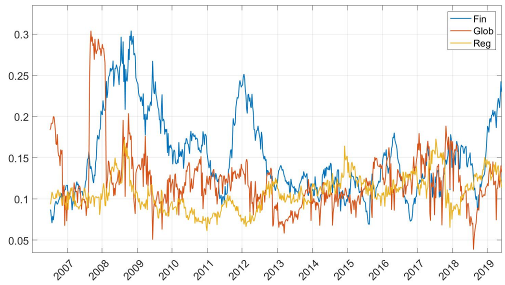
(I) *Consumer, Cyclical - Five-Year*



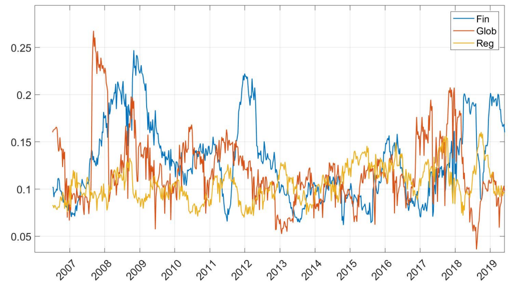
(J) *Consumer, Non-cyclical - Five-Year*



(K) *Consumer, Cyclical - Half-Year*



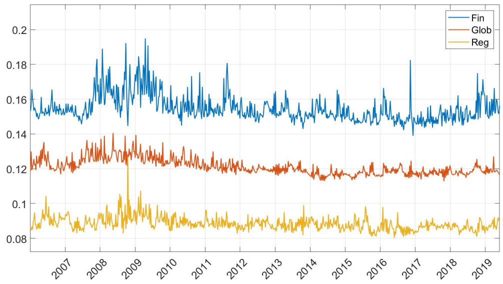
(L) *Consumer, Non-cyclical - Half-Year*



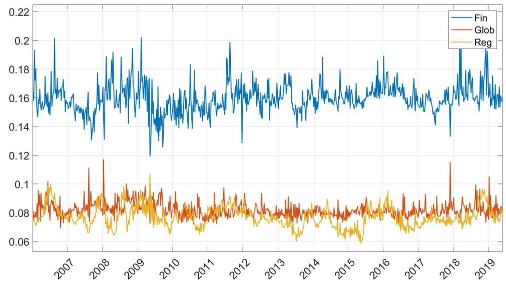
(To be continued)

(Continued)

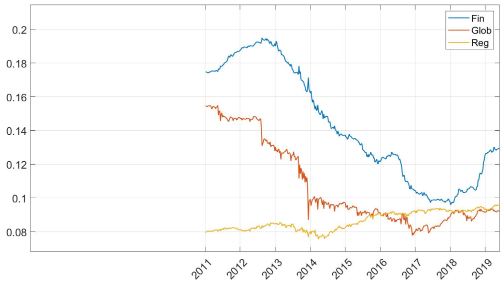
(M) *Industrial - KF*



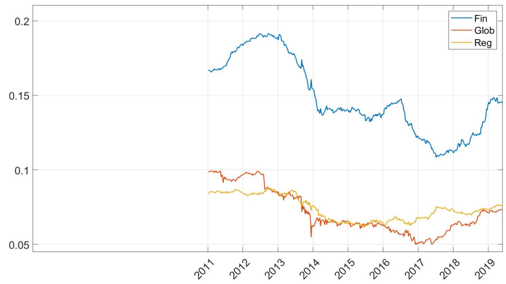
(N) *Technology - KF*



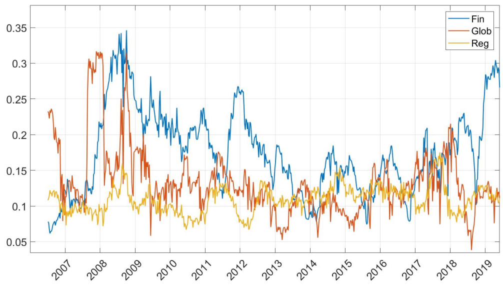
(O) *Industrial - Five-Year*



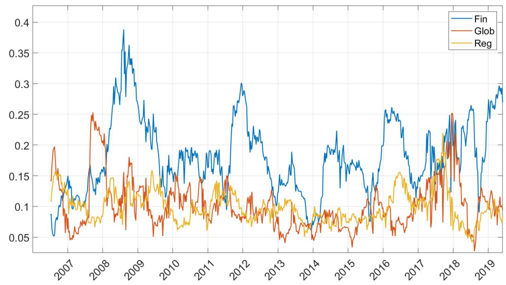
(P) *Technology - Five-Year*



(Q) *Industrial - Half-Year*



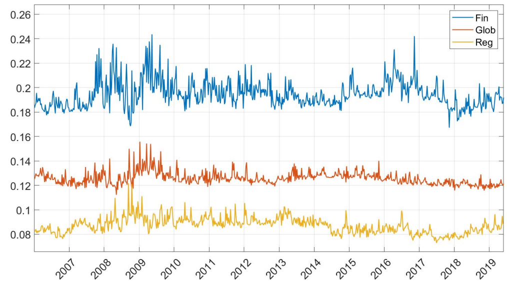
(R) *Technology - Half-Year*



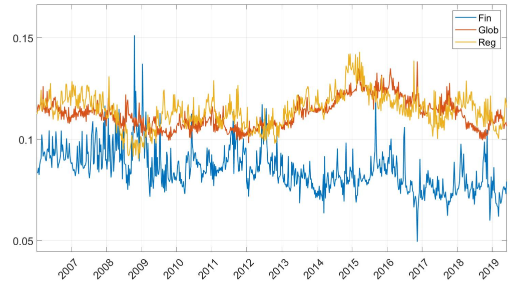
(To be continued)

(Continued)

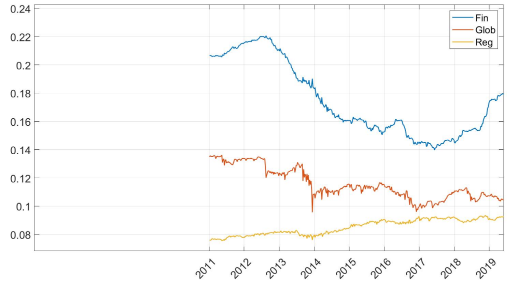
(S) *Financial - KF*



(T) *Utilities - KF*

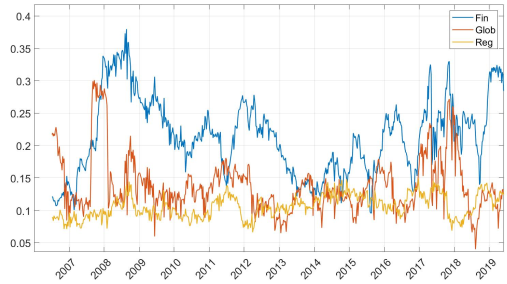


(U) *Financial - Five-Year*

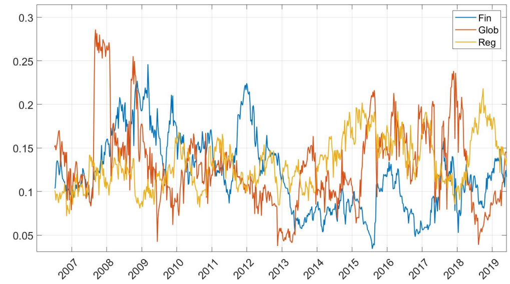


(V) *Utilitie - Five-Year*

(W) *Financial - Half-Year*



(X) *Utilities - Half-Year*



$N = 1692$
$T = 700$ (full-sample, 6th Jan 2006 - 31st May 2019)
$T = 440$ (five-year, 7th Jan 2011 - 31st May 2019)
$T = 674$ (half-year, 7th Jul 2006 - 31st May 2019)

Chapter 2

The Role of Rolling Betas in the Cross-Section of Stock Returns

Abstract

In this chapter I examine the out-of-sample predictive performance of conditional asset pricing models of stock returns. My models feature time-varying factor sensitivities, which are estimated on a real-time basis using rolling least squares, and static unobserved factors, that are estimated via PCA on the full sample. Conditional on the estimated factor space, I evaluate the out-of-sample predictive ability of time-varying loadings to explain patterns in expected future returns. I analyse the results using different measures to document the relative performance of the least square estimator for various choices of the window size. Considering from as little as 26 observations up to 520, I find a trade-off between window size and out-of-sample R^2, with a window made of two years of weekly data (104 observations) being optimal out-of-sample. Across model specifications, I find that the candidate model of Borghi et al. (2018) that combines observed and latent factors performs better from a statistical viewpoint than the commonly used CAPM and FF models under all measures. My findings are based on a large panel of stock returns coming from 40 different countries in the period January 2006 to May 2019.

2.1 Introduction

Rolling estimation is one of the most popular methods in out-of-sample forecasting as it provides a simple way to update the sample information based on the assumptions that the underlying model parameters are time-varying. In asset pricing this approach was popularised by the work of Ferson and Harvey (1991) and Fama and French (1997) who use rolling least squares to estimate the time-varying factor betas in their predictive models for stocks and bonds returns. Their methodology is based on the standard five-year window pattern with data sampled at monthly frequency, and a short one-period ahead forecasting horizon. Despite its widespread use in the literature, few studies have examine the properties of the rolling OLS estimator in predicting future return patterns.

The goal of this chapter is to study in isolation the role of time-varying factor sensitivities in explaining contemporaneous return patterns and predicting future stock returns. To do so I adopt a conditional factor model featuring static (ex-post) factors and dynamic (real-time) factor sensitivities. My modelling setup is inspired by Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021) who propose a conditional factor model featuring time-varying betas and latent factors to study the cross section of corporate bonds and stocks returns respectively. In their framework, factors and loadings are jointly estimated from a rich set of observable asset characteristics which drive the dynamics of the model. I modify their framework in two ways: firstly I only require the loadings to be time-varying, and secondly I only use information contained solely in the asset returns for their estimation. In this setup, the dynamics of systematic risk are driven exclusively by conditional factor exposures (and not by the combination of factors and betas), which are estimated via rolling least squares regression. The choice of the window size determines the characteristics of the conditioning set available at each time for the estimation of the factor exposures, which in turns shapes the dynamics of systematic risk in my model.

The answer to the question on how many observations should be used for the estimation of the factor sensitivities has often been judgemental and based on past experience. The five-year benchmark with monthly observations was pioneered by Fama and MacBeth (1973) two-pass methodology, similarly Ferson and Harvey (1991) study the cross-section of US stocks and bonds with factor sensitivities estimated every 60 months, and Fama and French (1997) also conform to the monthly rolling window scheme. The work of Fama and French (1997) differs from previous studies because it explicitly acknowledges the possibility of using a shorter window made three and four years of past returns to tackle the instability in the slope estimates. They find that for both the CAPM and three-factor model the quality of the one-period ahead forecasts is not sensitive to the length of the regression estimation period: averages of the forecast errors are similar for three-, four-, and five-year estimation periods, which in turn are similar to those produced on the full-sample. For the three-factor slopes the authors find some evidence of mean-reversion in the industry loadings on the MKT, SMB, and

HML factors, however this result is true for short-term forecasts of less than two years, after which the full-period constant-slope regressions are superior to the rolling regressions, and does not appear to be consistent across the group of stocks considered (industry-wise).

Petkova and Zhang (2005) also follow to the FF framework and use portfolios as test assets to investigate the dynamics of value-minus-growth strategies. They present their empirical findings using a standard 60-month window, and find similar results using 24-, 36-, 48-month rolling windows. The paper of Petkova and Zhang (2005) is related to the work of Lewellen and Nagel (2006)[LN thereafter] who test the unconditional CAPM and explore the idea of using short-window regressions to estimate time-varying alphas and betas. The key assumption is that beta is fairly stable during a month or quarter, so that each short-window regression can treat it as constant. LN estimate quarterly betas using daily returns, semi-annual betas using daily and weekly returns, and annual betas using monthly returns. From a statistical perspective, they find their short-window regressions capture nearly all of the impact of time-varying betas in their asset pricing tests. Their paper is considered to be one of the first to use higher-frequency data other than monthly to examine the explanatory ability of time-varying factor betas in the cross-section of stock returns.

In a similar fashion, the estimation methodology of time-varying betas via rolling OLS in Ang, Chen, and Xing (2006) is based on the use of short one-year samples using daily data (rather than using a single long sample of monthly data). Their approach also leverages higher frequency data to characterise the time-variation in the betas over shorter samples, and it is a key reference in the literature studying factor models with returns sampled at daily frequency. Similarly to LN and Ang, Chen, and Xing (2006), Bekaert, Hodrick, and Zhang (2009) also employs a short-window framework made of half-year of weekly observations in their 'time-varying beta' model to study international stock return comovements. Throughout this chapter I adopt a more flexible approach and estimate the factor sensitivities using a short-window approach ($W = 26$, and 52 weekly observations), but also consider longer (sub-) sample sizes ($W = 104, 260$, and 520 weekly observations). Table 2.1 reports a summary of the key features of the papers in the literature that I review.

[Table 2.1 about here.]

In my analysis I examine asset pricing models for individual stock returns featuring factors that are ad hoc pre-specified, as well as latent drivers that are estimated via PCA on the full sample. I use individual stock returns as base assets and include tickers listed in more than 40 countries. Given the international composition of my universe, I differentiate between global (strong) factors that load on all stocks, and local (weak) factors that affect only a portion of the cross-section of returns. The focus of this chapter is on the regional classification of stocks, which is taken from Bekaert et al. (2014). I partition my universe into five world regions and let the local factors represent region-specific sources of variation in the returns. This choice is motivated by Bekaert, Hodrick, and Zhang (2009) who find little evidence for the relevance of industry- versus regional-style factors in explaining international stock return comovements.

PCA treats the loadings as constant over time and I use it to extract the latent factors considering the information contained in the asset returns on the full sample. Standard PCA theory assumes that the latent factors have zero mean by construction, a feature which would not allow me to study the ability of statistical factor models to predict future return patterns based on the models' expected return decomposition. To overcome this problem, I propose an alternative rotation of the latent factors that recovers the information in the asset means and allow my factors to have non-zero prices of risk. This approach is inspired by the recent developments in the literature such as Lettau and Pelger (2020a) and Lettau and Pelger (2020b) who modify the standard PCA objective function and include a penalty based on the unexplained cross-sectional pricing errors (alphas). Their RP-PCA methodology has the advantage of recovering factors that are relevant both for the time-series, as in standard PCA, and the cross-section of asset returns, due to the imposed pricing penalty. The extension that I propose effectively imposes a mapping of the latent drivers to their observed counterparts. In particular, I estimate the risk premium of the global factor as the ex-post return on the equally-weighted portfolio comprised of all stocks in the universe, and equivalently for the local factors I estimate their premia as the realised return on the $1/N$ portfolio made of the stocks domiciled in a specific region.

The observed factor models against which I compare are taken from the international asset pricing literature and include the global CAPM model that features a single 'stock market' factor that loads on all assets in the universe, the region-specific Fama and French (2012)[FF3 thereafter] three-factor model, and the updated Fama and French (2017)[FF5 thereafter] five-factor specification. The candidate statistical factor model, 'Regional model', is taken from Borghi et al. (2018) and features one observed source of systematic variation, the financial factor, and two latent factors.

Conditional on the factor space, I estimate the time-varying sensitivities of the relevant factors via rolling least squares in an out-of-sample framework, and I use a suite of performance measures to assess the models' explanatory and predictive ability. The diagnostics that I use are adapted from Kelly, Palhares, and Pruitt (2021) and examine different aspects of the models' performance. I differentiate between in-sample (IS) measures, which track the performance in explaining contemporaneous stock returns based on the exposures to the common factors, out-of-sample (OOS) measures, which isolate the ability in predicting future stock returns, conditional (COND) measures which are inspired by the conditional MSFE function featured in Inoue, Jin, and Rossi (2017), and finally I also assess the performance of the models via predictive (PRED) measures, which quantify the error in predicting future return patterns based the factor models' expected return decomposition. These indicators are asset- and time- specific, and I construct R^2-based performance metrics similar to Kelly, Palhares, and Pruitt (2021), as well as canonical MEA and MSE metrics to aggregate the results across dimensions.

The time span for my analysis is January 2006 to end-of May 2019 ($T = 700$ weeks). This period includes financial and macro events that had an impact on virtually all stocks in my universe, as well as events that are relevant to the equity markets of specific world regions. To ease the economic interpretation of my results I use an economic calendar that distinguishes global and local turmoil periods in each region. These events include the global financial crisis (GFC), the European sovereign debt crisis (ESDC), the US presidential elections in 2012 and 2016, and the 2015-2016 Chinese stock market crash. My sample includes $N = 1686$ tickers listed in 40 different countries with sufficiently long price history.

I analyse my findings on different dimensions. Firstly I compare the information content of the factors across models, with a focus on assessing the differences between statistical factors and the benchmark observed factors. Secondly, I compare the results of the full-sample OLS beta estimates across models as a basis to interpret the dynamic-loadings analysis that follows. I analyse the statistical properties of the rolling beta estimates for varying window size, and I describe the findings on the models' in-sample and out-of-sample performance based on the metrics that I define. Finally, I also try and find an economic interpretation of my results based on the financial and macro events listed in my economic calendar.

Benchmark factors. Based on rolling correlation analysis, I corroborate the evidence in Fama and French (2015) that the value factor becomes redundant for describing average return patters with the addition of the profitability and investment factors. I also find that the CMA investment factor offers a modest level of protection against market-wide drops in all regions considered, this is true especially during the GFC. Investing into the RMW portfolio would have also protected a US-based investor during the Debt crisis, although the this does not apply during market turmoil periods in other markets, such as the Chinese stock market crash in early 2015 (Asia-Pacific region).

Statistical factors. A key difference of the FF factors with respect to the ones featured in the Regional model is that the former are not orthogonal to each other, which implies that they do not effectively isolate different sources of systematic variation in the returns. In the candidate model, the financial factor is identified with the S&P500 Financials equity index, which tracks the performance of the biggest US financial tickers, and the orthogonal global factor isolates sources of variation in the excess returns that are linked to financial and macro events that had a world-wide impact. For instance, the post-2009 rebound in the equity markets of all regions appears to be reflected in the series of returns of the global factor, which recovered much more quickly than the financials-only index after the GFC. Similarly, the returns on the global factor turned progressively negative during the 2015-2016 period, which coincides with the Oil glut and the Chinese stock market crash, while the financial factor dynamics appear to be little affected. The third set of factors featured in the Regional model include region-specific drivers that are closely related to the dynamics of an equity index proxy constructed with the relevant stocks of the region (which is also related to the first PC of all stocks in the group). This is true especially in developing markets such as MEA, Latin America, Asia-Pacific and Emerging Europe, which are less influenced by US-specific dynamics tracked by the financial factor.

Static factor exposures. I find that the model of Borghi et al. (2018) featuring latent factors dominates the benchmark models based on industry- and region-specific averages of the individual R^2s from the pricing equations, and t-statistics of the estimated loadings. The single-factor model's performance in explaining contemporaneous stock returns is positive in developed market regions such as North America and Western Europe, while it deteriorates in the developing markets. When I compare the results of the three-factor models, I find that the loadings on the SMB and HML factors in FF3 are on average one order of magnitude lower than the ones featured in the Regional model (and similarly for t-stats). I also find that the estimated loadings on the MKT factor of FF3 are all greater than the ones on the observed financial factor for the Regional model. In the latter specification, the global and regional factors are the predominant drivers of systematic variation in the stock returns for 9 out of the 17 groups (regions and industries), at the expenses of the local 'market' factor. This is true especially in the Asia-Pacific region where the improvement in average explanatory power is the highest across regions (about 9% increase in the average R^2). Adding the RMW and CMA factors to FF3 also does not appear to bring substantial benefits in terms of explained variation in the cross-section of returns. In fact, I notice that the average t-statistics on the HML factor for the developing Asia-Pacific region drop significantly from FF3 to FF5, a further corroboration of the redundancy of the HML factor in the FF5 framework. This is true for most of the regions and sectors, and to a lower extent in North America which maintain the highest (average) t-stat and magnitude of the value factor betas across all groups. I find that the RMW and CMA factors load negatively on financials stocks, with a statistically significant beta coefficient. The loadings on the CMA factor are negative and significant across virtually all industries and regions.

Dynamic factor exposures. Comparing the magnitude of the full-sample estimates with the time-varying counterparts suggests that the former tend to be higher, even when including as much as 520 weekly observations in the estimation window. This is true across all models and factors with the exception of the RMW factor. In fact, in FF5 the magnitude of the time-varying betas on the RMW factor is substantially higher than the corresponding full-sample estimate, indicating that the relevance of the robustness factor in the cross-section of returns becomes more pronounced when I allow for time variation in the loadings. However, the RMW beta estimates remain far from being statistically significant on average for all the stocks. Overall, I find that the sensitivities on the FF factors show little evidence of significant time-series variation, which suggest that their role may be negligible out-of-sample. Compared to the FF models, the estimates of the time-varying factor sensitivities on the Regional model's factors are statistically significant driver of stock return variation for all window sizes. This is true also for the loadings on the MKT model's factor. According to the average R^2s from the rolling time-series regression, I find that for all specification except the MKT model a short window made of as little as half-year observations provides the best performance.

Model performance. Based on the R^2 measures of Kelly, Palhares, and Pruitt (2021), I find that the models' performance to explain and predict future return patterns share a peculiar relation with the size of the estimation window. If I am interested in describing contemporaneous variation in stock returns, the short-window approach of Lewellen and Nagel (2006) has much to recommend. However, for predictive purposes including too little observations for estimation causes the conditional betas to be noisy, which in turn results in forecasts that show little predictive power. The trade-off between the length of the estimation window and the variance of the estimator is resolved around the two-year mark, this is true across all models. The choice of the window length alone accounts for about $\pm 10\%$ of the factor model's out-of-sample forecasting performance (R^2). Comparing across models, I find that FF5 maximises the in-sample R^2 with 46%, followed by the Regional model with $R^2 = 42\%$. The predictive performance of the MKT model based on the out-of-sample R^2 is about 20%, followed by the FF3 and FF5 models with approximately 25%. The model that is best suited to predict future return patterns is the candidate model with an out-of-sample R^2 of 29%. Based on the MSE and MSA measures, I find that for nearly all models the explanatory and forecasting performance is increasing in window size, with the five- and ten-year windows yielding the best results. My results are in line with the optimal window criteria of Inoue, Jin, and Rossi (2017), which features the COND MSE function and yields an average optimal window of about four years of weekly observations, but at odds with the measures of Kelly, Palhares, and Pruitt (2021). This leaves the research question open to further contributions, see Section 2.6.

Economic interpretation. When I interpret economically the beta estimates for the factors featured in the Regional model, I find that the loadings on the global factors (financial and global) tend to increase during turmoil periods. During the GFC and ESDC, for nearly all regions the loadings on the global factors increase considerably. The stocks belonging to the North America region started becoming more sensitive to changes in the financial factor at the start of 2007, and reached a peak towards the end of the GFC in mid 2009. In North America, I show that the sensibility of stock returns to changes in the financial factor increased in the weeks preceding the US presidential election in late 2016. Moreover, during the months corresponding to the US-China trade war, my results suggest that a meaningful decrease in the financial factor loadings in late 2018, at the expenses of the global factor whose loadings increase during the same period. Excluding the US-China trade war, the only other period in which the loadings on the financial factor are comparable in magnitude to the others is at the outset of the Chinese stock market crash in mid-2015. Some of the events considered for the stocks listed in North America are also relevant for those domiciled in the Asia-Pacific region, where I find a material increase in the (latent) global factor loadings during the Chinese stock market crash of 2015-2016, and in the weeks leading to the US presidential election in 2016. In the dates corresponding to the oil market crash, I find that the loadings on the regional factor increase notably at the start of 2015 for the equities in MEA, and similarly that the loadings on the global (latent) factor dominate the others in magnitude in Emerging Europe and Latin America during this period. In Western Europe, I find that in the weeks leading to the Brexit referendum of 2016 the loadings on the regional factor start to increase consistently, while average sensitivity of the stocks in this region to shocks to the financial and global factors decreases.

Organisation of the chapter. The remainder of this chapter is organised as follows. Section 2.2 reviews the properties of the baseline model featuring time-varying sensitivities and static factors, describes the benchmark models from the literature against which I compare, and presents the candidate model for the analysis together with the factor extraction procedure via PCA. Section 2.3 describes the methodology that I use to forecast future stock returns for each model, and details the various metrics I adopt to assess the model's explanatory and forecasting performance. Section 2.4 describes the data, and examine the factor spaces implied by the different model specifications. Section 2.5 provides the results of the rolling least squares estimation of the time-varying betas based on a variety of sample sizes, and assess each model's statistical performance. Finally, Section 2.6 provides the closing remarks and details future research developments. The chapter is accompanied by Appendix B.

2.2 Models

This section comprises two parts. In the first part, Section 2.2.1, I review the properties of the baseline model featuring time-varying sensitivities and static factors, which I use to assess the out-of-sample predictive ability of factor betas in the cross-section of stock returns. I then examine different specifications of the baseline model in Section 2.2.2. In particular, in Section 2.2.2 I consider three models from the empirical Asset Pricing literature as benchmarks in my analysis, and in Section 2.2.2 I describe the properties of the candidate model from Borghi et al. (2018) that disentangles global and local sources of systematic variation in the stock returns via PCA. I introduce a modification of their PC estimator which allow the factors to have non-zero risk premia.

2.2.1 Baseline Model

To investigate whether beta predictability account for the average excess return patterns at horizon h for stock $i \in (1, N)$, with information up to period $t \in (1, T)$, I analyse a generic conditional factor model featuring time-varying factor sensitivities of the form

$$r_{i,t+h} = \boldsymbol{\beta}_{i,t}^{\top}\, \boldsymbol{f}_{t+h} + \epsilon_{i,t+h}, \quad E_t[r_{i,t+h}] = \boldsymbol{\beta}_{i,t}^{\top}\, \boldsymbol{\lambda} \tag{2.1}$$

where $r_{i,t+h}$ is the h-period excess return on stock i, $\boldsymbol{\beta}_{i,t}$ is the $(K \times 1)$ vector of time-varying factor loadings, and $\boldsymbol{f}_{t+h}$ is the $(K \times 1)$ vector of relevant factors. In this framework, conditional expected returns are driven by time-varying factor exposures, $\boldsymbol{\beta}_{i,t}$, and unconditional factor risk premia, $\boldsymbol{\lambda} := E[\boldsymbol{f}_{1:T}]$, with $\boldsymbol{f}_{1:T}$ denoting the $(T \times K)$ matrix of relevant factors in the sample period. Similarly to the setup in Chapter 1, I also require a strong form of arbitrage pricing theory (APT) to hold, so that residual risk has a premium of zero. In cross-sectional form, the model reads

$$\boldsymbol{r}_{t+h} = \boldsymbol{\beta}_t\, \boldsymbol{f}_{t+h} + \boldsymbol{\epsilon}_{t+h} \tag{2.2}$$

where $\boldsymbol{r}_{t+h} = [r_{1,t+h}, ..., r_{N,t+h}]^{\top}$ is the $(N \times 1)$ vector of excess stock returns in period $t + h$, $\boldsymbol{\beta}_t = [\beta_{1,t}^{\top}, ..., \beta_{N,t}^{\top}]^{\top}$ is the $(N \times K)$ matrix of factor sensitivities for N test assets and K common factors, and $\boldsymbol{f}_{t+h}$ is the $(K \times 1)$ vector of future-period factor realisations. I work with a panel in which number of cross-sectional observations N is very large compared the number of time-series observations T, $T \ll N$.

This modelling setup closely follows Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021) who propose a conditional factor model featuring time-varying betas and latent factors to study the cross-section of corporate bond and stock returns respectively. In their framework, factors and loadings are jointly estimated from a rich set of observable asset characteristics which drive the dynamics of the model, motivated by the findings on the strong association between asset-level characteristics and future market betas.

My goal is to study the contribution of time-varying factor betas in predicting future excess returns and, contrarily to Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021), in my setup factors are assumed to be 'static'. They incorporate information contained solely in the returns data on the full sample T (i.e. 'ex-post' factors), and feature unconditional first and second moments. On the other hand, factor exposures are assumed to be time-varying and embody information contained in the most-recent asset return observations (i.e. 'real-time' betas). In this framework, the dynamics of systematic risk are driven exclusively by conditional factor exposures, and I employ a rolling regression framework to introduce the time-variation in betas, see Section 2.3 for details on the forecasting methodology.

In rolling out-of-sample forecasting, once the betas are estimated using the most-recent W data points, I roll one observation forward and re-assess the predictive ability of the model. This procedure is repeated T_W times until all observations T in the sample are considered[1], $T_W = (T - W - h) + 1$ for $(T - W - h) > 1$. If I denote with t^* the time index corresponding to the end of each estimation window, model (2.2) can be evaluated out-of-sample as

$$\boldsymbol{r}_{t^*+h} = \boldsymbol{\beta}_{t^*} \boldsymbol{f}_{t^*+h} + \boldsymbol{\epsilon}_{t^*+h}, \quad t^* = 1, ..., T_W \tag{2.3}$$

with $\boldsymbol{r}_{t^*+h}$ being the $(N \times 1)$ vector of excess returns for period $t^* + h$, $\boldsymbol{\beta}_{t^*}$ is the $(N \times K)$ matrix of factor sensitivities conditional on the information up to (and including) period t^*, and $\boldsymbol{f}_{t^*+h}$ is the K-dimensional vector of future-period factor realisations. Model (2.3) entails the expected return decomposition

$$E_{t^*}[\boldsymbol{r}_{t^*+h}] = \boldsymbol{\beta}_{t^*}^{\top} \boldsymbol{\lambda}, \quad t^* = 1, ..., T_W \tag{2.4}$$

where $E_{t*}[\boldsymbol{r}_{t*+h}]$ is the $(N \times 1)$ vector of expected future returns in period $t*$, and $\boldsymbol{\lambda}$ is the K-dimensional vector of unconditional risk premia. In appendix B.1 I present an extension of my model to analyse what is the role of time-varying betas in shaping the temporal evolution of the co-movements structure implied by the factor model in (2.2). The peculiarity of this setup is that the time-variation in the first- and second- moments of future excess returns is driven only by conditional factor exposures, which extends the Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021) framework to the analysis of covariances. The emphasis of this chapter is on how time-varying betas can forecast future asset returns, equation (2.2), and their first moments, equation (2.4), for various choices of W. I leave the analysis on the second moments for future research, see Section 2.6 for details on further research developments.

[1]The last observation T will always be considered for forecasting purposes only (and not for estimation). Moreover if the quantity $(T - W - h)$ is lower than zero the choice of W and h is infeasible given T observations. If $(T - W - h) = 1$, then $T_w = 1$.

I denote with W the choice of the window size, which determines the characteristics of the conditioning set available at each time t^* for the calculation of the factor exposures. Contrarily to Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021) who use information beyond asset returns to derive the dynamics of systematic risk, in my model the conditioning set Ω_t^* only includes information stored in the returns of the observed factors and test assets. This implies that I am assuming that I am able to accurately forecast factor betas at any given future horizon h given the information content of the most-recent W returns observations[2]

$$E[\boldsymbol{\beta}_{t^*+h} \mid \Omega_{t^*}] = \boldsymbol{\beta}_{t^*}, \quad t^* = 1, ..., T_W. \tag{2.5}$$

This assumption may be restrictive if I am interested in forecasting beta over longer horizons (where h may be a few months), whilst it may be appropriate if I believe that beta varies slowly over time (relative to the sampling frequency) and I am interested in short-horizon forecasts, as Robertson (2018) reports. In Section 2.3.1 I discuss how the choice of h, W and the sampling frequency of the data is intrinsically related to the assumptions I are prepared to make on the time-series evolution of $\boldsymbol{\beta}_{t^*}$, which in turn determine the preferred estimation procedure.

2.2.2 Factor Space

In this section I examine different specifications of the baseline model in equation (2.2). I start by considering three benchmark models from the empirical asset pricing literature in Section 2.2.2, and then describe the candidate model of Borghi et al. (2018) that features latent factors estimated via PCA in Section 2.2.2. The test assets comprise equities listed in multiple countries and as such I differentiate between global and local (region-specific) factors.

Benchmark Observable Factors

I consider the following benchmark models:

- **MKT model.** I construct my own stock market factor as the equal-weighted average of the excess stock returns in my data, considering the complete universe of stocks.

$$\boldsymbol{f}_t = [\ MKT_t\ ] \tag{2.6}$$

  This model is motivated by Harvey (1991) and Fama and French (1998) who modify the local CAPM framework originally designed for me stocks to include international assets for the calculation of the global market factor. The specification in (2.6) is the most parsimonious among the benchmarks and includes just one factor, $K = 1$.

[2] I use W and Ω_{t^*} interchangeably throughout this chapter. W refers to the fixed number of observations that are used for the calculation of conditional factor exposures in each period t^*, which coincides with the information set Ω_{t^*} in the rolling out-of-sample framework. I drop the subscript t^* from the notation W_{t^*} to indicate that the window size is fixed. However, shifting the estimation window one period forward clearly implies a different conditioning set for the calculation of the betas, thus the notation Ω_{t^*}.

- **FF3 model.** I consider the three-factor model of Fama and French (2012) that augments the CAPM model with two factors: SMB, the difference between the returns on a diversified portfolios of small stocks and big stocks, and HML, the difference between the returns on a diversified portfolio comprised of high book-to-market (value) stocks and low book-to-market (growth) stocks.

$$\boldsymbol{f}_t = [\ MKT_{r,t}\ \ SMB_{r,t}\ \ HML_{r,t}] \tag{2.7}$$

  The subscript r in (2.7) indicates that the set of factors is region-specific. In fact, FF3 use model (2.7) to study international stock returns and construct the local portfolios for four different world regions: North America, Japan, Asia-Pacific and Europe. In this chapter I partition my universe into five world regions and Japan is excluded. Thus, to ease the comparison across models I report the results for the FF3 model for three regions only, North America, Asia-Pacific, and Western Europe[3]. The total number of factors for the model is therefore $K = 9$.

- **FF5 model.** I also examine the five-factor model of Fama and French (2017)[FF5 thereafter] which features two additional factors with respect to FF3, RMW and CMA, that represent sources of variation in the returns associated to stock-specific profitability and investment patterns respectively.

$$\boldsymbol{f}_t = [\ MKT_{r,t}\ \ SMB_{r,t}\ \ HML_{r,t}\ \ RMW_{r,t}\ \ CMA_{r,t}\ ] \tag{2.8}$$

  Similarly to FF3, model (2.8) is tested using a large panel of stocks listed in different countries. The classification structure is identical to the one in FF3 and as such I report the results considering three different regions. The total number of factors is $K = 15$.

Candidate Model

The candidate model is taken from Borghi et al. (2018) and includes two global sources of systematic variation, an observed financial factor and a latent global factor, and one regional latent driver. Throughout this chapter I refer to this model as 'Regional model'.

$$\boldsymbol{f}_t = [F_t\ \ G_t\ \ R_{r,t}]. \tag{2.9}$$

F_t is the time-t value of the financial factor, G_t is the global latent factor, and $R_{r,t}$ is the time-t value of the regional factor for stock i belonging to region $r \in (1, R)$.

The peculiarity of the model in Borghi et al. (2018) is that it disentangles global and local sources of systematic variation of stock returns. The total number of factors is K, and includes K^{glob} factors that are relevant for all stocks, and K^{loc} factors that affect only a specific portion of

[3]Details on the regional classification of the stock universe are given in Section 2.4.1.

the cross-section according to a pre-determined pooling criterion, $K = K^{glob} + K^{loc}$. Let R be the total number of clusters that I use to partition the cross-section of N stocks, $\sum_{r=1}^{R} n_r = N$, then the number of local factors is $K^{loc} = R$, assuming for simplicity that a single statistical factor is the main driver of returns belonging to a particular cluster[4].

The focus of my analysis is on the regional classification of the stocks based on the company domicile, thus the local factors conform to region-specific drivers of systematic variation. An alternative would be to classify stocks based on the industry in which they operate and build the local factors accordingly, however Bekaert, Hodrick, and Zhang (2009) find that country-dominate industry-style factors in a model with time-varying weights applied to international stock returns. I leave the analysis on the importance of country versus industry factors for future research, see Section 2.6. The country-region composition that I adopt in this chapter is taken from Bekaert et al. (2014) who partitions their universe into six world regions, thus $R = 6$. Further details on the regional classification of my universe are given in Section 2.4.1.

The two-level factor structure in Borghi et al. (2018) is inspired by the work of Breitung and Eickmeier (2014) and Breitung and Eickmeier (2015) who study comprehensively the problem of modelling and estimating block structures in dynamic approximate factor models. The advantage of adopting a two-level structure is that the assets belonging to cluster r are not influenced by the shocks that are specific to other clusters, a situation that is common in an international context in which I expect certain factors (global) to link all variables in the model, whereas others (local factors) to be associated to some subgroups of assets. In Breitung and Eickmeier (2014) the authors employ a model that imposes zero restrictions on the factor loadings, with factors that are unrestricted, whilst in Breitung and Eickmeier (2015) they interchange the role of the factors and the loadings and impose blocks of zero restrictions on the factor matrix, with time-constant loadings. Concordant to the assumptions of my baseline model, I choose to test the model of Borghi et al. (2018) that leverages the two-level factor structure of Breitung and Eickmeier (2014), and imposes sparsity conditions on the factor loadings matrix as follows

$$\underset{(N\times 1)}{\begin{bmatrix} \boldsymbol{r}_{1,\cdot t+h} \\ \boldsymbol{r}_{2,\cdot t+h} \\ \vdots \\ \boldsymbol{r}_{R,\cdot t+h} \end{bmatrix}} = \underset{(N\times K)}{\begin{bmatrix} \boldsymbol{\beta}^f_{1,\cdot t} & \boldsymbol{\beta}^g_{1,\cdot t} & \boldsymbol{\beta}^r_{1,\cdot t} & \mathbf{0}_{n_r} & \cdots & \mathbf{0}_{n_r} \\ \boldsymbol{\beta}^f_{2,\cdot t} & \boldsymbol{\beta}^g_{1,\cdot t} & \mathbf{0}_{n_r} & \boldsymbol{\beta}^r_{2,\cdot t} & \cdots & \mathbf{0}_{n_r} \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \boldsymbol{\beta}^f_{R,\cdot t} & \boldsymbol{\beta}^g_{1,\cdot t} & \mathbf{0}_{n_r} & \cdots & \cdots & \boldsymbol{\beta}^r_{R,\cdot t} \end{bmatrix}} \underset{(K\times 1)}{\begin{bmatrix} F_{t+h} \\ G_{t+h} \\ R_{1,t+h} \\ R_{2,t+h} \\ \vdots \\ R_{R,t+h} \end{bmatrix}} + \underset{(N\times 1)}{\begin{bmatrix} \boldsymbol{e}_{1,\cdot t+h} \\ \boldsymbol{e}_{2,\cdot t+h} \\ \vdots \\ \boldsymbol{e}_{R,\cdot t+h} \end{bmatrix}} \tag{2.10}$$

where $\boldsymbol{r}_{r,\cdot t+h}$ is the $(n_r \times 1)$ vector of future-period excess returns on the n_r stocks in cluster r, $\boldsymbol{\beta}^f_{r,\cdot t}$, $\boldsymbol{\beta}^g_{r,\cdot t}$ and $\boldsymbol{\beta}^r_{r,\cdot t}$ are the $(n_r \times 1)$ vectors of time-t loadings on the financial, global and

[4]The model can accommodate multiple local risk drivers, for simplicity I limit my analysis to just one per cluster.

regional factors respectively. In total, the number of factors is $K = 8$, with $K^{glob} = 2$ and $R = 6$, and $K_i = 3$ for each stock.

In this chapter, I use PCA to extract the global and latent factors considering information on the full sample T. In conventional factor analysis, the PC estimator of the latent factors treats the loadings as static over time, and attempts to minimise the residual time-series variation in the returns by solving the following optimisation problem, see Stock and Watson (2002),

$$\underset{\hat{\boldsymbol{\beta}}^{PCA},\ \hat{\boldsymbol{f}}^{PCA}}{\operatorname{argmin}} \quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((r_{i,t}-\bar{r}_i)-\left(\boldsymbol{\beta}_i^{\top}\left(\boldsymbol{f}_t-\boldsymbol{\lambda}\right)\right)\right)^2 \tag{2.11}$$

with $\hat{\boldsymbol{\beta}}^{PCA}$ being the $(N \times K)$ matrix of estimated static factor sensitivities, and $\hat{\boldsymbol{f}}^{PCA}$ the $(T \times K)$ matrix of latent factors. The theoretical framework underpinning PCA is characterised by at least three features which in practice make the estimation of the latent factors problematic. The first is the assumption of time-invariant factor loadings, which is in sharp contrast with the conditional framework that I adopt in my baseline model. The second relate to the technical challenge that lies in estimating the relevant factors in a multi-level structure as in (2.10), and to take into account the zero restrictions imposed on the factor loadings matrix. The third is that the objective function in (2.11) depends on estimates of the first moments of asset, $\bar{r}_i$, and factor returns, $\boldsymbol{\lambda}$, two quantities that are crucial for assessing the pricing ability of the factor models considered. In what follows I will assess each problem separately and provide further details on the factor extraction procedure.

Minimising the sum of squared residual in (2.11) is equivalent to maximising the log-likelihood function in the standard Gaussian framework. In fact, the theory of PCA is developed under the assumption of time-invariant loadings, and independent normally distributed idiosyncratic errors $\boldsymbol{e}_{r,\cdot t+h}$, across all dimensions $i \in (1, n_r)$, $r \in (1, R)$, and $t, h \in (1, T)$. Two assumptions that are difficult to defend in empirical applications[5]. However as Breitung and Eickmeier (2014) report, the assumption of independent normally distributed idiosyncratic errors is used to obtain a simple quasi-likelihood function, and it is not strictly necessary for consistent estimation of the latent factors. In their paper, they leverage the results of Wang (2008) who prove that the PC estimator remains consistent if the errors are heteroskedastic and autocorrelated[6]. PCs also remain a consistent estimator of the unknown factors in the presence of time-varying loadings, provided that the panel is sufficiently large with $N, T \to \infty$, see Bates et al. (2013). More recently, Mikkelsen, Hillebrand, and Urga (2019) prove average uniform consistency in t if $T/N^2 \to 0$ is satisfied. As anticipated earlier in Section 2.2.1, in this chapter I take an ex-post view of the factors and consider the information

[5]Formally, $E[\boldsymbol{e}^2_{r,\cdot t+h}] = \sigma^2$ for $i \in (1, n_r)$, $r \in (1, R)$, and $t, h \in (1, T)$.

[6]It is important to recognise that although the estimator of the latent factors in Breitung and Eickmeier (2014) is consistent in the case of heteroskedastic errors, asymptotic efficiency may still be improved by using a generalised least squares approach as in Breitung and Tenhofen (2011)

contained on the full sample T for their estimation - assuming constant loadings. Conditional on the factor space (with factors either estimated via PCA or constructed from observable characteristics as in the benchmark models), I then introduce time variation in the factor loadings via rolling window regressions, see Section 2.3 for further details. This procedure allow me to study in isolation the role of time-varying betas and compare the benchmark models with the candidate model.

The technical challenge of estimating the relevant factors in model (2.10) is to take into consideration the zero restrictions on the factor loadings matrix, and in this chapter I employ the the sequential least square (LS) algorithm of Breitung and Eickmeier (2014) for this purpose. To ensure that the iterative algorithm converges quickly to the global minimum, I initialise the algorithm with the first PC of all stocks (for the global factor), and the first PC of each group of stocks clustered by region (for the regional factor), as in Borghi et al. (2018). The main difference of my factor extraction procedure with respect to Borghi et al. (2018) and Breitung and Eickmeier (2014) is the normalisation that I impose on the estimated factors to ensure that they can be effectively compared with the benchmark factors in terms of pricing ability.

The objective function in (2.11) depends on the means of the test assets, $\bar{r}_i$, as well as on the factor risk premia, $\boldsymbol{\lambda}$, however as Lettau and Pelger (2020b) reports the data matrix is usually demeaned before PCA is applied. This implies that the standard PCA estimator does not consider information contained in the asset means, resulting in factors that have zero mean by construction. This setup would effectively rule out any possibility of assessing the pricing ability of the candidate factor model against the benchmark models. In this chapter, I propose an alternative procedure to rotate the PCA factors and recover the information in the asset means. In particular, I firstly apply PCA to the standardised matrix of stock returns, and extract the unit-variance zero-premia global and local factors. The factors are estimated consistently up to some arbitrary rotation, and as in Breitung and Eickmeier (2014) I impose a positive correlation between the latent factors and the stock market index proxies constructed as the equally-weighted portfolios made of the selected stocks (full universe for the global factor, and region-specific equities for the regional drivers). I then estimate the ex-post full-sample risk premia of these portfolios, and require the factors to have mean equal to the estimated quantities (see IR3 below)[7]. This procedure maps the regional factors to local stock market indeces, and the global factor to the stock market factor in the global CAPM model in equation (2.6), and allows me to resolve the rotation indeterminancy of the PCA factors while easing their economic interpretation.

[7] Applying PCA to a matrix of stock returns suggests that the first PC is highly correlated to the $1/N$ portfolio formed with the relevant tickers.

In summary, I impose the following idenfitying conditions to obtain a unique solution to the minimisation problem in (2.11):

IR1 $T^{-1}\sum_{t=1}^{T}(G_{k,t})^2 = 1$ for all $k \in K^{glob}$, and $T^{-1}\sum_{t=1}^{T}(R_{r,t})^2 = 1$ for all $r \in R$. Normalising the latent factors to have unit length allows me to effectively compare them.

IR2 $T^{-1}\sum_{t=1}^{T} R_{r,t}^{\top}\ G_{k,t} = 0$ for all $r \in R$ and $k \in K^{glob}$. This ensures that local factors are orthogonal to the global ones.

IR3 $\sum_{t=1}^{T} R_{r,t}^{\top}\ S_{r,t} > 0 \ \wedge \ E[R_r] = E[S_r]$, where $S_{r,t}$ is the return on the equally-weighted portfolio made of the relevant tickers in group/cluster r. This condition identifies the sign of the local factors by imposing a positive correlation with the most important source of systematic variation of the relevant group, and normalises the local factors to have mean equal to the one of the equity index proxy. The same applies to the global factor, considering the full stock universe.

Note that I do not need to assume orthogonality of regional factors as in standard factor analysis, $T^{-1}\sum_{t=1}^{T} R_{r,t}^{\top}\ R_{r,t} = I_R$, thus local drivers can be correlated with one other. This suggests that the model-implied co-movements structure of returns is determined only by the covariance between the local shocks, being orthogonal to the global universe-wide drivers. Refer to Section 2.4.2 for the results on the factor extraction procedure.

2.3 Forecasting Methodology

In this section I describe the methodology that I use to forecast future stock returns based on the models outlined earlier. In Section 2.3.1 I provide an in-depth discussion of the rolling least squares estimator of conditional betas which drives the dynamics in my model. Firstly, I review the choices that are made in the literature on the selection of the estimation window size W, the sampling frequency, and the forecasting horizon h, and secondly I describe the MSFE-based optimal window selection criteria of Inoue, Jin, and Rossi (2017). In Section 2.3.2 I report the statistical performance measures that I use to evaluate the in-sample and out-of-sample performance of the conditional factor models.

2.3.1 Rolling Least Squares

Conditional on the ex-post factor realisations for the models outlined in Section 2.2.2, I introduce the time-variation in the factor loadings via rolling least squares

$$\hat{\beta}_{i,t^*} = \left(\sum_{s=t^*-W}^{t^*} \hat{f}_s^{\,2} \right)^{-1} \left(\sum_{s=t^*-W}^{t^*} \hat{f}_s \, r_{i,s} \right) \tag{2.12}$$

$$t^* = 1, ..., T_W,$$

where $\hat{f}_t$ is the time-t realisation of a generic factor $k \in K$ for each model (i.e. the MKT model, the FF3 model, the FF5 model and the Regional model, see Section 2.2.2), and $\hat{\beta}_{i,t^*}$ is the respective loading estimated in period t^*, which comprises W observations at a given frequency.

It is important to recognise that the behavior of the rolling OLS estimator is driven by the choice of the parameter W, which in turns depends on the sampling frequency of the data and the forecasting horizon. As Robertson (2018) reports, the choice of W is intrinsically related to the assumptions that I am prepared to make on the dynamics of the beta parameters. If I am interested in forecasting betas over short horizons with data sampled at high frequency (intra-daily, daily, and weekly) then the rolling OLS approach is a valid solution. This is because in the out-of-sample rolling OLS framework, conditional on the appropriate choice of W, I am implicitly assuming that the next-period betas are consistently estimated by the current-period betas, see equation (2.5). On the other hand, if I am interested in longer-term forecasts I have to take into account the possibility of a mean-reverting behavior in the beta parameters, and this is similar to the modelling approach that I follow in chapter 1 where I explicitly define the equations for the dynamics of the beta process.

In this chapter, I a adopt a non-parametric approach and study the behavior of the rolling OLS estimator for various choices of W. In fact, similar to the techniques found in the literature on realised volatility[8], I model $\boldsymbol{\beta}_{t^*}$ as a function of the estimation window alone and treat the OLS estimator as being locally (within-window) constant, where the estimation window plays the role of the bandwith.

Crucially, the amount of local information on which the estimator depends increases suitably as the sample size W increases. Standard OLS theory assumes that the factor loadings remain constant in the sample size T, and that the variance of the beta estimator decreases with T. The idea is that, if beta is constant within each window, a simple OLS regression would produce an unbiased estimate of the true conditional beta. However in practice the assumption of constant loadings is difficult to defend especially when T comprises several years of data. Including too distant observations in the estimation of current-period betas results in estimates that are not representative of the present economic conditions, and possibly mis-specified due to the presence of structural breaks in the series. On the other hand, including few recent data points ease the economic interpretation of the (short-term) estimates, but necessarily increases the variance of the OLS estimator. In this chapter, I let the sample size for the estimation of the beta parameters vary by including as little as 26 observations (half-year window), up to 520 data points (five-year window). To isolate the role of W in shaping the dynamics of the factor loadings, I fix the sampling frequency to weekly and the forecasting horizon to one-period ahead, $h = 1$.

Individual Optimal Window

In the individual window selection approach, I estimate the optimal window size W_i for each stock following the methodology in Inoue, Jin, and Rossi (2017), which draws from Pesaran and Timmermann (2007). The optimal window criterion in Inoue, Jin, and Rossi (2017) is achieved by minimising the conditional MSFE at the end of the sample, $t = T$

$$\underset{\hat{W}_i}{\operatorname{argmin}} \quad E_T[\ (r_{i,T+h} - \boldsymbol{\beta}_{i,T}^{\top}\ \boldsymbol{f}_{T+h})^2\ ] \tag{2.13}$$

with $\boldsymbol{f}_{T+h}$ being the K-dimensional vector of factor realisation at time T (excluding constant), and $\boldsymbol{\beta}_{i,T}$ the respective time-varying factor sensitivities considering information up to time $t = T$. I estimate the time-varying betas via rolling least squares regressions using the last W_i observations, and replace the unknown parameter $\boldsymbol{\beta}_{i,T}$ with the OLS estimate $\hat{\boldsymbol{\beta}}_{i,W}$

$$\underset{\hat{W}_i}{\operatorname{argmin}} \quad E_T[\ (\hat{\boldsymbol{\beta}}_{i,W} - \boldsymbol{\beta}_{i,T})^{\top}\ \boldsymbol{f}_{T+h}\ \boldsymbol{f}_{T+h}^{\top}\ (\hat{\boldsymbol{\beta}}_{i,W} - \boldsymbol{\beta}_{i,T})\ ] \tag{2.14}$$

Inoue, Jin, and Rossi (2017) suggest replacing the unknown $\boldsymbol{\beta}_{i,T}$ with the local linear estimate $\hat{\boldsymbol{\beta}}_{i,W_0}$ computed on a pilot window that considers the most recent W_0 observations. The pilot window is calculated according to the cross-validation method in Pesaran and Timmermann

[8] See for instance the seminal contribution of Andersen et al. (2006) who is one of the first to provide a theoretical framework to analyse the dynamics of conditional betas estimated using non-parametric techniques.

(2007), which is originally designed for forecasting models conditioned on structural breaks in a discrete settings. On the other hand, the framework of Inoue, Jin, and Rossi (2017) assumes a smooth variation of the parameters over time, concordant with the assumptions underpinning the rolling OLS estimator of the factor betas[9].

Theorem 2 in Inoue, Jin, and Rossi (2017) provides the conditions for the asymptotic optimality of their MSFE criterion, so that the error introduced by replacing $\boldsymbol{\beta}_{i,T}$ with the sample counterpart $\hat{\boldsymbol{\beta}}_{i,W_0}$ is negligible. This allows me to quantify the bias induced by the rolling OLS estimator using the most recent W observations. In this framework, the minimisation problem is independent of the unknown parameter $\boldsymbol{\beta}_{i,T}$, replaced by the pilot window estimate $\hat{\boldsymbol{\beta}}_{i,W_0}$, and of the unknown error $e_{i,T+h}$, which does not enter the equation due to the property of conditional expectations. The feasible criterion reads

$$\underset{\hat{W}_i}{\text{argmin}} \quad E_T[\ (\hat{\boldsymbol{\beta}}_{i,W} - \hat{\boldsymbol{\beta}}_{i,W_0})^\top\ \boldsymbol{f}_{T+h}\ \boldsymbol{f}_{T+h}^\top\ (\hat{\boldsymbol{\beta}}_{i,W} - \hat{\boldsymbol{\beta}}_{i,W_0})\ ] \tag{2.15}$$

with $\hat{W}_i$ being the estimated optimal window for stock i. I report the results on the estimated optimal window in Section 2.5.2.

2.3.2 Performance Measures

I document the performance of the factor models with respect to three main features: describing the contemporaneous variation in stock returns, predicting future stock return patterns at horizon h, and predicting future stock returns based on the model's conditional expected return decomposition. For this purpose I define the following error measures that that draw from Kelly, Palhares, and Pruitt (2021):

- In-Sample error: the measure quantifies a model's success in describing contemporaneous stock excess returns based on the conditional betas estimated in period t^* and ex-post contemporaneous factor realisations, $\hat{\boldsymbol{f}}_t$.

$$\text{IS error}_{i,t^*} = r_{i,t^*} - \hat{\boldsymbol{\beta}}_{i,t^*}^\top\ \hat{\boldsymbol{f}}_{t^*} \tag{2.16}$$

- Out-of-Sample error: the measures quantifies a model's success in predicting future excess return patterns based on the conditional betas estimated in period t^*, conditioning on future ex-post factor realisations, $\hat{\boldsymbol{f}}_{t+h}$. This metric isolates the ability of the time-varying betas in forecasting excess stock returns at horizon h.

$$\text{OOS error}_{i,t^*} = r_{i,t^*+h} - \hat{\boldsymbol{\beta}}_{i,t^*}^\top\ \hat{\boldsymbol{f}}_{t^*+h} \tag{2.17}$$

[9]Pesaran and Timmermann (2007) consider a discrete breaks framework to accommodate recursive least squares regressions. They develop a suite of alternative window selection criteria based on a combination of rolling- and recursive- OLS (out-of-sample) forecasts.

- Conditional error: the measure is inspired by the conditional MSFE criterion in Inoue, Jin, and Rossi (2017), which is best suited to describe the model's ability in predicting future excess returns with factor betas estimated in period t^*, conditioning on contemporaneous ex-post factor realisations, $\hat{\boldsymbol{f}}_t$.

$$\text{COND error}_{i,t^*} = r_{i,t^*+h} - \hat{\boldsymbol{\beta}}_{i,t^*}^{\top} \hat{\boldsymbol{f}}_{t^*} \tag{2.18}$$

- Predictive error: the measure quantifies a model's ability in predicting future excess returns based on the conditional expected return process, $E_{t^*}[\boldsymbol{r}_{t^*+h}] = \boldsymbol{\beta}_{t^*}^{\top} \boldsymbol{\lambda}$. In my framework, the time-variation in expected returns is driven by the conditional factor exposures estimated in period t^*, with static ex-post factor risk premia, $\hat{\lambda}$.

$$\text{PRED error}_{i,t^*} = r_{i,t^*+h} - \hat{\boldsymbol{\beta}}_{i,t^*}^{\top} \hat{\boldsymbol{\lambda}} \tag{2.19}$$

The error measures defined above are time- and asset- specific, and in order to study the models' performance in the cross-section of asset returns I consider three classes of objective functions that aggregate the results differently:

- Mean-Squared Error: I construct mean-squared error measures for the quantities defined above by taking T_W-averages of the asset- specific squared errors, and then N-averages:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T_W}\sum_{t^*=1}^{T_W} \text{error}_{i,t^*}^2\right) \tag{2.20}$$

- Mean-Absolute Error: I calculate the deviation of the model-implied common component from the realised return in absolute value, and similarly to MSE I firstly take T_w-averages and then N-averages:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T_W}\sum_{t^*=1}^{T_W} |\text{error}_{i,t^*}|\right) \tag{2.21}$$

- R^2: the suite of R^2-based measures that I adopt for performance evaluation is taken from Gu, Kelly, and Xiu (2020) and Kelly, Palhares, and Pruitt (2021). The intuition behind the metric is to quantify the model's explained panel covariation with respect to the total variance proxied by the squared-return process.

$$R^2 = 1 - \frac{\sum_{i,t^*} \text{error}_{i,t^*}^2}{\sum_{i,t^*} r_{i,t^*}^2.} \tag{2.22}$$

The notation $\sum_{i,t^*}$ denotes aggregation over time and assets, with $r_{i,t^*}^2.$ being the squared realised return either at time t^*, for the case of IS error, or at time $t^* + h$ for the forward-looking error measures (OOS, COND, and PRED error). I follow Gu, Kelly, and Xiu (2020) and calculate the denominator in (2.22) as the sum of squared returns without demeaning.

They find that historical averages severely underperform a naive zero benchmark in out-of-sample fits using single-name stocks, resulting in estimates of the time-varying alphas that are so noisy to unduly inflate the R^2.

$$\text{error} \in \{\text{IS}, \text{OOS}, \text{COND}, \text{PRED}\}$$

The performance measures are calculated for all out-of-sample periods, $t^* = 1, ..., T_W$, and in the notation above, N may include equities belonging to a particular region, GICS industry group, or potentially the full cross-section.

2.4 Data and Factors

This section is split into two parts. In the first part, Section 2.4.1, I describe the data sources and report the summary statistics on the panel of weekly excess (log-) returns. In the second part, Section 2.4.2, I examine the factor space implied by the models outlined in Section 2.2, which include the three benchmark models (MKT, FF3 and FF5), as well as the candidate Regional model featuring latent factors. Further details on the data cleaning procedure as well as on the regional classification of my stock universe are found in Appendix B.3.

2.4.1 Data Description

The data source is Bloomberg. We download end-of-week (Friday) closing prices for all the tickers that entered the 48 national stock market indeces at some point during the sample period, January 2006 to end of May 2019. Historical member composition of the stock market indeces is available at monthly frequency. All prices are expressed in US dollars, are adjusted for splits, and include dividend payments. The emphasis in this chapter is on the regional classification of stocks, and we follow Bekaert et al. (2014) who partition their universe into six regions based on the company's domicile: North America, Latin America, Asia-Pacific, Western Europe, Emerging Europe, and Middle East & Africa.

Table 2.2 reports the countries that are considered for each region, and for each country it shows the total number of tickers that entered the index in the sample period, column #*Stocks*, the number of stocks with complete price series, at least one year of weekly data and no more than 12 consecutive missing observations, column #*Selected*, as well as the tickers that remain listed in the stock market indeces during the entire sample, column #*Full*. The total number of tickers with complete price series in the 13-year sample is $N = 1686$, and throughout this chapter I present the results only for this cross-section of stocks. In the rolling out-of-sample forecasting framework, the cross-sectional dimension is assumed to be time-varying, N_{t^*} with $t^* = 1, ..., T_W$, and reflects the composition of the index members during the W weeks that make up each window. However, to ease the comparison across models and window sizes I keep N fixed and present a portion of my results (which are available for a much larger cross-section of 2905 tickers, see column #*Selected*).

[Table 2.2 about here.]

I calculate week-on-week log-returns for all stocks in excess of the USD overnight index swap rate. Unless stated otherwise, throughout this chapter I refer to returns as the ratio of consecutive log-prices in excess of the risk free rate proxy. The latter is available in annualised terms, and I divide it by 52 (number of weeks in a fiscal year) to obtain an equivalent weekly figure. Once the 'raw' week-on-week excess returns are computed for the observed factors (local FF factors, MKT factor, and the S&P500 Financials) and stocks, I winsorise the individual time series at 99% level.

[Table 2.3 about here.]

Table 2.3 reports N-averages of the summary statistics based on the weekly stock returns, together with cross-sectional averages of market capitalisation, total assets and total debt, values in \$$B$. I present the results region- and industry-wise, using the 11 GICS industry groups for the latter[10]. Throughout the sample, the stocks in the Asia-Pacific region have returned an average of 12bp a week, followed by the MEA, and North America stocks with 10.78bp and 7.17bp respectively. The average excess return for the stocks in the Emerging Europe region is the lowest with -7.2bp. In Section 2.4.2 I describe how these figures determine the risk premia of the regional latent factors, which are assumed to have zero mean in standard PCA theory. Looking at the median estimates, the stocks with the highest median weekly return are found on average in developed markets such as North America and Western Europe, with 9.6bp and 7.5bp respectively. For the stocks listed in the Latin America, Asia-Pacific, and Emerging Europe region I report the highest volatility figures, which is expected given the higher risk-return profile of the tickers operating in emerging market regions. Based on the estimated intra-group dependencies, *Pearson*, I also find that the stocks belonging to developed markets such as Western Europe and North America are more integrated and show a high correlation level, while the lowest figures are found in the developing Asia-Pacific and MEA regions.

Balance sheet data show that the companies listed in the Western Euorpe and North America regions have highest market capitalisation, total assets and total debt. Similarly to Chapter 1, the issue of micro-cap stocks does not seem to be relevant in my universe given the a minimum average market capitalisation across regions of about 1.5 Billion USD, for MEA and Emerging Europe.

Industry-wise, my results show that the companies operating in sectors such as Energy and Financials are the ones with the highest intra-group correlation figures. Macro factors such as fluctuations in interest rates and commodity prices tend to have a wide-spread impact on the equities operating in these sectors. Excluding Financials, the companies in the Energy, Communication Services and Utilities sectors have highest assets and debt, in line with the level of infrastructures needed for the businesses.

2.4.2 Estimated Factor Space

I now examine the factor space implied by the specifications described in Section 2.2, and compare the benchmark observed factors of the MKT, FF3 and FF5 models against the set of observed and latent factors of the Regional model.

[10]From the industry-wise results I exclude 4 tickers with missing industry classification.

Benchmark Observed Factors

MKT model. The benchmark observed factor that I examine first is the equal-weighted average portfolio of all stocks in my universe, which serves as an ad-hoc proxy of the 'stock market factor' for the panel of international stock returns. Figure 2.1 shows the cumulative returns of the global $1/N$ portfolio and reports the estimated summary statistics based on the full sample. Mapping the cumulative factor returns to the financial and macro events listed in the economic calendar reveals that the MKT factor responds negatively to the GFC of 2008-2009, and the ESDC of 2011-2013, see the grey bands in panel 2.1a. A further set of events that appears to affect the market portfolio is the crash of the Chinese stock market in mid-2015, with cumulative factor returns turning negative for the first time after the GFC, together with the oil glut initiated at the end of 2014. From the end of 2016 factor returns enjoyed positive momentum, reaching a peak at the beginning of 2018. However, throughout 2018 the world stock markets suffered one of the worst years due to escalation of the trade war between China and the me, a major slowdown in global economic growth, and mounting concerns that the Federal Reserve was raising interest rates too quickly. Investing in the (ex-post) global $1/N$ portfolio since 2006 yields an average yearly return of 1%.

[Figure 2.1 about here.]

FF models. I now analyse the dynamics of factor returns for the FF3 model, featuring the MKT, SMB, and HML factors, as well as for the FF5 model with the additional robustness, RMW, and profitability, CMA, factors. Results are region specific, and details on the FF regional classification are found in Appendix B.3. Figure 2.2 shows the cumulative returns on the FF factors, and reports the estimated in-sample summary statistics based on the weekly excess (log-) returns.

[Figure 2.2 about here.]

North America. In the North America region, the MKT factor cumulative returns turn progressively negative at the beginning of 2009, recovering to pre-crisis levels only in early 2014. Returns on a portfolio of small-minus-large stocks together with the value-minus-growth portfolio enter a downward spiral during the GCF, albeit to a lower extent than the MKT factor. Investing into the RMW profitability portfolio would have hedged against the market-wide drop at the outset of the crisis in 2008, before following the bear market in the first half of 2009. On the other hand, the CMA investment value factor would have protected a me-based investor during the entire two-year period. During the ESDC, returns on the MKT and SMB factors turn progressively negative, while returns on the HML and CMA portfolios enjoy positive momentum. In the weeks prior the presidential me elections of 2016 returns on the MKT, SMB and RMW portfolios enter a downward phase, while the benefits of investing into stocks with high book-to-market ratios and stocks with a conservative investment approach (relative to the low B/M and aggressive investment style counterparts) increase significantly.

Overall, I find a negative relationship in the sample period between the cumulative returns of the MKT and the RMW portfolios (-0.82), confirming the role of the latter in hedging against market-wide fluctuations. On the other hand, the CMA investment value factor appears to be redundant with respect to the HML book-to-market value factor, with the correlation between cumulative factor returns yielding 0.83 in the sample period.

From the estimated risk premia, I find that the MKT factor's average return in the sample period is 8% annualised, which is in line with the figures commonly reported in the asset pricing literature. The RMW factor is the second-best factor in terms of average annualised returns, 3%, followed by CMA with a near-zero value, while the SMB and HML factors earn negative lambdas of -1% and -3% respectively. The variance of the factor is normalised to unity in the sample period, and I analyse the higher moments of the returns distribution to gauge the risk profile of each factor. The HML and CMA factors are the only ones showing evidence of positively skewed return distributions, with the MKT factor being the one with the most negative value among the five. Between the HML and CMA factors, the former appears to be the riskier in terms of estimated fourth moments. I find that the MKT and the HML factors show evidence of heavy tails in their return distribution, with a sample kurtosis near 4.5. I also notice that the returns of the SMB and the RMW factors are normally distributed in the sample period according to the JB test, p-values of 0.19 and 0.15 respectively.

Asia Pacific. Moving on to set of FF factors for the Asia-Pacific region, I find that the MKT and SMB factors share similar dynamics from the beginning of the sample up to mid-2012. From the outset of the Chinese stock market crash, I find that the value-minus-growth portfolio starts to track the performance of the MKT factor up to mid-2018, which marks the start of the trade war between the me and China. Similarly to the case of North America, the CMA factor returns are negatively correlated with the market-wide factor returns, with a Pearson correlation coefficient of -0.65 in the sample period. The CMA factor returns show a 0.41 Pearson correlation with the returns on the HML portfolio, and the RMA factor returns are negatively correlated with the HML portfolio gains, with an estimated in-sample coefficient of -0.55. Investing into the small-minus-big portfolio at the outset of the Chinese stock market crash would have effectively protected against the market-wide drop, as well as against the loss realised on the RMW portfolio. While the hedging potential of the RMW factor is strong in the North America region, for the equities listed in the Asia-Pacific region this is not apparent. During the period mid-2015 to mid-2016 the only portfolio outperforming the market is the small-minus-large stocks portfolio.

The estimated risk premium on the MKT is similar to the figures reported for the North America region, however the HML risk premium yields 4% annual, while it is negative in North America. This indicates that investing in value stocks would have outperformed an equivalent investment into growth stocks in the Asia-Pacific region. The returns on the HML factor are also the ones which are influenced the least among the FF3 factors by extreme swings, with an in-sample kurtosis of 3.6. The excess return on the RMW portfolio is on average 2% yearly, and both the HML and RMW portfolios show evidence of little deviation from the Gaussian pattern in terms of third and fourth moments. The CMA investment factor yields a premium similar to the HML factor, but has fatter tails in the returns distribution, indicating a higher risk profile. Similarly to the North America case, the CMA factor returns are positively skewed with a coefficient of 0.25.

Western Europe. The MKT factor dynamics for the equities listed in the Western Europe region are in line with what reported earlier for North America and Asia-Pacific. After the GCF, the MKT factor recovers much quicker in the European region than in North America, and at the outset of the ESDC it suffered more in the European region with respect to the others. Similarly to the case of North America and Asia-Pacific, the returns on the CMA investment factor increased in mid-2009, providing an hedge against market-wide fluctuations. During the ESDC the RMW profitability portfolio is the best performer among all five, and this is also true at the outset of the Chinese market slowdown in 2015 when the returns on the RMW factor consistently increase throughout the turmoil phase. I also find that in the weeks preceding the Brexit referendum the CMA and RMW cumulative factor returns increase, while the SMB and HML factors entere a downward phase.

In the Western Europe region the average market premium is 5% annualised, similar to the excess returned earned on the RMW profitability portfolio. In this region the value premium is negative with a coefficient of -2%, which indicates little benefits in investing into stocks with a high B/M ratio. In Western Europe, factors other than the MKT share a similar risk profile based on the estimated in-sample kurtosis of 3.4 (RMW) and 3.9 (SMB, HML, CMA). In Western Europe the MKT factor follows closely the performance of the HML portfolio in the sample period, with a correlation coefficient of 0.48 between returns and a similar figure for the cumulative returns. The RMW factor returns are negatively correlated with the HML returns with a Pearson coefficient of -0.72, and of -0.80 between the respective cumulative series.

Overview. To summarise, the key takeaways from the analysis of the region- and factor- specific results for the FF models are as follows.

- During the GCF, the returns on the MKT, SMB and HML portfolios turn progressively negative with respect to pre-2008 levels in all regions. In North America and Asia-Pacific, investing into the HML value portfolio would have provided a moderate level of protection against the market-wide drop at the beginning of 2009. Notably, the returns on the CMA investment factor increase significantly during this period across all regions. In the North America region, the CMA factor returns also enjoy positive momentum during the ESDC, following the performance of the HML factor (only in this region). At the outset of the Chinese stock market crash, in the Asia-Pacific region CMA and SMB factors are the only ones hedging against market-wide fluctuations. In-sample correlations between the MKT and CMA factor returns are all negative, respectively -0.26 for the North America, -0.56 for Asia-Pacific, and -0.27 for Western Europe. This finding suggests that investing into a portfolio that sorts stocks based on total asset growth (better performance for companies with low total asset growth, 'conservative', with respect to the the high- growth counterparts, 'aggressive') would have protected a me-based investor against the market-wide drop in the GFC, and to a lesser extent also in the ESDC.

- The RMW factor provides an hedge against market turmoil of the GFC in the North America region. In the midst of the ESDC, the returns on RMW factor also increase for the stocks domiciled in North America, together with those in the Western Europe region. In the latter, all factors other than the RMW enter a downward phase during this period. Overall, I find that for the developed markets, a portfolio that sorts stocks based on the company's operating profitability (better performance for companies with a higher profitability, 'robust', with respect to the low- profitability counterparts, 'weak') would have protected a me-based investor during the Debt crisis. This does not apply to the developing Asia-Pacific equity markets, where the RMW factor returns turn progressively negative during the Chinese stock market crash in early 2015.

- The FF factors are not orthogonal to each other, which implies that they do not effectively isolate different sources of systematic variation in the excess returns. Based on the estimated in-sample correlations, in the North America region the HML and CMA factors share very similar dynamics, with estimates of 0.83 between the cumulative returns, and 0.46 between the weekly returns. In Asia-Pacific the figures drop to 0.58 and 0.41 respectively, but are still the highest across all factor pairs for the region. In Western Europe, the correlation between the HML and CMA factor returns yields a modest 0.21 coefficient, the lowest figure across all regions. This finding is in line with Fama and French (2015) who report that with the addition of profitability and investment factors, the HML value factor becomes redundant for describing average return patterns.

Contrarily to the set of FF factors, the factor extraction procedure that I use to identify the financial factor and estimate the latent factors (for the Regional model of Borghi et al. (2018)) imposes orthogonality among the factors. By construction, the latent factors estimated via PCA are orthogonal to each other and maximize the share of explained covariation in the excess stock returns. In Section 2.4.2, I study how the identification of the 'stock market factor' with different equity indeces plays a crucial role in determining the information content of the estimated factor space. I present the factor extraction results in Section 2.4.2.

Observed Financial Factor

I identify the financial factor F_t of model (2.9) with the S&P500 Financials. My procedure examines the information content of the residual matrix of returns, after orthogonalisation against F_t, to determine which index is best suited to represent the broad 'stock market factor'. Figure 2.3 reports the results.

[Figure 2.3 about here.]

Panel 2.3a shows the time-series plot of the cumulative factor returns during the sample period. At the outset of the GFC, the Financials indeces have a steeper decline with respect to the others, while the MSCI World and the EW factors are the best performer given their diverse (global) composition. In panel 2.3b I report the ratio of consecutive eigenvalues estimated via PCA based on the $(T \times N)$ matrix of excess returns after orthogonalisation against each of the candidate factors. In standard PCA theory, the largest eigenvalues of the in-sample covariance matrix of asset returns are driven by the factors, which motivates the use of PCA for factor extraction purposes[11]. Traditional factor models as in Stock and Watson (2002) are based on the assumption that all factors are 'strong', in the sense that the strength of systematic factors is given by their corresponding eigenvalue. Following the interpretation of Lettau and Pelger (2020b), weak factors can be thought as either factors with only a weak effect on many assets, or factors with a strong effects on a few assets. In their paper, the authors analyse the empirical spectrum of the eigenvalues extracted from equity (portfolio) data, and find that the first eigenvalue of the sample covariance matrix is usually very large, while the second and third eigenvalues have only magnitude around twice or three times of the average of the residual spectrum (i.e. the eigenvalues of e). I adopt a similar approach to Lettau and Pelger (2020b) and report the ratio of consecutive eigenvalues estimated from the residuals of a regression of the excess returns matrix against the candidate equity indeces, see panel 2.3a.

I find that when I identify the financial factor with the S&P500 Financials, which tracks the performance of me Financials stocks, the evidence for the existence of a latent dominant factor in the residual matrix is strong, with a ratio between the first and second eigenvalues of 4.5. This is true to a lesser extent for the S&P500 index, which is confined to represent the performance of the regional me stock market despite including companies belonging to sectors other than Financials. When I move to the MSCI indeces, due to their global composition, the magnitude of the first eigenvalue drops to 1.5 to 2 times the one of the second characteristic root. My results suggest that the identification of the financial factor with the S&P500 Financials index eases the estimation of the global latent factor, which is considered to be a 'strong' factor affecting all stocks in the panel. For economic interpretation, this allows me to isolate sources of systematic variation in the stock returns that are not related (linearly) to the American financial

[11] In the equivalent static-loadings version of equation B.1, Appendix B.1, the eigenvalues of $cov(\boldsymbol{r}_{1:T})$ are driven by the relevant factors.

sector's performance. In the next section, I describe in detail the factor extraction procedure of the global and regional factors.

Latent Factors

In this section, I report the results of the factor extraction procedure for the global latent factor and the six regional factors featured in the Regional model. Figure 2.4 reports the results.

[Figure 2.4 about here.]

After orthogonalisation against the S&P500 Financials, I estimate the global latent factor via PCA. Panel 2.4a shows the time-series plot of the cumulative global factor returns during the sample period. The global factor dynamics are influenced by the financial and macro events that I consider in my economic calendar. During the GFC, global factor returns turn progressively negative from the second half of 2008 up to the beginning of new year, and by the end of 2009, the rebound in equity markets is significant for most of the regions considered. In Europe, the FTSE saw its biggest annual gain since 1997, rising 22% over the course of the year, Germany's DAX rose 23%, while France's CAC added 22%. In Asia, the Shanghai Composite jumped by a staggering 80% year-on-year, and the broad me stock market also enjoyed a strong performance. The global factor isolates sources of variation in the excess stock returns that are not linked to the performance of American financials stocks, thus given the broad composition of my universe the rally is in line with my expectation. At the outset of the ESDC, both the financial and global factors enter a downward phase, however the former rebounds much quicker than the latter during the two-year period from mid 2011 to the beginning of 2013. The most relevant drop in the global factor returns is experienced during the period mid-2015 to end of 2016, which coincides with a major slowdown in economic growth, the Chinese stock market crash and the oil glut. Towards June of the same year, the events linked to UK's exit from the EU system appear to influence negatively the performance of the global factor, while the returns on the financial factor enjoy positive momentum. Finally, during 2018 global factor returns dip amid the uncertainty related to the me-China trade war, a fear that continued interest rate increases could trigger a recession, and the slowest growth figures in China since the global financial crisis.

Moving on to the region-specific factors, panel 2.4b reports the cumulative returns of the six estimated local factors, and panel 2.4c shows the in-sample summary statistics. In-sample variances are normalised to one, and I refer to the sample kurtosis figures to gauge the risk profile of the factors. The factors with returns showing evidence of fat tails are the ones corresponding to to emerging market regions such as Latin America, Emerging Europe, MEA and Asia-Pacific, listed in descending order. The risk premia are estimated as the realised return on an equally-weighted portfolio made of all stocks in the relevant region, I plot the time-series of this portfolio which I refer to as 'index proxy' in figure 2.5. The local factors are estimated via PCA based on the matrix of orthogonal returns with respect to the financial and global factors, and are calculated as the first principal component of the region-specific matrix of (orthogonal) excess returns. In the same exhibits, I also calculate the rolling correlations between the factor and index proxy returns to interpret the results economically. The regional factor dynamics are related to the performance of the index proxies, and I analyse how their (linear) relationship changes over time.

[Figure 2.5 about here.]

North America. I start with panel 2.5a that reports the results for North America. During the GFC, the correlation between the factor and index proxy drops significantly to near zero territory, and even turns negative in the first half of 2009, indicating that the regional factor is effectively pricing a different source of systematic variation in the returns other than the market. Looking at the cumulative factor returns, I find that they turn progressively positive during this period, while the broad North American market factor experiences a significant drop. This behavior is similar to the one of the RMW factor for the same region and, across all FF factors, the North America regional factor returns has an in-sample correlation of 0.35 with the RMW returns, the highest across all factor pairs. From mid-2008 up to mid-2009, the North America regional factor and the RMW portfolio share very close dynamics, indicating that the estimated factor is effectively mimicking the performance of the robust-minus-week (profitability) factor. Similarly, at the outset of the ESDC the correlation between the factor and index proxy turns gradually negative, and the North America regional factor starts to track the returns on the HML portfolio. From the end of the ESDC up to the start of 2016, the regional factors performance follows the one of the index proxy before dropping consistently in the months leading to the US presidential elections, this represents the biggest drop in cumulative factor returns in the sample period.

Latin America. Panel 2.5b shows the results for the Latin America region, where I find that the in-sample correlation between the between the returns of the latent factor and index proxy is the highest across all regions, with 0.51. During the GFC, the returns on the equally-weighted portfolio made of all tickers in the regions turn progressively negative, while the factor returns are stable. The correlation between the two drops to a low of -0.01 in mid-2009, indicating that the behavior of the factor was not related to the one of the (regional) market in this period. From the end of GFC, the latent factor starts to mimic closely the performance of the equity index proxy, this can be seen from the short-term correlation estimates using one and two years of the most recent data. I highlight in red the dates corresponding to the oil price crisis in panel 2.5b, and I find that during this period the correlations between the factor and index proxy drop significantly. Latin America as a region is an important contributor to the world supply of petroleum [12], which is why I focus on the impact of the oil price turmoil in this region. I notice a diverging behavior of the index proxy and the regional factor, with

[12] Brazil is currently estimated to be among the top top countries for oil production, Argentina is at the 26th place and Peru at the 41th place

the returns on the latter increasing consistently during the period while the market index is affected by the drop in oil prices. This behavior is similar to the one reported for the CMA factor in the other regions, i.e. showing hedging potential against market-wide drops, however a direct comparison with the FF framework is not possible due to the limitation in the stock universe considered. See Appendix B.3 for further details.

Emerging Europe. In the Emerging Europe region, panel 2.5d, the correlation between the factor and the index proxy returns drops to negative territory at the end of 2009, a behavior which is similar to the one of Latin America. The post-crisis rebound in economic activity can be seen from the factor and index proxy returns enjoying positive momentum up to the start of the ESDC, when I find again that the index proxy and the factor de-correlate. This is also true during the oil price crisis, when the index returns enter a downward spiral and the factor returns increase consistently. The in-sample correlation between the cumulative returns of the Emerging Europe factor and the Western Europe-specific RMW factor (the 'closest' region against which I can compare my estimates) yields 0.73, indicating that a similar role of the latent factor for this region with respect to the profitability portfolio of FF5.

Western Europe. Panel 2.5c reports the results for Western Europe. During the ESDC, factor returns turn progressively positive after a serious drop in the series of cumulative returns at the outset of the crisis in late 2011. In the months that follow, the correlation between the proxy and the factor returns is the highest in the sample period, reaching 0.8 in late 2013. This indicates that the latent factor is tracking the performance of the broad stock market index for the region during this period. At the beginning of 2015, the factor starts to price a different source of systematic variation, which is not captured by the market performance. While the latter decrease from the highs of mid 2014, the regional factor returns enjoy positive momentum, reaching an high at the beginning of 2016. The correlations between the factor and the proxy drop by more than 0.7 in absolute value in the period 2015-2016. I also find that the factor returns are negatively influenced by the outcome of the Brexit referendum in the summer of 2016, and this is also true for the index proxy. In the weeks preceding the vote, the factor returns closely follow the performance of the stock market index.

Asia Pacific. I analyse the set of results related to the Asia-Pacific region in panel 2.5e. Similarly for all the other regions, during the GFC the correlation between the factor and the index proxy peaks and drops to near zero (and even negative) territory by the beginning of the following year. In the year 2009, the estimated factor rebounds quicker than the index proxy in the Asia-Pacific region, but enter a downward phase which continues up to the beginning of 2015. On the other hand, the post-crisis performance of the index proxy is positive throughout this period. During the Chinese stock market turmoil, the factor returns follow market dynamics and this can be seen from the short-term estimates of linear relationship between the two. After 2016, the factor and index proxy's trajectory are similar. Throughout the whole sample, the OLS estimate of linear relationship between the two is the third-highest across all regions, 0.44, indicating a strong mapping between the latent factor and the observed index proxy.

MEA. For the MEA region, panel 2.5f, the correlation between the index proxy and the factor is 0.86 during the sample. This case is peculiar to the composition of the region, which includes just one country and as such the factor extraction procedure yields a latent factor (first PC) that has a strong correlation with the broad stock market index for the region.

To complement the time-varying analysis on the relationship between the latent factors and the observed index proxies, I calculate the in-sample correlations between the regional index proxies and the latent factors. By construction, the regional factors can be correlated with one another and as such they shape the comovement structure implied by the Regional model[13]. Table 2.4 reports the results.

I observe on average high correlations among the returns on the $1/N$ portfolios, the most prominent being the one between the two developed markets in my sample, North America and Western Europe, with a in-sample coefficient of 0.89, highlighted in black in table 2.4. I also find a strong (linear) relationship between the returns of the equity indeces of Western Europe and the neighbouring Emerging Europe regions, 0.82, followed by the figure specific for the Latin and North America factors, 0.77. These values are highlighted in blue in table 2.4. On the other hand, I see little evidence of market integration between the equities belonging to the North America and MEA regions, with the correlation between the two being the lowest across all pairs, highlighted in red in the table reporting the estimates for the equity index proxies. In panel 2.4a I report the statistically significant Pearson correlation estimates for the regional factors[14]. I find that for North America, all correlations with respect to the other five regional factors are different from zero, the highest being the one with Western Europe, 0.24, and the lowest with MEA, -0.15. The level of commonality between the regional factor returns of the neighboring Latin and North America region is 0.08, the latter also includes Mexico, a country that has strong link with the economies in the southern American continent. I also find a negative correlation between the returns of the North America and Asia-Pacific factors, -0.06, although significant only at 10% level. A similar case is for the Western Euorpe region, which show statistically significant links with all the other regions, Latin America excluded. The ex-post factor returns of the Western Europe factor are negatively correlated with the Asia-Pacific returns and mildly correlated with the Emerging Europe region. The second-highest correlation estimates are the ones between emerging economies. For instance Emerging Europe and Latin America show a 0.13 coefficient, similarly Western Europe and MEA due to the proximity of Morocco to southern European countries such as Spain and France, and in general to its

[13]On the other hand, the financial and global factors are orthogonal to each other and to each of the local factors - see the sparsity conditions imposed on the block-diagonal matrix of the factor loadings in (2.10).

[14]This corresponds to the quantity $cov(f_{1:T})$ in equation (B.1).

special partnership status with the EU.

[Table 2.4 about here.]

2.5 Model Comparison

In this section, I compare the results of the rolling beta estimates using different statistical measures that examine the models' ability to explain and predict excess equity returns. I firstly present the results of OLS beta estimates on the full sample, Section 2.5.1, and then analyse the properties of the rolling least squares estimator in Section 2.5.2. I estimate the out-of-sample rolling betas considering the historical member composition of the national indeces, which effectively makes the dimension N time-varying. However, to ease the comparison across models, in what follows I report the results only for the 1686 unique tickers with complete price series in the 13-year sample considered[15]. Moreover, before estimation I winsorise the data (factors and single stock returns) at 99% level, and standardise them to have unit variance and zero mean in the sample period. Figures on the goodness of fit of the models, and on the relative statistical significance of the factor betas are not affected by this normalisation.

It should be noted that while one of the advantages of my setup is that I am able to isolate the relative error from the rolling estimation of time-varying betas for different window size, when I take the factor returns of the Regional model as given I am implicitly conditioning on future knowledge. The estimation of global and regional factors via PCA is in fact conducted using data on the full sample. The bias in the performance measures becomes relevant when I compare other models against the Regional model for a given window size. However when I compare the performance of Regional model across window sizes I am conditioning on the same set of factor returns at time t^* and the relative error in the factors be considered constant. The results for the MKT model and the FF models are not influenced by the look-ahead bias since the factors are observed, and not estimated.

2.5.1 Static Loadings

I analyse the results of the least square beta estimates on the full sample by model specification and report aggregate statistical measures across regions and industries. For each factor, I report the average beta magnitude, and the average absolute value of the t-statistic, and for each region and industry I show the N-average R^2 for the stocks in the relevant group. Table 2.5 shows the results for the single-factor MKT model, panel 2.5b, the three-factor Regional model, panel 2.5a, and panels 2.5c and 2.5d for the FF models.

[Table 2.5 about here.]

MKT model. I start by analysing the most parsimonious model. With a single factor constructed as the $1/N$ portfolio of all the stocks in the universe, the static-loadings MKT model

[15] As I anticipated earlier, the FF factors are region-specific, and are constructed using a classification criterion that closely matches the one in Bekaert et al. (2014). In their original work, FF3 and FF5 limit their analysis to three distinct world regions. I use their factors as benchmarks in my analysis, and as such I can only report the results for the cross-section of companies belonging to the North America, Asia-Pacific, and Western Europe regions. The number of stocks considered is reduced from 1686 to 1232. Details on the country-region classification against which I compare are found in Appendix B.3. Overall the grouping in FF is very similar to the one I consider.

is able to explain a substantial portion of the contemporaneous returns in the developed European region, $R^2 = 37\%$. The estimated beta parameter is on average 20 times larger than its corresponding error, t-stat = 20.6, and the average beta magnitude is substantially higher in this region, $\hat{\beta}^{mkt} = 0.595$. In terms of average goodness of fit, I find that the MKT model performs best in developed markets such as Western Europe and North America compared to all other regions. Sector-wise, I find that nearly 30% of stock returns comovements are explained by exposures to the market factor in the Financials, Energy and Materials sectors. I also record the highest beta magnitude and t-stats for these groups. The performance of the companies operating in these sectors tend to follow the business cycle on the economy. For instance, Financials stocks are negatively affected by a low-interest rate environment, a situation typical of recessionary phases, while their valuation increases with rising interest rates. Similarly, demand for energy and basic materials tend to be cyclical, which may help explain why I see a better performance of the single-factor model in these sectors. As I anticipated in Section 2.4.2, the equally-weighted portfolio of all stocks (EW) appears to be among the best-suited to represent wide cross-sectional variation of stock returns for my panel. Figure 2.3b shows that the ratio of consecutive eigenvalues estimated from the data matrix after orthognalisation against the $1/N$ portfolios is one of the lowest, together with the MSCI World index. Due to their composition, these indeces are less prone to fluctuations in regional market performances, and are highly correlated with the first PC of all stocks, which by definition maximises the share of explained co-variation in the stock returns.

Regional model. The candidate model of Borghi et al. (2018) features three factors, two of which are latent. Stocks with a relatively higher sensibility to shocks to the financial factor, proxied by the S&P500 Financials, are found in regions such as North America, and Western Europe. I estimate an average factor beta of $\hat{\beta}^{fin} = 0.485$, and 0.456 respectively. The average t-stats for the loadings on these factors are the highest across all regions considered, 17.4 for North America, and 16.9 for Western Europe. For the three countries that make up the North America region (Canada, USA, and Mexico) I find that only 31 of the 231 companies are classified as Financials using the GICS convention, effectively ruling out the possibility of my results being explained by the peculiar industry composition of the region. A similar result is found in Western Europe as well, in which 16% of all stocks analysed operate in the financial sector. For the other regions, I find considerable improvements in the average goodness of fit (R^2) with respect to the single-factor model, the biggest being the one in the MEA region of about 20%, followed by Latin America and Asia-Pacific with a 10% increase, and finally Emerging Europe with roughly 8%. Across all regions, the loadings on the observed financial factor for the developing regions listed above are on average the lowest. The latent global factor dominates the other two factors in terms of average beta mangnitude and t-stat in these regions, similarly the regional factor becomes a substantial driver of stock return variation and complements the role of the observed financial factor. In fact, the financial factor isolates shocks peculiar to the US financial sector, while the local factor tends to follow the aggregate performance of the national equity indeces that make up each region. The role of the regional factor is similar to the one of the financial factor in the Latin America region, this can be seen from the estimated beta magnitudes and t-stats of respectively $\hat{\beta}^{reg} = 0.239$ and t-stat= 8.2, and $\hat{\beta}^{fin} = 0.288$ and t-stat= 9.5. The equities domiciled in this region share close trade links with the neighbouring North America, and are in general influenced by the performance of the dollar-denominated stock market. However, there are also peculiar characteristics for the tickers operating in this region that call for the addition of a factor able to represent region-wide cross-sectional differences. In Asia-Pacific I find a similar case as the one for Latin America, the stock returns load with an average $\hat{\beta}^{fin} = 0.227$ coefficient on the financial factor, and $\hat{\beta}^{reg} = 0.182$ on the regional factor.

Industry-wise, I find the biggest increases in the R^2 with respect to the benchmark MKT model in the Utilities sectors, with a delta of 12% on average, followed by Financials and Industrials, with an increase of 10%, and by the Materials sector with a 9% gain. Interestingly, the biggest gain in R^2 across all groups is the one in the Utilities sector, which is commonly referred to as a 'defensive sector' because of its anti-cyclical behavior. Companies in this sector tend to offer investors with a stable and consistent dividend flow, and are also less prone to price volatility with respect to other industry groups. This is also confirmed by the average estimated standard deviation of the returns in table 2.3a being the lowest across all regions and industries considered. The case for Utilities is peculiar because it is one of the few groups in which the global and regional factor loadings play a bigger role than the observed factors, in terms of average beta magnitude and t-stat. In fact, the betas on the global factor for Utilities are the highest in magnitude with a $\hat{\beta}^{glob} = 0.314$ coefficient, followed by the regional and financial factor with $\hat{\beta}^{reg} = 0.265$ and $\hat{\beta}^{fin} = 0.264$. This ranking is analogous considering t-stats.

FF models. Across all groups of stocks, the FF3 model explains a substantial portion of the contemporaneous variation in the returns of the stocks domiciled in the North America and Western Europe regions, $R^2 = 0.34$ and 0.41, and of stocks operating in industries such as Financials, Industrials, and Materials, with an average goodness of fit of $R^2 = 0.42$, 0.33, and 0.32 respectively. Compared to the MKT model, the gain in performance in the Asia-Pacific region is negligible, with about a 1% increase in the R^2, indicating that the 'local' FF3 model is not able to accurately explain the variation in stock returns for this region. Industry-wise, the R^2s of the FF3 model are on average 5% bigger than the corresponding figures for the MKT model, the biggest gain is found for Financial stocks, with a delta of 11%. For the latter group I find that the estimated betas on the size factor are on average statistically insignificant at 5% level, with a t-stat of 1.9, and that average beta magnitude of the value factor $\hat{\beta}^{HML} = 0.126$ dwarf the beta estimates across all other groups. In fact, across all groups with the exception of the value factor for Financials, the loads on the SMB and HML factors are at least one order of magnitude lower with respect to the ones for the MKT factor. Average statistical significance of the estimated parameters is also one order of magnitude lower. The highest figures for the SMB factor loadings are the ones corresponding to the Materials and Consumer Staples sectors, t-stat= 2.8, followed by Western Europe, Health Care, and Industrials with a t-stat on the size factor of about 2.7. For the HML factor with the exception of Financials, the highest t-stats are the ones in Health Care, t-stat= 2.9, and Western Europe. I find a statistically significant average negative loading of $\hat{\beta}^{SMB} = -0.015$ on the size factor in North America, which suggests that the benefits of investing into small-cap stocks with respect to the highly capitalised counterparts are slim in this region. Negative loadings on the SMB factor are also found in typical 'defensive' sectors such as Consumer Staples, and Health Care, which tend to protect an investor against downside risk and offer a lower volatility profile than other industries. In table 2.3a I record the lowest standard deviation of weekly returns for these sectors. Loadings on the value factor for sectors such as Consumer Staples and Health Care are also negative.

Comparing the FF3 model with FF5 suggests that the addition of the profitability and investment factors does not bring substantial gains in terms of average goodness of fit. Across all groups, the improvement in R^2 is about 1 to 2%. The RMW factor appears to be a statistical significant driver of stock returns only for few groups, namely North America with a t-stat of 2.2 and Financials with t-stat= 3. The CMA factor yields t-stats greater than 2 only for the developed North America and Western Europe regions, and the Energy and Financials sectors. For these regions, the loadings on the CMA factor, albeit relatively small compared to the HML factor, are all negative. This result builds on the empirical evidence presented earlier on the anti-cyclical behavior of the CMA factor returns during market turmoils. I corroborate the evidence that the HML factor is redundant with the addition of the RMW and CMA factors. Across all groups, the decline in the magnitude of the estimated beta parameters with respect to their standard errors (t-stats) is apparent for the HML factor, when moving from the FF3 to the FF5 specification. In the FF5 specification the average t-stat on the HML factor for Financials drops significantly from 5.2 (the highest across all groups for FF3) to 3.1. The North American region is one of the few exception, together with Energy stocks, with $\hat{\beta}^{HML}$ and its relative t-stat increasing from FF3 to FF5. The loadings on the value factor for these group of stocks are about 1/5th of the magnitude of the corresponding MKT factor laoads, and are the highest across groups for all non-market factors (SMB, HML, RMW and CMA).

Overall, I find that the FF models' performance decreases considerably in the Asian region with respect to the others by a factor of 15 to 20%, in terms of average R^2. My results are in line with Kubota and Takehara (2018) who test the FF5 model in Japan and find little relation of average returns with profitability and investment patterns, suggesting that the model cannot be an adequate benchmark for the country in the sample period from 1978 to 2014. Among the 484 stocks classified in the Asia-Pacific region, almost 1/4th of the stocks are domiciled in Japan, followed by 90 domiciled in China. In Asia-Pacific, the addition of the RMW and CMA factors makes the HML factor redundant (average t-stat), while the HML factor remains statistically significant in the FF3 specification. My results are also related to Guo et al. (2017) who test the FF5 model for the Chinese stock market and find strong profitability patterns in average returns, and weak investment patterns. They corroborate the evidence that the significance of the HML factor is weakened in the more recent part of their sample (1997 to 2014), although they find that the value factor is non-redundant out-of-sample.

In the next section, I examine the results on the out-of-sample analysis on the beta loadings for the models and I complement the static loadings analysis.

2.5.2 Dynamic Loadings

I now examine the dynamics of the estimated time-varying factor betas for varying window size. I firstly compare the results from a statistical viewpoint by referring to the suite of performance measures defined in Section 2.3.2, and then interpret the estimates economically.

Performance Evaluation

[Figure 2.6 about here.]

MKT model. I start by analysing the benchmark MKT model. Figure 2.6 shows N-averages of the performance measures for the rolling OLS betas estimated with varying window sizes. Based on the in-sample estimates, I find that a window made of the most recent five-year observations gives the lowest MSE and MAE, IS MSE $=$ 22.9%% and IS MAE $=$ 2.9%, while the R^2-based measure of Kelly, Palhares, and Pruitt (2021) is maximised using a short window made of as little as 26 weekly observations, IS $R^2 = 26.8\%$. The out-of-sample estimates suggest that in order to predict future return patterns a window size of two years maximises the out-of-sample R^2, OOS $R^2 = 20.1\%$ in panel 2.6c. However, the MSE and MAE measures indicate that a much longer window is needed to minimise the out-of-sample errors, OOS MSE $=$ 23.2%% with a five-year window, and OOS MAE $\approx$ 2.9% for the five- and ten-year windows. Based on the MSE and MAE functions, I find that the five-year window pattern finds strong support from the data. Contrarily, if I analyse the estimates from the R^2 measures, a clear pattern emerges. If I am interested in explaining contemporaneous-return variation, a short-window approach is to be recommended, and this is very much in line with the results of Lewellen and Nagel (2006). In a short-horizon forecasting settings (with $h = 1$ week), I find support for the existence of a trade-off between length of the estimation window and predictive power, based solely on the properties of time-varying rolling betas. In fact, panel 2.6c suggests that the OOS R^2s are increasing in window size up to the two-year mark, before deteriorating considerably with a longer estimation window. Based on the R^2 measures, the choice of the window length alone accounts for about $\pm 10\%$ of the model's performance (the difference between the R^2s for the two- and ten-year windows).

Based on the summary statistics of the estimated MKT betas, panel 2.6d, I find that the magnitude of the loadings on the MKT factors is increasing with the window size, and that their variance decreases with the sample size. This result is partly in line with standard OLS theory, i.e. the variance of the beta OLS estimator is in fact decreasing in T as I anticipated in Section 2.3.1, however the average beta value should not vary considerably. With steadily increasing beta magnitude for longer sample sizes, the t-stats are also increasingly higher and almost one order of magnitude apart at the extremes (half- and ten-year windows). Choosing the window size based on the 'strength' of the beta signal leads to the use of an-ever increasing window size, however the use of a long sample size can sometimes lead to estimates that are not representative of current market conditions, due to the presence of breaks in the times series. In panel 2.6d I also report the figures based on T_W-averages of the R^2s from the time-series (rolling) regressions at varying window size[16]. The results are mixed and do not suggest that the existence of a trade-off between average goodness of fit of the regressions (in the estimation phase) and window length. I find support for the use of a short half-year window (second-best alternative), but a ten-year window performs better. The average R^2 from the short-window regressions is 25.6% (comparable to the five-year benchmark, $R^2 = 25.4\%$), while the one for the ten-year regressions is the highest at 26.6%.

I now analyse the role of time-varying MKT betas in predicting future return patterns based on the conditional and predictive error measures defined in Section 2.3.2. For the calculation of the IS and OOS measures, I condition on the contemporaneous (time t^*) and future (time $t^* + h$) factor realisations respectively. I do so to isolate the channel through which the estimation error of the time-varying betas influences the model's ability to explain or predict future return patterns. In the conditional error measures (COND), I take into account the possibility of forecasting future return patterns based on the time-t^* factor realisations (ex-post), and on the time-t^* betas.

The difference between a model's OOS and COND measures quantifies the error from assuming that no change occurs in the evolution of the relevant factors from time t^* to $t^* + h$. This difference is higher for a short window size, accounting for a 57% change in MSEs in relative terms, while it is minimised when a long window is employed, 24% change[17]. This result suggests that even in a short-horizon forecasting framework ($h = 1$ week), the assumption of constant factor realisations from period t^* to $t^* + h$ is responsible for a substantial portion of the model's forecasting error, and this is exacerbated with the use of a short estimation sample. In a conditional out-of-sample forecasting framework as in Inoue, Jin, and Rossi (2017), I expect

[16]Note that this measure does not quantify the forecasting ability of time-varying betas, rather the average goodness of fit of the time-series rolling regressions.

[17]The difference between the COND and OOS MSEs for the half-year window, 'delta', is about 15%%, considering the MAE measures it is approximately 1%. Compared to the OOS measures, the COND measures are inflated by a factor of 57% for the MSE, and by 30% for the MAE (half-year window case). Similarly for the ten-year window, the COND measures are greater than the OOS counterparts by a factor of 24% for the mean-squared errors, and 20% for the mean-absolute errors.

their objective function (COND MSE) to be minimised with the use of a longer estimation sample. Based on the COND MSE and MAE measures, I find that a ten-year window minimises the out-of-sample conditional error measure with COND MSE= 26.8%% and COND MAE= 3.2%. In relative terms, this represents a 28% decrease in the COND MSEs, and a 20% decrease in the COND MAEs, due solely to the different window size. Moving on the PRED measures, I find that based on the expected return process the MKT model's forecasting performance is higher the longer the sample. The MSE and MAE are minimised with the use of a ten-year window. Based on the difference between PRED and OOS measures (with window fixed), I find that using the model's pricing equation is always recommended over the expected return equation for forecasting purposes. Interestingly, I find that using the model's expected return process with time-varying betas estimated every ten-year provides a similar forecasting performance as using the model's pricing equation with a short estimation window.

[Figure 2.7 about here.]

Regional model. I now analyse the performance of the candidate model. Based on the in-sample measures, I find strong support for the use of a short estimation window with as little as 26 observations. The MSE is minimised with a half-year estimation window, IS MSE = 19.5%%, while the MAE is the third lowest and comparable to the five-year figure (the MAE is minimised with a ten-year window). Contrarily to the MKT model, the IS R^2 agrees with the MSE and hints at the use of a short half-year estimation to explain contemporaneous return patterns. The IS R^2 is highest across windows with $\approx$ 42%. The out-of-sample estimates suggest the existence of a pattern between average explained variation and window size, and similarly to the MKT model I find that the OOS R^2 is increasing in window size up to the two-year mark, before decreasing sharply with longer time samples. The OOS MSE indicate that a five-year window is the best suited, OOS MSE= 21.1%%, while the OOS MAE is minimised with the longest window. All in all, I find strong support for the use of a short estimation window in a contemporaneous-equations setting (in-sample), and for the two-year window in a forecasting exercise (out-of-sample). Based on the R^2 measures, the choice of the window length alone accounts for about $\pm$11% of the model's out-of-sample forecasting performance (the difference between the OOS R^2s for the two- and ten-year windows, panel 2.7c).

From the summary statistics of the estimated factor betas, panel 2.7d, I find that the magnitude of the loadings on all factors tend to increase with the sample size, and this is true especially for the loadings on the observed factor. For the latent factors, I notice that the average beta value across all stocks does not show evidence of a positive (monotonic) relation with window size. On the other hand, I corroborate the evidence that the variance of the beta OLS estimator is decreasing in T, and this is true for the loadings on the observed and latent factors. Looking at the summary statistics, I find that although the half-year window size provides the best forecasting performance according to the R^2, the average t-stats of the betas indicate a slim margin of significance. For instance, the regional factor appears to be on average not statistically significant at 5% level with a short estimation window. The t-stats of the estimated loadings on the global factors fluctuate approximately around ≈ 1.9. In panel 2.7d I report the R^2 figures based on T_W-averages of the R^2s from the time-series (rolling) regressions at varying window size, and I find support for the existence of a trade-off between average goodness of fit of the regressions (in the estimation phase) and window length. The estimated betas are more volatile with a short estimation window, causing the t-stat to decline. In this context I find that the choosing the optimal window based on the 'strength' of the beta signal leads to the use of an-ever increasing window size, which contradicts the performance measures. The average R^2 from the short-window regressions is 40%, which is considerably higher than all the other window size alternatives, and it is comparable in absolute value to the in-sample R^2-based error measure of Kelly, Palhares, and Pruitt (2021).

I analyse the role of time-varying betas on the financial, global and regional factors in forecasting future return patterns based on the conditional and predictive error measures. Taking an ex-post view of the factors (estimated on the full sample), the OOS measure uses knowledge of the next-period 'exact' factor realisation (time $t^* + h$) to quantify the error in the beta estimates. Contrarily, if I condition on the current-period factor realisation (time t^*), the COND measure gives me a comparable estimate to assess the error that occurs in assuming time-invariant factor realisations.

The difference between a model's OOS and COND measures is much higher in relative terms than in the MKT model, and it is again minimised when a long estimation window is employed. The 'delta' between the model's OOS and COND errors is approximately 80% for the MSEs and 40% for the MAEs, with a short half-year window observations, while it is much lower with the use of a ten-year window[18]. The COND mean-squared-error function is also featured in Inoue, Jin, and Rossi (2017), and based on my results in panels 2.7a and 2.7b I expect their criterion to be minimised with the use of a long estimation window. The pooled COND MSE for all stocks is minimised with the a ten-year window, with COND MSE = 32.1%% and COND MAE = 3.6%, and similar results are achieved with a five-year window, which is the second-best alternative, COND MSE = 34.3%% and COND MAE = 3.8%.

[Table 2.6 about here.]

In table 2.6 I report the results of the estimated optimal windows according to the criteria of Inoue, Jin, and Rossi (2017), panel 2.6b, and Pesaran and Timmermann (2007), panel 2.6a. For each stock, the estimate from the latter is an input of the objective function in Inoue, Jin, and Rossi (2017). Across all stocks, the average estimated window size with the Pesaran and Timmermann (2007) criterion is 5.1 years, while it is roughly 4 years of the most recent data

[18]Half-year window 'delta'= 21%% for the MSE measures and 'delta'= 1.3%. Compared to the OOS measures, the COND mesures are inflated by a factor of 80% for the MSE and 40% for the MAE. For the ten-year window, the COND measures are greater than the OOS ones by a factor of 41% for the MSE and 33% for the MAE.

in Inoue, Jin, and Rossi (2017). The corresponding median values are 3.6 and 2.9 respectively. Looking at the measures on the cross-sectional dispersion, I notice that for virtually all groups the distribution of estimated optimal windows is heavily skewed to to the spectrum of long window sizes. The 95% quantile in the distribution of the optimal windows is usually associated to the ten-year window mark, while the 5% conforms to the two-year window. The results are broadly in line with my expectations, and suggest that a relatively long estimation window (three to five-years of weekly data) minimises the conditional forecasting MSE. Moving on the PRED measures, I find that based on the expected return process, the Regional model's forecasting performance is higher the longer the sample. Similarly to the MKT model, the PRED MSE and PRED MAE are minimised with the use of a ten-year window. Comparing PRED and OOS measures (with window fixed), I find that using the model's pricing equation is always recommended over the expected return equation for forecasting purposes. The PRED MSEs are comparable for the five- and ten-year windows with $\approx$ 27%%, which is also equal to the OOS MSE with a half-year estimation window. Thus, using the model's expected return process with time-varying betas estimated every ten-year provides a similar performance as using the model's pricing equation with a short estimation window.

[Figure 2.8 about here.]

[Figure 2.9 about here.]

FF models. I now analyse the performance of the FF3 model, figure 2.8, and the FF5 model, figure 2.9. Based on the MSE and MAE, the in-sample performance of both models is maximised with the a long estimation window. For the FF3 model, the MSEs for the five- and ten-year window patterns are the lowest, IS MSE$\approx$ 17.5%%, followed by the the MSE corresponding to the half-year window, IS MSE $=$ 18.4%%. With the addition of the profitability and investment factors, the MSE is minimised with a ten-year estimation window, IS MSE$=$ 25.9%%, although the in-sample mean-squared error is higher by a factor of 50% with respect to the three-factor model. On the other hand, the variation in the in-sample MAEs between the two FF models is minimal for the preferred window size (ten-year). Focusing on the R^2 measures, I corroborate for the FF models as well the evidence of a trade-off between sample size and the model's performance in explaining contemporaneous return patterns. The short half-year window size maximises the share of explained return variation. The IS R^2 of the FF5 is highest across all model specifications, IS $R^2 = 45.9\%$, while the IS R^2 of the FF3 model is lower than the figure corresponding to the Regional model, FF3 IS $R^2 = 39.1\%$.

Analysing the out-of-sample performance of the FF models, I find similar results as to what reported earlier. The OOS MSE is minimised with a long estimation window (five-year window for FF3 is the optimal one, while it is ten-year for the FF5 model), and the R^2 measures of Kelly, Palhares, and Pruitt (2021), OOS R^2, are increasing in window size up to the two-year mark, which is the optimal window size. Similarly to the MSE, the MAE also points to the use of a long estimation sample for forecasting purposes. The variation in the metrics for different window sizes is certainly greater for the R^2-based measures than the mean- squared or absolute errors. Based on the former suite of measures, I corroborate the evidence on the use of a short estimation window in a contemporaneous-equations setting (in-sample), and for the two-year window in a forecasting setting (out-of-sample). This is true for all models considered. Based on the R^2 measures, the choice of the window length alone accounts for about $\pm 10\%$ of the FF3 forecasting performance (the difference between the OOS R^2s for the two- and ten-year windows, panel 2.8c) and for about 15% of the FF5 performance (the difference between the OOS R^2s for the two- and half-year windows, panel 2.9c).

Based on the summary statistics of the estimated factor betas, panels 2.8d for FF3 and panel 2.9d for FF5, I find that the magnitude of the loadings on factors other than the MKT is not increasing in window size. On average, the estimated loadings on these factors at all frequencies are at least one order of magnitude lower than the MKT factor sensitivities. Moving from the FF3 to FF5 specification, I notice that the loadings on the SMB and HML factors are consistently higher in magnitude, although not statistically significant at the 5% level except for the HML factor loadings estimated with a ten-year window. Overall, from the average t-stats I find little evidence of significant variation in the time-varying loadings of factors other than the (local) MKT. The loadings on the SMB factor are not on average statistically significant at the 5% level for both models, while the loadings on the HML, RMW, and CMA factors are only significant when a ten-year rolling estimation window is employed. Although for all factors I find that the the variance of the beta OLS estimator is decreasing in T, the magnitude is so low that the t-stats are considerably smaller than the figures for the Regional and MKT model. Looking at the R^2s in panels 2.8d and 2.9d, I corroborate the evidence on the improved in-sample performance of the models with betas estimated with a half-year window. For both models in fact, the average average goodness of fit of the regressions (in the estimation phase) is higher with the use of the shortest estimation window. The average R^2 from the half-year window regressions is 36% for the FF3, and 44% for the FF5.

I now discuss the results in relation to the conditional and predictive error measures. The COND MSE measures for both models suggest a significant gain in forecasting performance with the use of a long estimation window (five to ten years) with respect to a shorter one (two years or less). The COND MSEs and COND MAEs are minimised for both models when a ten-yer estimation window is used. Similarly to the Regional and MKT models considered earlier, the 'delta' between the FF models' OOS and COND errors is substantial even with the use of ten-year window, approximately 40% considering the MSEs and 28% for the MAE. Similarly to the benchmark models, the PRED MSE and PRED MAE are minimised with the use of a ten-year window. Comparing PRED and OOS measures (with window fixed) for FF3 and FF5, I corroborate the evidence that using the models' expected return equations with time-varying betas estimated every ten-year provides a similar forecasting performance as using the pricing equations with a short estimation window.

Overview. The performance measures defined in Section 2.3.2 help me study the contribution of rolling betas in forecasting future stock return patterns, and the key takeaways from the performance analysis are as follows.

- Based on the R^2 measures of Kelly, Palhares, and Pruitt (2021), I find that the models' performance to explain and predict future return patterns share a peculiar relation with the size of the estimation window. If I am interested in explaining contemporaneous return patterns, a short estimation window made of as little as 26 (weekly) observations provides the best performance out-of-sample. On the other hand, in a short-horizon forecasting framework ($h = 1$) I find evidence for the existence of a trade-off between performance and window length, with the optimal window being the one made with 104 observations (two-year). The choice of the window length alone accounts for about $\pm 10\%$ of the factor model's out-of-sample forecasting performance, this is true for all specifications considered. Comparing across models, I find that the model that performs best in-sample is FF5, with IS $R^2 = 46\%$, followed by the Regional model with $R^2 = 42\%$. The third-best performing model is FF3 with IS $R^2 = 39\%$, and the worst performer is the single-factor MKT model with IS $R^2 = 27\%$. The predictive performance of the MKT model based on the OOS R^2 is about 20%, followed by the FF3 and FF5 models which show a similar average OOS R^2 of about 25% for the three-factor model, and 24% for the five-factor counterpart. The model that shows the best capabilities in predicting future return patterns is the candidate model with OOS $R^2 = 29\%$.

- If I analyse the MSE and MAE measures, I find that for nearly all models the explanatory and forecasting performance is increasing with window length, the longest (ten-year) window minimises these error functions and a similar performance is achieved with the five-year window. The exception is the Regional model, in which the IS R^2 agrees with the IS MSE and hints at the use of a short estimation window (half-year). Comparing the models' performance according to the MSE measures, I find that the in-sample (explanatory) mean-squared error of the MKT model is 22.9%% (five-year window), followed by IS MSE = 25.9%% for FF5 (ten-year window), and finally the Regional model's IS MSE = 19.5%% (half-year window), and FF3 model with IS MSE = 18.4%% which is the best performer. Out-of-sample, I find that the FF models show better forecasting performance with respect to the other models. In fact, in decreasing order I find the MKT model, OOS MSE = 23%% (five-year window), the regional model with OOS MSE = 21.1%% (five-year window), and finally the FF3 and FF5 with similar figures of $\approx$ 17%% using a five-year window as well. The COND MSEs and MAEs are also minimised with a long estimation window, ten years. My results are broadly in line with the optimal window criteria of Inoue, Jin, and Rossi (2017), which features the COND MSE function and yields an average optimal window of about four years of weekly data. Based on the distribution of estimated optimal windows, I find that it is heavily skewed to the spectrum of longer (ten-year) windows. To quantify the cross-sectional dispersion in the estimates I calculate the 5% and 95% quantiles of the distribution of individual optimal windows, and find that the lower quantile conforms to the two-year benchmark while the upper one to the ten-year window. The median value is about 3 years. Moreover, based on the difference between the models COND and OOS measures, I find that the assumption of constant factor realisations from t^* to $t^* + h$ is more justified with models featuring solely observed factors (MKT, FF3, and FF5), than with models such as the Regional one with latent and observed factors.

- For all factor models featured in this chapter I also study the ability of time-varying factor sensitivities to explain future return patterns based on the expected return decomposition. The latter features static (ex-post) prices of risk, λ in equation (2.4), and dynamic factor betas which are estimated using a variety of window sizes. In Section 2.3.2 I define the PRED error measure to evaluate the forecasting performance based on the models' expected return equations. The metric is taken from Kelly, Palhares, and Pruitt (2021). Based on the mean-squared prediction error (PRED MSE), I find that using the models' expected return equations with time-varying betas estimated every ten-year provides a similar forecasting performance as using the pricing equations with betas calculated on a short estimation window (which is not preferred). The FF models perform best with PRED MSEs of approximately 20.1%%, while the Regional model performance is similar to the MKT model with PRED MSEs $\approx$ 26.7%%, indicating that the risk premia on the global and regional latent factors estimated using the factor rotation (IR3) proposed in Section 2.2.2 do not improve the model's forecasting performance.

Economic Interpretation

I proceed to examine the rolling beta estimates for the FF3 and Regional models and study how their time-series evolution is related to the financial and macro events defined in the economic calendar. The focus of this section is on the comparison of the loading dynamics for the competing three-factor models, fixing the estimation window. I report the estimates for the short half-year window, which maximises the in-sample R^2, and for the two-year window, which gives the best performance out-of-sample.

Regional model. Figure 2.10 shows the rolling factor sensitivities for the Regional model estimated using a half-year and a two-year window. I focus on the analysis of the short-window estimates which are the most volatile and help me better map the changes in the factors' relative significance to macro events that are sometimes short-lived. Using the two-year estimates for economic interpretation is difficult because most of the time-series variation in the betas is averaged out.

During the GFC and ESDC, I find that for nearly all regions the loadings on the global factors (financial and global) increase considerably. From 2009 to mid 2011 the loadings on the financial and global factors decrease, this is apparent from the two-year window estimates in figure 2.10, before increasing rapidly at the outset of the ESDC. In fact, in mid 2011 I find a substantial change in the trajectory of the financial factor loadings for regions such as North America, Western and Emerging Europe, and somewhat surprisingly Latin America as well. Referring to the region-specific events in North America, I show that the sensibility of stock returns to changes in the financial factor increase in the weeks preceding the election of us presidential elections in late 2016. Moreover, during the months corresponding to the US-China trade war, my results suggest that a meaningful decrease in the financial factor loadings (in late 2018), at the expenses of the global factor whose loadings rise during the same period.

In North America, there are few events that are responsible for a substantial decrease in the relative importance of the financial factor with respect to the global and regional ones. Excluding the US-China trade war, the only other period in which the loadings on the financial factor are comparable in magnitude to the others is at the outset of the Chinese stock market crash in mid-2015. Some of the events considered for the stocks listed in North America are also relevant for those domiciled in the Asia-Pacific region, where I find a material increase in the global factor loadings during the Chinese stock market crash of 2015-2016. Similarly to the behavior observed for the North American equities, at the outset of the US-China trade war the loadings on the global and financial factors decrease considerably, and stocks became equally sensitive to regional factor shocks.

Another event that I include in my calendar is the oil market crash initiated in late 2014, which I assume runs throughout the year 2015. The start and end dates are shaded in red for the Latin America, Emerging Europe and MEA regions. I find that the loadings on the regional factor increase at the start of 2015 for the equities in MEA, while the global factor dominates the others in terms of magnitude for the stocks listed in Emerging Europe and Latin America. Moving on to the Western Europe region, I find that in the weeks leading to the Brexit referendum of 2016 the loadings on the regional factor start to rise consistently, while the average sensitivity of the stocks in this region to shocks to the financial and global factors decrease significantly. This is true up until the start of November, which coincides with the month of the 2016 US presidential elections, after which I notice a reversal in the the relative significance of the factors, with the loadings on the global and financial factors growing from an average of 0.3 to more than 0.5 by the end of 2017.

[Figure 2.10 about here.]

FF3 model. Figure 2.11 shows the time-series plots of the rolling factor sensitivities for the FF3 model. For each region, I report the estimates using a half-year window, left panels, and a two-year window, right panels. Compared to the Regional model, I notice that the the loadings on the observed HML and SMB factors are materially lower that the ones on the MKT factor, which dominate in all regions in terms of average magnitude and statistical significance. As reported in panel 2.8d, the loadings on factors other than the MKT do not show evidence of statistically significant changes over time (all windows), and in fact I notice from figure 2.11 that they tend to fluctuate around the zero mark and often fall into negative territory. During the Chinese stock market crash of 2015 in the Asia-Pacific region, I find that the sensitivities of the stocks to all factors increase considerably, which implies greater covariance estimates using the factor model variance decomposition (ceteris paribus), a phenomenon which is in line with stylised factors in the literature (i.e. during turmoil periods stock co-movements increase). Overall, I find that when I allow for time-varying factor sensitivities in the FF3 model, the loadings on the SMB and HML factors are not on average statistically different from zero, and are difficult to interpret economically. In terms of average magnitude, I do not find any evidence of substantial changes in the relative importance of the three factors over time for the stocks domiciled in the North America, Asia-Pacific and Western Europe regions.

[Figure 2.11 about here.]

2.6 Conclusion and Further Research

In this chapter, I studied how the choice of the estimation window influences the explanatory and predictive power of conditional asset pricing models of stock returns. I used a model inspired by Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021) that features static factors known ex-post and dynamic betas estimated out-of-sample via rolling least squares. I use single stocks as test assets, including more than 1500 of the biggest companies listed in 40 different national stock indeces in the period January 2006 to end-of May 2019. I use various factor models from the literature as benchmarks in my analysis, and I also review the properties of the model in Borghi et al. (2018) that features a combination of observed and latent factors. Conditional on the ex-post factors, I estimate the real-time betas for varying window sizes and compare the models' in- and out-of-sample performance based on a variety of measures.

The main results of my analysis can be summerised as follows. In relation to the factor space, a key difference of the FF factors with respect to the combination of observed and latent factors in the model of Borghi et al. (2018) is that the former are not orthogonal to each other, which implies that they do not isolate different sources of systematic variation in the returns. This results in factor loadings that fail to be statistically relevant out-of-sample in the cross-section of stock returns. For instance, while the full-sample estimates suggest that the loadings on the SMB and HML in FF3 are on average statistically different from zero across all regions (and for the vast majority of the industries), based on the out-of-sample analysis I find that their performance deteriorates significantly, to the point that the average beta estimates constructed for window sizes up to the ten-year mark are not statistically different from zero. In FF5, the in-sample loadings on the RMW and CMA are statistically significant only in few of the groups considered, and out-of-sample they fail to be relevant for virtually all groups. I corroborate the evidence in Fama and French (2015) that for stocks listed in North America (which includes the United States) the HML factor becomes redundant with the addition of the RMW and CMA factors, and extend their results internationally.

In the candidate model of Borghi et al. (2018), the financial factor is identified with the S&P500 Financials equity index, which tracks the performance of the biggest US financial tickers, and the orthogonal global factor isolates sources of variation in the excess returns that are linked to financial and macro events that have a world-wide impact. The third set of factors featured in the model include region-specific drivers that I find to be closely related to the dynamics of an equity index proxy constructed with the relevant stocks of the region. The loadings on these factors remain statistically significant out-of-sample for the combination of different window sizes.

Secondly, with respect to the estimation of rolling betas in linear asset pricing models, I find a dual role of the rolling least square estimator in the cross-section of stock returns. The short-window approach of Lewellen and Nagel (2006) (half-year of weekly data) has much to recommend if I am interested in describing contemporaneous stock return variation, however for predictive purposes including too little observations for estimation causes the betas to be noisy, which in turn results in forecasts from the factor model that show little predictive power according to the R^2-based measures in Kelly, Palhares, and Pruitt (2021). Out-of-sample, the trade-off between the length of the estimation window and the variance of the estimator is resolved around the two-year benchmark, and this is true across all models considered. The choice of the window length alone accounts for about $\pm 10\%$ of the factor model's out-of-sample forecasting performance (R^2). On the other hand, based on the MSE and MSA measures, I find that for nearly all models the explanatory and forecasting performance is increasing in window size, with the five- and ten-year windows yielding the best results. My results are in line with the optimal window criteria of Inoue, Jin, and Rossi (2017), which features the COND MSE function and, when applied to my data, yields an average optimal window of about four years of weekly observations. Rolling estimates of the time-varying betas constructed using a short estimation window (which maximise the in-sample R^2) help us map the changes in the factors' relative significance to macro and financial events. Despite their statistical properties, using the two-year estimates (which maximise the out-of-sample R^2) for economic interpretation is difficult because most of the time-series variation in the betas is averaged out. I extend the in-sample results of Borghi et al. (2018) and consider a rich set of region specific and global macro events, corroborating their evidence that during turmoil periods stocks tend to become more sensitive to changes in the financial and global factors.

2.6.1 Further Research

The analysis in this chapter provides a further contribution to the understanding of how factor models with time-varying rolling betas can explain future return patterns in international asset prices, and complements the work in Chapter 1 which focuses on the analysis of contemporaneous return patterns.

In this study I focus on explaining expected stock returns, however the setup outlined in Section 2.2 can also be used to study the contribution of time-varying betas in shaping the comovements structure implied by the factor model. The peculiarity of the factor model that I employ is that the time-variation in the first- and second- moments of (excess) stock return is driven solely by conditional factor exposures, which extends the Kelly, Moskowitz, and Pruitt (2021) and Kelly, Palhares, and Pruitt (2021) framework to the analysis of covariances. In Appendix B.1 I expand on this idea and present a preliminary modelling framework. I leave this research question for future studies.

Further research on this chapter include examining alternative objective functions to study the models' performance. In fact, while I find the existence of a trade-off between window length and predictive power (two-year window), this depends on the choice of the objective function. The R^2 metrics inspired by Kelly, Palhares, and Pruitt (2021) do not agree with the commonly-used mean- and absolute-squared measures, even when they are constructed from the same 'errors', as per definition in Section 2.3.2. MSE functions are also used in optimal window selection criteria such as Inoue, Jin, and Rossi (2017) and Pesaran and Timmermann (2007) and my findings suggest that these objective functions tend to agree with the use of a longer window (about four years of recent data) when applied to my dataset. This remains an open research question. Additionally, a key limitation of my study is that I only let the window size vary, for fixed sampling frequency, which excludes a simultaneous analysis of the two as in Lewellen and Nagel (2006). I leave this research question for future studies.

The benchmark models against I compare include the local FF models and the global CAPM, however I exclude competing alternatives in the asset pricing literature such as four-factor model of Carhart (1997), which augments the FF3 specification with the momentum factor, and the q-factor model of Hou, Xue, and Zhang (2014), which features a market factor, a size factor, an investment factor, and a profitability factor. The latter approach is similar to excluding HML with respect to FF5, which I show to be redundant with respect to the CMA investment factor.

Finally, further research based on a more recent time frame is needed to assess the impact of the COVID-19 outbreak on the dynamics of factor sensitivities. This is also true for the US-China trade conflict, which I am able to consider up to stage four[19], thereby excluding the phase from February 2020 to today (stage five) in which tariffs between the two countries remain elevated, above pre-conflict levels although lower than what I analyse in stages one to four. In regards to the COVID-19 outbreak, I expect the loadings on the global factor in the model of Borghi et al. (2018) to increase, probably more than the ones on the observed financial factor, due to the world-wide impact of the event which does not affect only US financial stocks. Similarly, I expect stocks to be relatively more exposed to region-specific shocks during the period, due to the variety of the containment measures adopted across countries.

[19]Source: Peterson Institute for International Economics.

TABLE 2.1: **Rolling Window in the Literature**

The table summarises the characteristics of the papers in the finance literature that I review with respect to the choices of the window size, W, sampling frequency, *Frequency*, and forecasting horizon, h. I also report if portfolios or individual securities are used as test assets in the empirical studies of the papers, *Test Asset*. The table shows the focus asset class, *Asset Class*, the complete time span of the rolling analysis (without considering any sub-samples), *Time Span*, and finally the regions considered, *Region*.

Fama and French (1997): multiple forecasting horizons (one month, and one to five years) as well as rolling window sizes (three to ten years) are considered in the paper, in the table below I only report the set of parameters for which the results are shown.

Lewellen and Nagel (2006): contemporaneous analysis thus $h = 0$. Short-window regressions based on quarterly betas with daily observations ($W = 63$, Frequency=Daily), semiannually using daily ($W = 126$, Frequency=Daily) and weekly returns ($W = 26$, Frequency=Weekly), and annually using monthly returns ($W = 12$, Frequency=Annually).

Ang, Chen, and Xing (2006): multiple forecasting horizons (from one to 12 years), all at yearly frequency. The standard approach is one year of daily data, ($W = 250$, Frequency=Daily) however results are also checked against using intervals of 24 months with weekly frequency ($W = 104$, Frequency=Weekly).

Paper	Test Asset	Asset Class	Time Span	Region	Frequency	h	W	W (Years)
Ferson and Harvey, 1991	Portfolios	Stocks, Bonds	1964-1986	US	Monthly	1	60	5
Fama and French, 1997	Portfolios	Stocks	1968-1994	US	Monthly	1	36, 48, 60*	3, 4, 5*
Petkova and Zhang, 2005	Portfolios	Stocks	1927-2001	US	Monthly	1	36, 48, 60	3, 4, 5
Lewellen and Nagel, 2006	Portfolios	Stocks	1964-2001	US	*Multiple*	0	*Multiple*	*Multiple*
Ang, Chen, and Xing, 2006	Securities	Stocks	1963-2001	US	Daily	1*	250*	1
Bekaert, Hodrick, and Zhang, 2009	Portfolios	Stocks	1980-2005	World	Weekly	0	26	0.5

TABLE 2.2: **Universe of Securities**

The table reports the countries that are considered for each region, and for each country it reports the following variables: *#Stocks* is the number of companies that entered the index during the period from January 6th 2006 to May 31st 2019, *Avg Active* is the average number index of members at the beginning of every month in the sample period, *#Selected* is the number of stocks with complete price series that are considered for the out-of-sample analysis. These tickers have no more than 12 consecutive missing observations, and at least one year of weekly data. Finally, *#Full* is the number of equities with complete price series that remained listed in the national indexes throughout the entire sample period.

Index	Country	Region	#Stocks	AvgActive	#Selected	#Full
SPTSX60	Canada	North America	106	60	98	64
OEX	US	North America	179	101	174	129
MEXBOL	Mexico	North America	76	35	72	38
MERVAL	Argentina	Latin America	87	17	83	62
IBOV	Brazil	Latin America	139	65	126	53
IPSA	Chile	Latin America	86	37	80	64
SPBLPGPT	Peru	Latin America	90	34	60	38
TPXL70	Japan	Asia-Pacific	128	70	126	104
SSE50	China	Asia-Pacific	158	50	154	90
HSCEI	HongKong	Asia-Pacific	93	41	90	56
SENSEX	India	Asia-Pacific	85	30	71	0
LQ45	Indonesia	Asia-Pacific	122	45	116	70
KOSPI50	Korea	Asia-Pacific	87	50	63	0
SET50	Thailand	Asia-Pacific	107	50	107	64
NZSE50FG	NewZealand	Asia-Pacific	96	50	87	45
AS31	Australia	Asia-Pacific	94	50	89	55
ATX	Austria	Western Europe	41	20	40	23
BEL20	Belgium	Western Europe	38	20	36	26
KFX	Denmark	Western Europe	34	20	32	26
HEX25	Finland	Western Europe	35	25	33	26
CAC	France	Western Europe	68	40	65	48
DAX	Germany	Western Europe	49	30	44	37
ISEQ	Ireland	Western Europe	94	51	80	22
AEX	Netherlands	Western Europe	61	25	50	29
OBX	Norway	Western Europe	67	25	60	28
PSI20	Portugal	Western Europe	38	19	34	21
IBEX	Spain	Western Europe	62	35	61	33
OMX	Sweden	Western Europe	43	30	42	36
SMI	Switzerland	Western Europe	55	21	53	43
UKX	UK	Western Europe	208	101	195	119
CRO	Croatia	Emerging Europe	94	23	86	38
CCTX	CzechRepublic	Emerging Europe	15	9	14	6
TALSE	Estonia	Emerging Europe	23	16	20	7
BUX	Hungary	Emerging Europe	33	14	29	16
RIGSE	Latvia	Emerging Europe	73	28	43	20
MALTEX	Malta	Emerging Europe	33	18	19	11
VILSE	Lithuania	Emerging Europe	47	27	39	14
WIG20	Poland	Emerging Europe	45	20	44	27
ROTXEUR	Romania	Emerging Europe	25	11	21	10
CRTX	Russia	Emerging Europe	49	13	9	0
BELEX15	Serbia	Emerging Europe	26	13	11	7
XU030	Turkey	Emerging Europe	78	30	67	49
PFTS	Ukraine	Emerging Europe	45	17	8	0
MOSEMDX	Morocco	MEA	82	48	74	32
Total			3294	1534	2905	1686

TABLE 2.3: **Summary Statistic**

The table reports the summary statistics for the 1686 tickers that are part of the national stock indexes in the sample period, January 2006 to May 2019. Panel 2.3a reports cross-sectional averages of the summary statistics for the weekly log-returns, and panel 2.3b reports average market capitalisation, total assets and debt. *Mean* is the cross-sectional average weekly return (in basis points) for the stocks belonging to a particular region or sector, analogously *Med* is the average median return (in bp), *Min* the most-negative weekly return for the cross-section, and *Max* the highest weekly return. *Std* is the average weekly standard deviation, *Skew* and *Kurt* are the average skewness and kurtosis. $\rho(1)$ is the average OLS estimate of the first autocorrelation coefficients, and *ADF* is the Augmented Dickey-Fuller test statistics, which is run with a constant, time trend and one lag. The critical value at 95% significance is -3.41, with the null hypothesis being the presence of a unit root. Lastly, *Pearson* is the average pair-wise Pearson correlation of the stocks in the relevant group.

(A) *Stock Returns*

Group	Mean (bp)	Med (bp)	Min	Max	Std	Skew	Kurt	$\rho(1)$	ADF	Pearson	N
North America	7.17	9.6	-0.335	0.252	0.047	-0.62	13.953	-0.049	-18.799	0.357	231
Latin America	6.06	-2.29	-0.386	0.313	0.059	-0.237	11.219	-0.022	-18.166	0.278	217
Asia-Pacific	12.57	2.95	-0.325	0.28	0.054	-0.222	9.745	-0.023	-18.586	0.257	484
Western Europe	2.64	7.5	-0.344	0.26	0.052	-0.649	11.726	-0.054	-18.923	0.415	517
Emerging Europe	-7.2	-2.85	-0.455	0.413	0.065	-0.351	18.066	-0.012	-18	0.301	205
MEA	10.78	-0.23	-0.193	0.214	0.04	0.218	7.861	-0.073	-19.807	0.187	32
Communication Services	-0.09	1.83	-0.322	0.265	0.052	-0.381	9.973	-0.047	-18.716	0.259	114
Consumer Discretionary	5.11	4.49	-0.392	0.338	0.058	-0.311	14.121	-0.024	-18.442	0.238	167
Consumer Staples	11.71	3.82	-0.301	0.257	0.047	-0.26	10.973	-0.049	-18.933	0.199	158
Energy	-1.58	1.15	-0.381	0.319	0.061	-0.39	10.099	-0.027	-18.749	0.339	119
Financials	3.13	2.84	-0.384	0.302	0.054	-0.56	15.428	-0.046	-18.677	0.317	259
Health Care	15.82	9.2	-0.284	0.223	0.045	-0.496	9.798	-0.049	-18.845	0.245	91
Industrials	3.88	4.58	-0.356	0.284	0.055	-0.426	11.78	-0.029	-18.469	0.27	293
Information Technology	9.54	7.88	-0.316	0.269	0.052	-0.359	10.773	-0.027	-18.677	0.23	86
Materials	5.09	0.39	-0.369	0.314	0.061	-0.24	9.613	-0.014	-18.182	0.279	212
Real Estate	7.52	5.93	-0.42	0.337	0.059	-0.863	19.658	-0.047	-18.573	0.244	71
Utilities	8.57	4.88	-0.304	0.235	0.046	-0.47	10.261	-0.053	-19.055	0.254	112

(B) *Balance Sheet*

Group	Mkt Cap ($B)	Tot Assets ($B)	Tot Debt ($B)	N
North America	51.15	111.148	28.712	231
Latin America	5.416	17.494	5.807	217
Asia-Pacific	9.465	38.578	9.665	484
Western Europe	18.968	106.694	30.563	517
Emerging Europe	1.711	7.378	1.653	205
MEA	1.439	3.963	0.779	32
Communication Services	22.194	29.212	9.476	114
Consumer Discretionary	12.231	18.36	6.3	167
Consumer Staples	19.912	15.729	4.395	158
Energy	22.567	40.136	8.129	119
Financials	20.872	290.629	78.226	259
Health Care	32.402	22.542	5.397	91
Industrials	9.555	18.319	5.797	293
Information Technology	37.387	22.336	4.553	86
Materials	7.593	12.977	3.401	212
Real Estate	4.459	8.483	3.37	71
Utilities	9.438	26.522	9.308	112

$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 2.4: **Mapping Local Factors Correlations**

The table reports the in-sample correlation matrix of the estimated regional factors, panel 2.4a, and of the $1/N$ equity index proxies for the regions, panel 2.4b. I test the significance of the estimated parameters using a t-statistic with $T-2$ degrees of freedom based on a transformation of the correlation coefficient. The transformation is exact when the variables are normally distributed. * denotes significance at 10% level, ** at 5% level, and *** at 1% level.

(A) *Regional Factors*

	North America	Latin America	Asia-Pacific	Western Europe	Emerging Europe	MEA
North America	1					
Latin America	0.08**	1				
Asia-Pacific	-0.06*	0.08**	1			
Western Europe	**0.24*****	0	-0.12***	1		
Emerging Europe	0.1**	**0.13*****	0	0.08**	1	
MEA	**-0.15*****	0	0	**0.13*****	0	1

(B) *Index Proxies*

	North America	Latin America	Asia-Pacific	Western Europe	Emerging Europe	MEA
North America	1					
Latin America	**0.77*****	1				
Asia-Pacific	0.71***	0.72***	1			
Western Europe	**0.89*****	0.75***	0.72***	1		
Emerging Europe	0.75***	0.74***	0.72***	**0.82*****	1	
MEA	**0.29*****	0.36***	0.35***	0.41***	0.46***	1

$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 2.5: **Static Loadings Estimation**

The table reports cross-sectional averages of the OLS beta estimates on the full sample for the models considered. In particular, for each factor I report the average beta magnitude and the average absolute value of the t-statistic, and for each region and industry I show the N-average R^2 for the stocks in the relevant group. Before estimation, stock and factor returns are winsorised at 99% level and then standardised to have unit variance and zero mean. Panel 2.5a reports the results for three-factor Regional model, and panel 2.5b for single-factor MKT model. To ease the comparison across models, I report the estimates only for the 1686 companies that remain listed in the respective national equity indeces during the 13 years considered. Panel 2.5c shows the results for the FF3 model, and panel 2.5d for the FF5 model. The results on the FF models are available only for the cross-section of companies belonging to the North America, Asia-Pacific, and Western Europe regions, 1232 tickers among the 1686.

(A) *Regional Model*

	Financial		**Global**		**Regional**			
Group	β^{fin}	t-stat	β^{glob}	t-stat	β^{reg}	t-stat	R^2	N
North America	0.485	17.4	0.242	8.4	0.193	6.8	0.39	231
Latin America	0.288	9.5	0.327	10.6	0.239	8.2	0.31	217
Asia-Pacific	0.227	7.3	0.319	10.1	0.182	6.4	0.28	484
Western Europe	0.456	16.9	0.365	13.3	0.245	9.1	0.43	517
Emerging Europe	0.264	9.2	0.316	10.6	0.152	7.2	0.28	205
MEA	0.074	2.3	0.213	6.7	0.42	13.3	0.25	32
Communication Services	0.339	11.3	0.312	10.1	0.194	6.6	0.3	114
Consumer Discretionary	0.343	11.8	0.294	9.8	0.197	7.2	0.32	167
Consumer Staples	0.27	8.6	0.294	9.2	0.196	6.5	0.25	158
Energy	0.354	12.2	0.404	13.9	0.169	6.1	0.38	119
Financials	0.428	16.6	0.32	11.8	0.218	8.9	0.43	259
Health Care	0.291	9.2	0.253	7.8	0.192	6.3	0.25	91
Industrials	0.357	12.8	0.324	11.2	0.237	8.9	0.37	293
Information Technology	0.353	11.7	0.265	8.6	0.227	7.8	0.31	86
Materials	0.33	11.8	0.369	12.8	0.211	7.6	0.37	212
Real Estate	0.311	10.6	0.329	10.9	0.165	6.4	0.32	71
Utilities	0.264	8.8	0.314	10.3	0.265	9.3	0.32	112

(B) *MKT Model*

	Market			
Group	β^{mkt}	t-stat	R^2	N
North America	0.51	16.3	0.27	231
Latin America	0.441	13.5	0.21	217
Asia-Pacific	0.389	11.8	0.18	484
Western Europe	0.595	20.6	0.37	517
Emerging Europe	0.416	12.7	0.2	205
MEA	0.207	5.6	0.05	32
Communication Services	0.466	14.5	0.23	114
Consumer Discretionary	0.46	14.7	0.24	167
Consumer Staples	0.399	12	0.18	158
Energy	0.528	17.2	0.3	119
Financials	0.537	18.1	0.31	259
Health Care	0.392	11.7	0.17	91
Industrials	0.494	16.1	0.27	293
Information Technology	0.451	13.9	0.22	86
Materials	0.499	16.5	0.28	212
Real Estate	0.454	14.3	0.23	71
Utilities	0.418	12.8	0.2	112

(To be continued)

(Continued)

(C) *FF3 Model*

	MKT		SMB		HML			
Group	β^{fin}	t-stat	β^{glob}	t-stat	β^{reg}	t-stat	R^2	N
North America	0.547	16.9	-0.015	2.3	0.048	3.5	0.34	231
Asia-Pacific	0.414	11.3	0.06	2.2	0.035	2.1	0.19	484
Western Europe	0.621	17.5	0.042	2.7	0.022	2.9	0.41	517
Communication Services	0.491	13.3	0.013	2	0.025	1.9	0.27	88
Consumer Discretionary	0.518	14.3	0.046	2.1	0.022	2.2	0.29	124
Consumer Staples	0.476	12.6	-0.009	2.8	-0.027	2.3	0.23	103
Energy	0.542	15.5	0.067	2.6	0.056	2.5	0.32	92
Financials	0.568	17.8	0.008	1.9	0.126	5.2	0.42	173
Health Care	0.465	12.4	-0.007	2.7	-0.074	2.9	0.23	81
Industrials	0.564	16.2	0.071	2.7	0.038	2.3	0.33	222
Information Technology	0.526	14.8	0.04	2.5	-0.024	2.3	0.29	76
Materials	0.534	15.3	0.086	2.8	0.043	2.5	0.32	150
Real Estate	0.505	13.9	0.051	2.3	0.029	1.9	0.28	57
Utilities	0.452	12.2	-0.01	2.6	0.03	2.1	0.25	63

(D) *FF5 Model*

	MKT		SMB		HML		RMW		CMA			
Group	β^{MKT}	t-stat	β^{SMB}	t-stat	β^{HML}	t-stat	β^{RMW}	t-stat	β^{CMA}	t-stat	R^2	N
North America	0.548	15.8	-0.009	2.1	0.065	3.6	0.03	2.2	-0.01	2.7	0.36	231
Asia-Pacific	0.415	9.7	0.059	2.1	0.038	1.4	0.012	1.4	-0.004	1.4	0.2	484
Western Europe	0.604	15.4	0.041	2.7	0.039	2.2	0.004	1.6	-0.033	2.2	0.42	517
Communication Services	0.508	12.4	0.019	1.9	0.017	1.5	0.011	1	0.029	2	0.28	88
Consumer Discretionary	0.501	12.4	0.051	2.1	0.056	2	0.032	1.3	-0.037	2	0.3	124
Consumer Staples	0.494	11.8	0.004	2.4	-0.021	1.7	0.041	1.6	0.039	1.9	0.24	103
Energy	0.51	13.1	0.062	2.3	0.108	3.2	0.041	1.8	-0.076	2.5	0.34	92
Financials	0.548	15.5	-0.005	2.1	0.113	3.1	-0.055	3	-0.054	2.5	0.44	173
Health Care	0.489	11.7	-0.006	2.5	-0.107	2.8	-0.023	1.5	0.051	2	0.24	81
Industrials	0.555	14.2	0.074	2.6	0.055	1.8	0.022	1.4	-0.016	1.6	0.34	222
Information Technology	0.524	13.3	0.043	2.2	-0.016	1.8	0.014	1.1	-0.01	1.6	0.29	76
Materials	0.514	13.1	0.084	2.6	0.074	2.3	0.028	1.3	-0.05	1.8	0.33	150
Real Estate	0.498	12.1	0.053	2.3	0.049	1.3	0.023	1.6	-0.025	1.7	0.29	57
Utilities	0.467	11.3	-0.003	2.3	0.031	1.1	0.028	1.9	0.034	1.7	0.26	63

$N = 1232$

$T = 700$ (6th Jan 2006 - 31st May 2019)

TABLE 2.6: **Individual Optimal Window**

The table reports the results of the individual optimal selection criteria of Inoue, Jin, and Rossi (2017), panel 2.6b, and Pesaran and Timmermann (2007), panel 2.6a, for the Regional model. The figures are expressed in years, and the frequency of the data is weekly. For each group, I report the N-average estimated window size, *Mean*, the median window size, *Median*, and estimates of the cross-sectional dispersion calculated as the 5% and 95% quantiles of the distribution of individual optimal windows.

(A) *Pesaran and Timmermann (2007)*

Group	Mean	5%	Med	95%	N
North America	5.7	2.1	3.8	10.8	231
Latin America	5.2	2	3.6	11.2	217
Asia-Pacific	5.2	2.1	3.7	11.1	484
Western Europe	4.8	2	2.9	10.9	517
Emerging Europe	5.4	2	3.8	11.4	205
MEA	4.5	2.1	2.8	10.7	32
Communication Services	5.3	2	3.6	10.8	114
Consumer Discretionary	5.5	2	3.8	11	167
Consumer Staples	5.4	2	3.8	11	158
Energy	4.7	2	3.4	10.8	119
Financials	5.6	2	3.8	11	259
Health Care	5.2	2	3.8	11.2	91
Industrials	5	2	3.7	11.3	293
Information Technology	4.8	2	3.4	11.1	86
Materials	5	2	3.5	11.1	212
Real Estate	5.1	2.1	3.5	11	71
Utilities	4.5	2	3.3	11.2	112
Total	5.1	2	3.6	11	1686

(B) *Inoue, Jin, and Rossi (2017)*

Group	Mean	5%	Med	95%	N
North America	4.3	2.2	3.1	9.7	231
Latin America	5	2	3.1	10.9	217
Asia-Pacific	3.5	2.1	2.8	7.5	484
Western Europe	3.7	2.2	2.8	7.4	517
Emerging Europe	4.9	2.1	3.6	10.4	205
MEA	4.7	2	3.1	11.3	32
Communication Services	3.9	2.1	2.9	7.5	114
Consumer Discretionary	3.9	2.1	2.9	9.1	167
Consumer Staples	4.5	2.1	3.1	10.3	158
Energy	3.9	2.2	2.9	10	119
Financials	3.7	2.1	2.7	7.5	259
Health Care	5	2.2	3.6	10	91
Industrials	4.2	2.2	3.1	10.1	293
Information Technology	4.5	2.3	3.3	10.2	86
Materials	3.6	2.1	2.7	9	212
Real Estate	3.2	2.3	2.8	6.4	71
Utilities	4.6	2	2.8	10.5	112
Total	4.1	2.1	2.9	9.9	1686

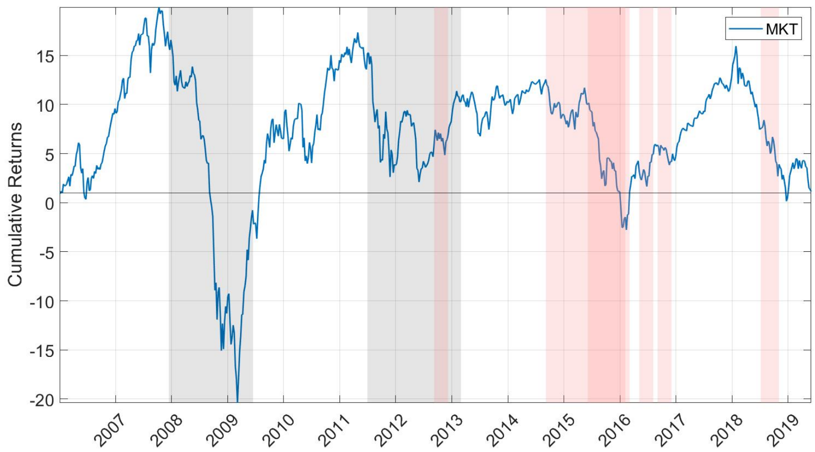
$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

FIGURE 2.1: **MKT Model**

The figure shows the cumulative returns of the equally weighted portfolio formed by all stocks in my universe, panel 2.1a, and reports summary statistics for the factor's weekly log- returns, panel 2.1b. $\hat{\lambda}$ refers to the (annualised) estimated risk premium based on the factor's weekly excess returns in the sample period. Skew is the sample skewness, Kurt the sample kurtosis, and JB is the p-value of a Jarque-Bera test for normality. Factor returns are winsorised at 99% and normalised to have unit variance in the sample period. All 1686 tickers contribute to the calculation of the MKT factor. The grey bands refer to the GFC of 2007-2009 and the ESDC of 2011-2013, respectively in chronological order. The light red bars isolate the start and end dates corresponding to region-specific financial events. Details on the economic calendar are found in Appendix B.2.

(A) *MKT*



(B) *Summary Statistics*

Factor	$\hat{\lambda}$ (%)	Skew	Kurt	JB	N
MKT	1	-0.58	5.2	0	1686

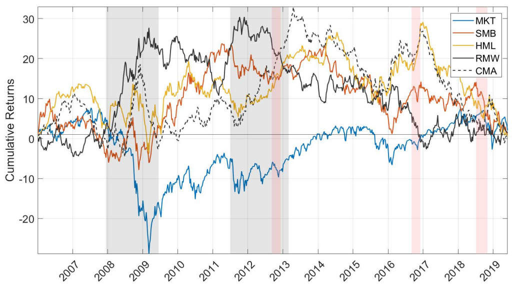
$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

FIGURE 2.2: **FF Models**

The figure shows the cumulative returns and summary statistics of the MKT, SMB, HML factors of the FF3 model, together with the two additional factors, RMW, and CMW, featured in the FF5 model. Results are region-specific. Panels 2.2a and 2.2b refer to the North America region, panels 2.2c and 2.2d to the Asia-Pacific region, and finally panels 2.2e and 2.2f to Western Europe. In the summary statistics tables, $\hat{\lambda}$ indicates the (annualised) estimated risk premium based on the factor's weekly excess returns in the sample period. Skew is the sample skewness, Kurt the sample kurtosis, and JB is the p-value of a Jarque-Bera test for normality. Factor returns are winsorised at 99% and normalised to have unit variance in the sample period. The grey bands refer to the GFC of 2007-2009 and the ESDC of 2011-2013, respectively in chronological order. The light red bars isolate the start and end dates corresponding to region-specific financial events. Details on the economic calendar are found in Appendix B.2.

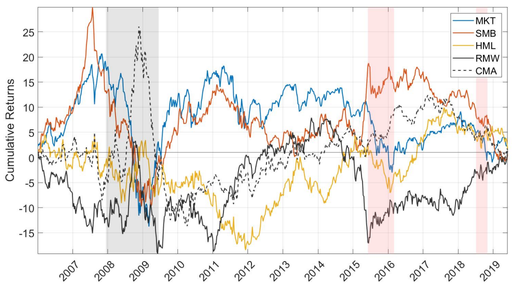
(A) *North America*



(B) *Summary Statistics*

Factor	$\hat{\lambda}$ (%)	Skew	Kurt	JB
MKT	8	-0.58	4.5	0
SMB	-1	-0.16	3	0.19
HML	-3	0.41	4.4	0
RMW	3	-0.07	3.3	0.15
CMA	0	0.26	3.5	0

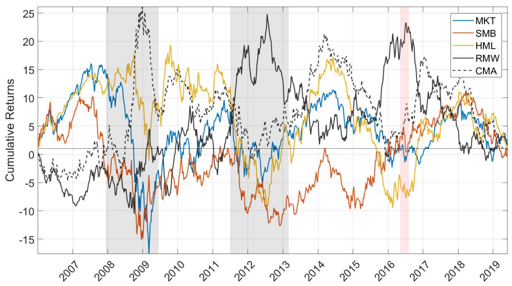
(C) *Asia-Pacific*



(D) *Summary Statistics*

Factor	$\hat{\lambda}$ (%)	Skew	Kurt	JB
MKT	7	-0.67	5.4	0
SMB	-3	-0.62	5.1	0
HML	4	-0.08	3.6	0.01
RMW	2	0	3.4	0.08
CMA	4	0.25	5.2	0

(E) *Western Europe*



(F) *Summary Statistics*

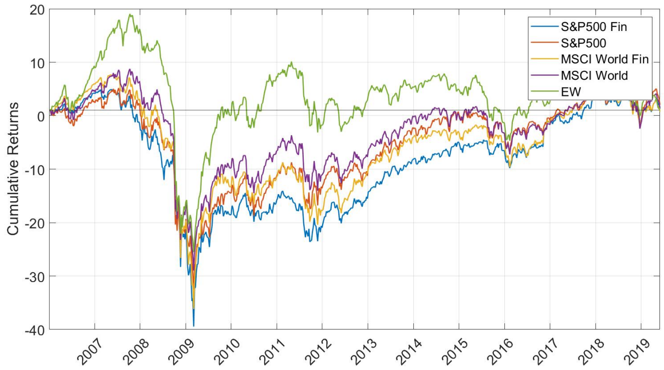
Factor	$\hat{\lambda}$ (%)	Skew	Kurt	JB
MKT	5	-0.49	4.3	0
SMB	0	-0.33	3.9	0
HML	-2	-0.11	3.9	0
RMW	5	-0.1	3.4	0.05
CMA	1	0.16	3.9	0

$T = 700$ (6th Jan 2006 - 31st May 2019)

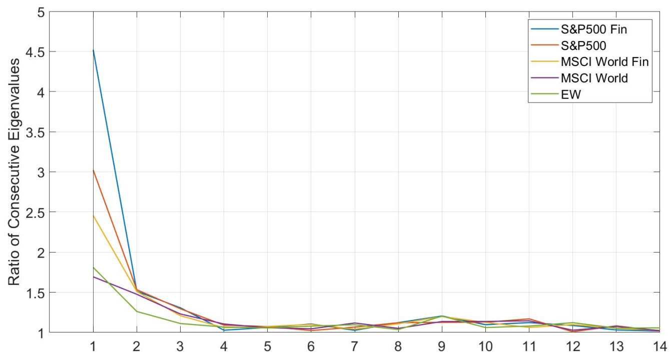
FIGURE 2.3: **Financial Factor Identification**

The figure reports the results of the identification procedure of the observed financial factor F_t of model (2.9). Panel 2.3a shows the time-series plot of the cumulative factor returns during the sample period, January 2006 to May 2019, before winsorisation. The candidate factors include the S&P500 index, its Financials-only counterpart, the S&P500 Financials, two broad global indeces of stock market performance, MSCI World and MSCI World Financials, together with the equally-weighted portfolio of all stocks in my universe, EW. Panel 2.3b reports the ratio of consecutive eigenvalues estimated from the matrix of excess return after orthogonalisation against the candidate equity indeces. Finally, panel 2.3c reports the estimated summary statistics of the candidate market indeces, before winsorisation. $\hat{\lambda}$ indicates the (annualised) estimated risk premium based on the factor's weekly excess returns in the sample period. Skew is the sample skewness, Kurt the sample kurtosis, and JB is the p-value of a Jarque-Bera test for normality.

(A) *Candidate Factors*



(B) *Ratio of Consecutive Eigenvalues*



(C) *Summary Statistics*

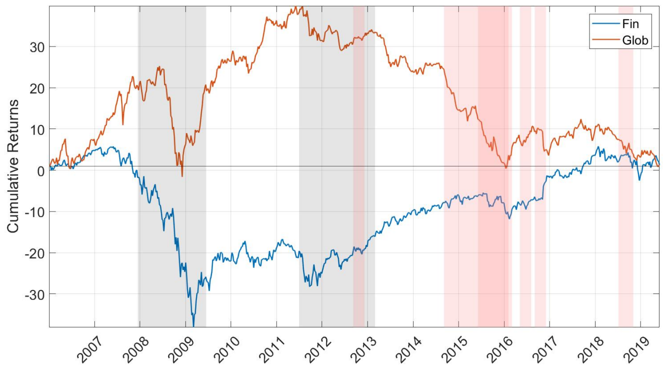
Index	$\hat{\lambda}$ (%)	Skew	Kurt	JB
S&P500 Fin	2	-0.17	17	0
S&P500	8	-0.98	12.1	0
MSCI World Fin	-2	-1.06	14.9	0
MSCI World	3	-1.42	15.1	0
EW	1	-2.08	20.4	0

$T = 700$ (6th Jan 2006 - 31st May 2019)

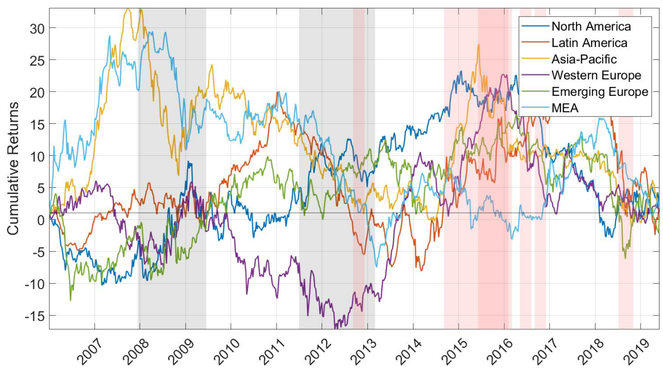
FIGURE 2.4: **Regional Model**

The figure reports the results of the factor extraction procedure via PCA for the latent factors of the Regional model. Panel 2.4a shows the time-series plot of the cumulative returns during the sample period of the financial factor and the orthogonal global factor. Panel 2.4b shows the analogous considering the six estimated local factors, one for each world region. Panel 2.4c reports the estimated in-sample summary statistics of all factors featured in the Regional model, $K = 8$. The risk premia of the latent factors are estimated as the in-sample means of the $1/N$ proxy portfolios formed by all stocks in the universe (global factor), and by the region-specific stocks (local factors) - see IR3 in Section 2.2.2 for further details. The grey bands refer to the GFC of 2007-2009 and the ESDC of 2011-2013, respectively in chronological order. The light red bars isolate the start and end dates corresponding to region-specific financial events. Details on the economic calendar are found in Appendix B.2.

(A) *Global factors*



(B) *Regional factors*



(C) *Summary Statistics*

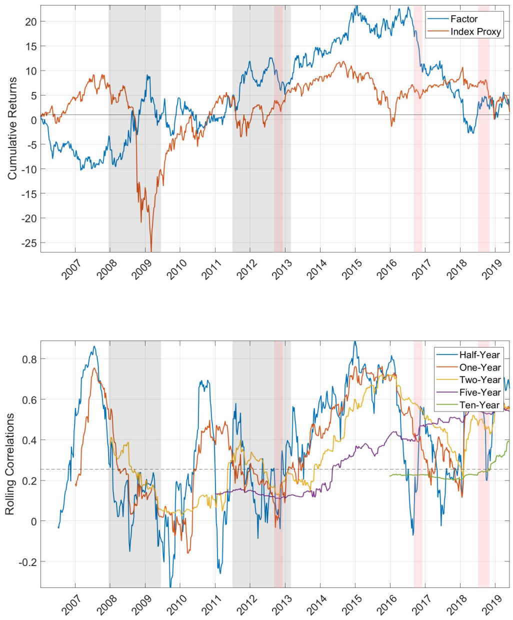
Factor	$\hat{\lambda}$ (%)	Skew	Kurt	JB	N
Financial	2	-0.4	5.5	0	
Global	3	-0.44	6.4	0	1686
North America	4	-0.27	4.4	0	231
Latin America	3	-0.4	6.5	0	217
Asia-Pacific	7	-0.15	4.9	0	484
Western Europe	1	-0.06	4.7	0	517
Emerging Europe	-4	-0.78	6.2	0	205
MEA	6	-0.18	5.5	0	32

$T = 700$ (6th Jan 2006 - 31st May 2019)

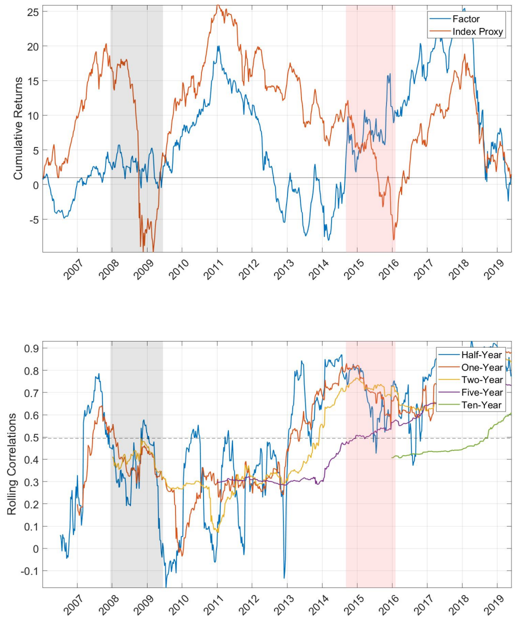
FIGURE 2.5: **Mapping Regional Factors to Equity Indeces**

The figure reports the results of the mapping exercise of the estimated latent local factors to the equity index proxies constructed as the $1/N$ portfolios of all stocks in the relevant region. For each of the six world regions considered, the upper panels show the cumulative in-sample returns of the factors and the index proxy, and the lower panels report the estimated rolling correlations for varying window size. In this exhibit I report the results for North America, panel 2.5a, Latin America, panel 2.5b, Western Europe, panel 2.5c, and Emerging Europe, 2.5d, Asia-Pacific, panel 2.5e, and MEA, panel 2.5f. Panel 2.5g reports the estimated average correlations and their significance based on the time-varying estimates, together with the full-sample OLS estimates of linear dependence. * denotes significance at 10% level, ** at 5% level, and *** at 1% level.

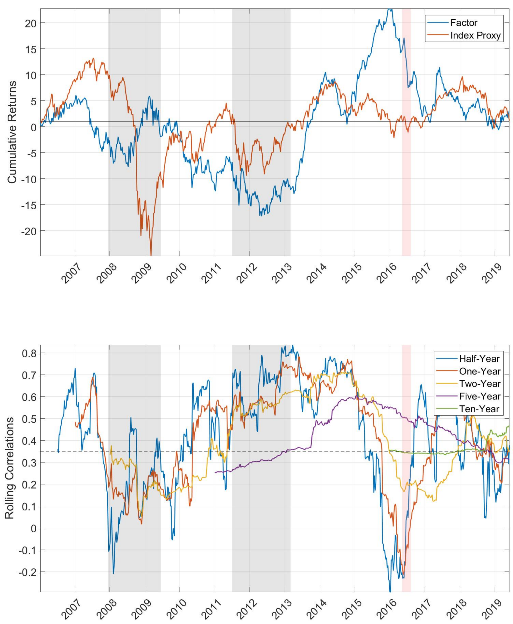
(A) *North America*



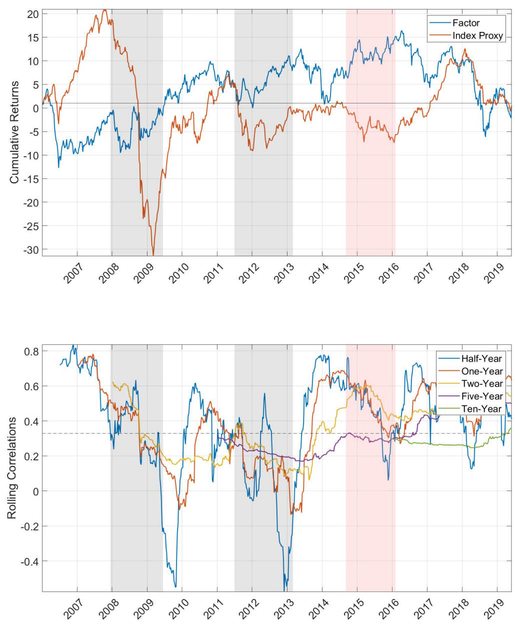
(B) *Latin America*



(C) *Western Europe*



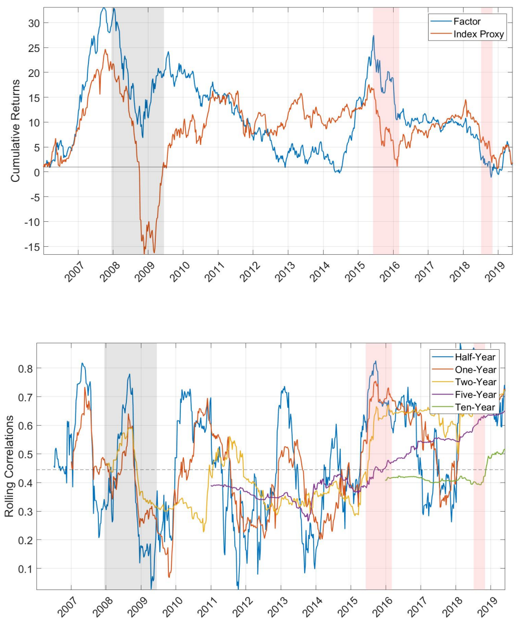
(D) *Emerging Europe*



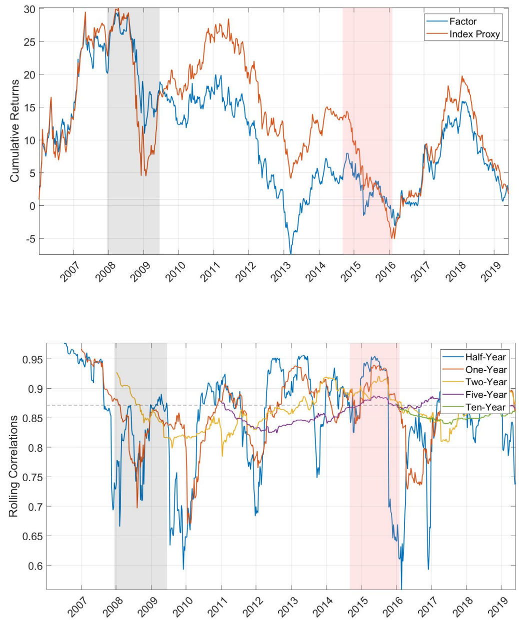
(To be continued)

(Continued)

(E) *Asia-Pacific*



(F) *MEA*



(G) *Summary Statistics*

Window Size	North America	Latin America	Asia-Pacific	Western Europe	Emerging Europe	MEA
Half-Year	0.34	0.54**	0.49**	0.41	0.37	0.85***
One-Year	0.34**	0.55***	0.48***	0.4**	0.37	0.86***
Two-Year	0.32**	0.53***	0.47***	0.38**	0.35**	0.86***
Five-Year	0.31***	0.5***	0.44***	0.39***	0.31***	0.86***
Ten-Year	0.27***	0.48***	0.42***	0.33***	0.27***	0.85***
OLS	0.27	0.51	0.44	0.32	0.31	0.86

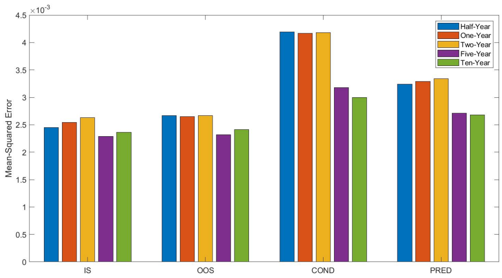
$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

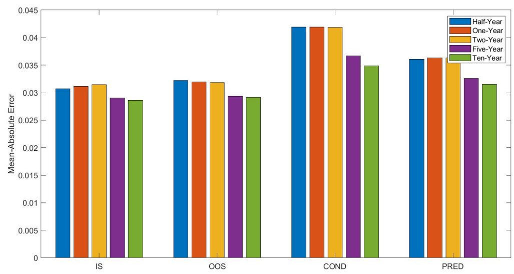
FIGURE 2.6: **Performance Measures - MKT Model**

The figure reports N-averages of the performance measures for the estimated rolling OLS betas of the MKT model, considering five different windows made of the most-recent half-, one-, two-, five- and ten-years of data. The performance measures are asset- and time-specific and I report the aggregate results across time first, and then assets. Panel 2.6a shows the bar plots corresponding to the in-sample, out-of-sample, conditional, and predictive MSE measures (respectively from left to right) for different window lengths. The blue bars refer to the half-year, the green to the one-year, the yellow to the two-year, the purple to the five-year, and the green to the ten-year window-specific results. Similarly, panel 2.6b reports the MAE measures, and panel 2.6c the R^2 measures based on Kelly, Palhares, and Pruitt (2021). Finally, panel 2.6d shows the estimated beta magnitude, standard deviation and absolute value of the t-statistics for the estimates at varying sampling frequency. The statistics in this panel are calculated from the results of the time-series regressions for varying window sizes and are thus T_W-specific (out-of-sample). The last row in panel 2.6d shows the respective in-sample OLS results. To ease the comparison across model specifications, I report the performance measures only for the 1686 tickers with complete price series in in the 13 years considered.

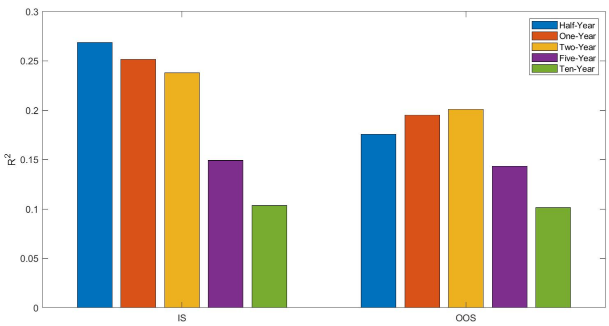
(A) *Mean-Squared Error*



(B) *Mean-Absolute Error*



(C) R^2



(D) *Summary Statistics*

	Market			
Window Size	β^{mkt}	Std	t-stat	R^2
Half-Year	0.435	0.21	2.8	0.2562
One-Year	0.445	0.162	4.0	0.2476
Two-Year	0.457	0.127	5.8	0.2502
Five-Year	0.468	0.084	9.3	0.2543
Ten-Year	0.485	0.021	13.7	0.2655
OLS	0.475	0.033	15.3	0.2529

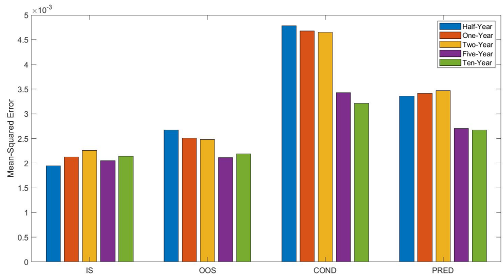
$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

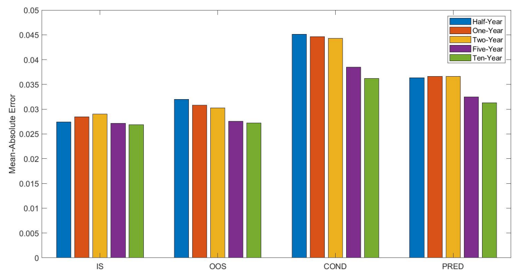
FIGURE 2.7: **Performance Measures - Regional Model**

The figure reports N-averages of the performance measures for the estimated rolling OLS betas of the regional model, considering five different windows made of the most-recent half-, one-, two-, five- and ten-years of data. The performance measures are asset- and time-specific and I report the aggregate results across time first, and then assets. Panel 2.7a shows the bar plots corresponding to the in-sample, out-of-sample, conditional out-of-sample, and predictive MSE measures (respectively from left to right) for different window lengths. The blue bars refer to the half-year, the green to the one-year, the yellow to the two-year, the purple to the five-year, and the green to the ten-year window-specific results. Similarly, panel 2.7b reports the MAE measures, and panel 2.7c the R^2 measures based on Kelly, Palhares, and Pruitt (2021). Finally, panel 2.7d shows the estimated beta magnitude, standard deviation and absolute value of the t-statistics for the estimates at varying sampling frequency. The statistics in this panel are calculated from the results of the time-series regressions for varying window sizes and are thus T_W-specific (out-of-sample). The last row in panel 2.7d shows the respective full-sample OLS results. To ease the comparison across model specifications, I report the performance measures only for the 1686 tickers with complete price series in in the 13 years considered.

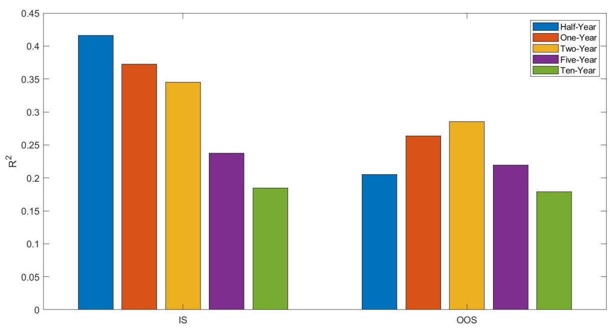
(A) *Mean-Squared Error*



(B) *Mean-Absolute Error*



(C) R^2



(D) *Summary Statistics*

	Financial			**Global**			**Regional**			
Window Size	β^{fin}	Std	t-stat	β^{glob}	Std	t-stat	β^{reg}	Std	t-stat	R^2
Half-Year	0.305	0.224	1.9	0.309	0.204	1.9	0.211	0.199	1.4	0.4042
One-Year	0.308	0.163	2.8	0.312	0.143	2.8	0.212	0.137	2	0.367
Two-Year	0.318	0.122	4.3	0.313	0.101	4	0.212	0.095	2.9	0.3565
Five-Year	0.319	0.08	6.9	0.303	0.057	6.3	0.209	0.052	4.5	0.3484
Ten-Year	0.343	0.022	10.5	0.314	0.021	9.3	0.21	0.016	6.5	0.3536
OLS	0.342	0.03	12	0.321	0.03	10.9	0.211	0.03	7.5	0.3433

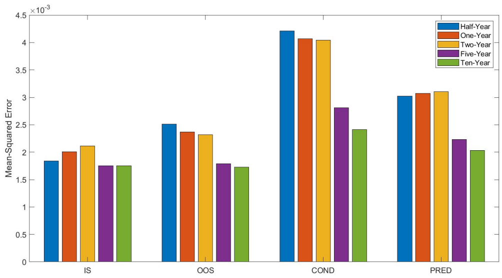
$N = 1686$

$T = 700$ (6th Jan 2006 - 31st May 2019)

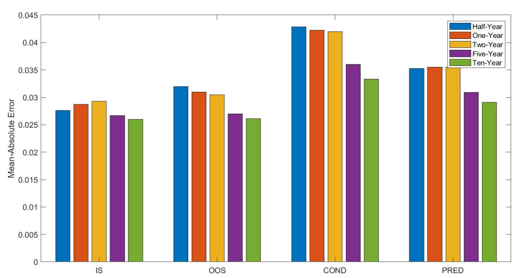
FIGURE 2.8: **Performance Measures - FF3 Model**

The figure reports N-averages of the performance measures for the estimated rolling OLS betas of the FF3 model, considering five different windows made of the most-recent half-, one-, two-, five- and ten-years of data. The performance measures are asset- and time-specific and I report the aggregate results across time first, and then assets. Panel 2.8a shows the bar plots corresponding to the in-sample, out-of-sample, conditional out-of-sample, and predictive MSE measures (respectively from left to right) for different window lengths. The blue bars refer to the half-year, the green to the one-year, the yellow to the two-year, the purple to the five-year, and the green to the ten-year window-specific results. Similarly, panel 2.8b reports the MAE measures, and panel 2.8c the R^2 measures based on Kelly, Palhares, and Pruitt (2021). Finally, panel 2.8d shows the estimated beta magnitude, standard deviation and absolute value of the t-statistics for the estimates at varying sampling frequency. The statistics in this panel are calculated from the results of the time-series regressions for varying window sizes and are thus T_W-specific (out-of-sample). The last row in panel 2.8d shows the respective full-sample OLS results. To ease the comparison with the Regional and MKT models, I report the performance measures for 1232 (out of 1686) tickers with complete price series that remain listed the national equity indeces making up the North America, Asia-Pacific and Western Europe regions. Details on the regional classification of FF are found in Appendix B.3.

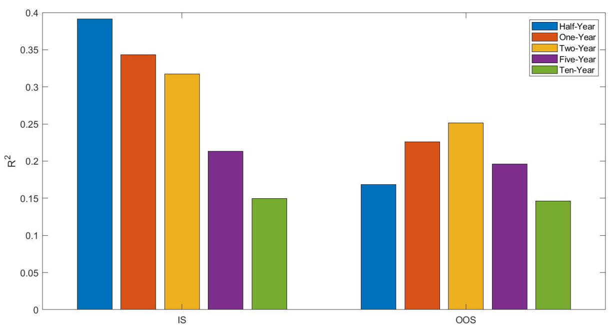
(A) *Mean-Squared Error*



(B) *Mean-Absolute Error*



(C) R^2



(D) *Summary Statistics*

	MKT			SMB			HML			
Window Size	β^{MKT}	Std	t-stat	β^{SMB}	Std	t-stat	β^{HML}	Std	t-stat	R^2
HalfYear	0.475	0.241	2.7	0.009	0.207	1	-0.005	0.214	1.1	0.3692
OneYear	0.481	0.17	3.8	0.006	0.133	1	-0.008	0.144	1.2	0.3308
TwoYear	0.501	0.123	5.7	-0.002	0.087	1.2	-0.008	0.099	1.5	0.3242
FiveYear	0.527	0.073	9.5	-0.008	0.044	1.5	0.002	0.057	2	0.3315
TenYear	0.551	0.021	14.3	-0.004	0.016	1.8	0.01	0.02	2.7	0.3496
OLS	0.526	0.037	15	0.038	0.034	2.4	0.032	0.034	2.7	0.3117

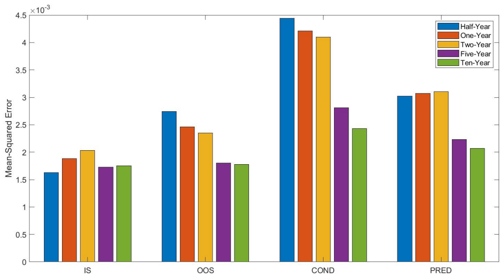
$N = 1232$

$T = 700$ (6th Jan 2006 - 31st May 2019)

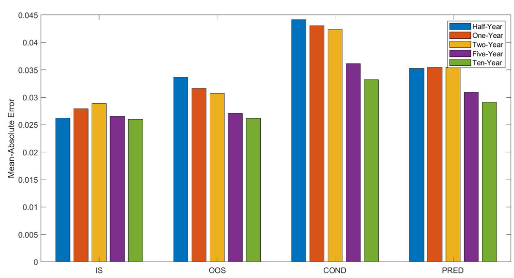
FIGURE 2.9: **Performance Measures - FF5 Model**

The figure reports N-averages of the performance measures for the estimated rolling OLS betas of the FF3 model, considering five different windows made of the most-recent half-, one-, two-, five- and ten-years of data. The performance measures are asset- and time-specific and I report the aggregate results across time first, and then assets. Panel 2.9a shows the bar plots corresponding to the in-sample, out-of-sample, conditional out-of-sample, and predictive MSE measures (respectively from left to right) for different window lengths. The blue bars refer to the half-year, the green to the one-year, the yellow to the two-year, the purple to the five-year, and the green to the ten-year window-specific results. Similarly, panel 2.9b reports the MAE measures, and panel 2.9c the R^2 measures based on Kelly, Palhares, and Pruitt (2021). Finally, panel 2.9d shows the estimated beta magnitude, standard deviation and absolute value of the t-statistics for the estimates at varying sampling frequency. The statistics in this panel are calculated from the results of the time-series regressions for varying window sizes and are thus T_W-specific (out-of-sample). The last row in panel 2.9d shows the respective full-sample OLS results. To ease the comparison with the Regional and MKT models, I report the performance measures for 1232 (out of 1686) tickers with complete price series that remain listed the national equity indeces making up the North America, Asia-Pacific and Western Europe regions. Details on the regional classification of FF are found in Appendix B.3.

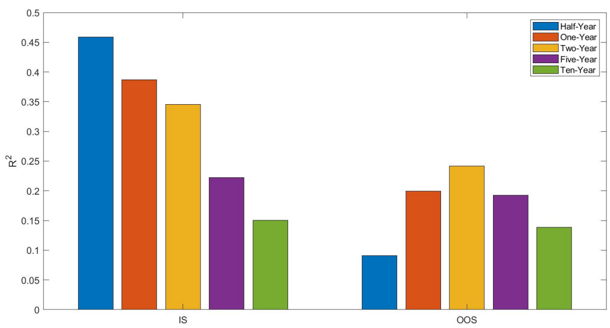
(A) *Mean-Squared Error*



(B) *Mean-Absolute Error*



(C) R^2



(D) *Summary Statistics*

	MKT			SMB			HML			RMW			CMA			
Window Size	β^{MKT}	Std	t-stat	β^{SMB}	Std	t-stat	β^{HML}	Std	t-stat	β^{RMW}	Std	t-stat	β^{CMA}	Std	t-stat	R^2
HalfYear	0.488	0.275	2.4	0.023	0.225	0.9	0.017	0.316	1	0.043	0.27	1	0.019	0.289	1	0.4466
OneYear	0.492	0.185	3.5	0.018	0.137	1	0.014	0.198	1.1	0.043	0.166	1	0.015	0.185	1.1	0.3774
TwoYear	0.511	0.13	5.2	0.01	0.088	1.1	0.016	0.131	1.3	0.042	0.108	1.2	0.011	0.125	1.3	0.3561
FiveYear	0.536	0.076	9	0.004	0.044	1.3	0.028	0.069	1.7	0.047	0.052	1.5	0.006	0.061	1.8	0.3525
TenYear	0.56	0.022	13.3	0.007	0.016	1.7	0.041	0.022	2.4	0.052	0.021	2.1	-0.003	0.024	2.1	0.3659
OLS	0.519	0.041	13.2	0.039	0.035	2.3	0.044	0.046	2.1	0.012	0.042	1.6	-0.017	0.039	2	0.3212

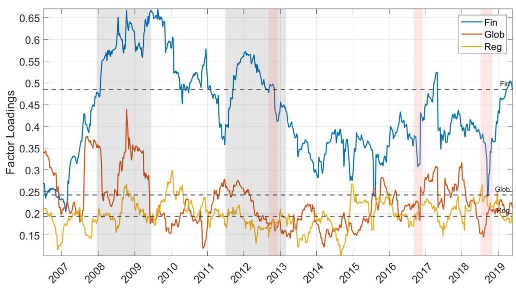
$N = 1232$

$T = 700$ (6th Jan 2006 - 31st May 2019)

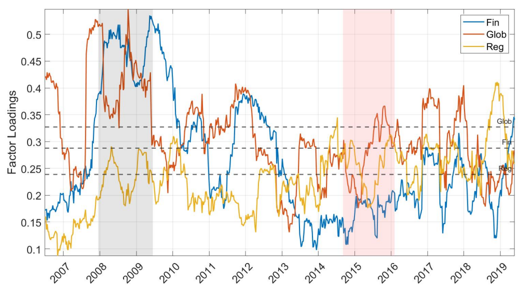
FIGURE 2.10: **Dynamic Loadings - Regional Model**

The figure shows the estimated rolling factor sensitivities for the Regional model using a half-year window, panels 2.10a-2.10f, and a two-year window, panels 2.10g-2.10l. For each region, I report the time series evolution of the loadings of the financial factor, blue line, global factor, red line, and regional factor, yellow line. The shaded horizontal lines indicate OLS full-sample estimates of the loadings. The grey bands refer to the GFC of 2007-2009 and the ESDC of 2011-2013, respectively in chronological order. The light red bars isolate the start and end dates corresponding to region-specific financial events. Details on the economic calendar are found in Appendix B.2.

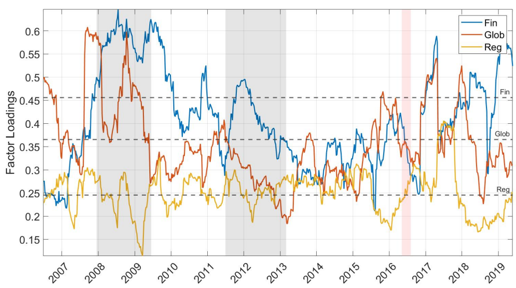
(A) *North America - Half-Year*



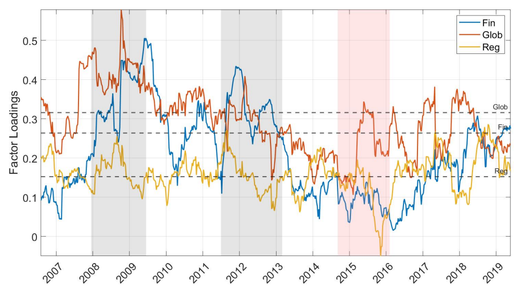
(B) *Latin America - Half-Year*



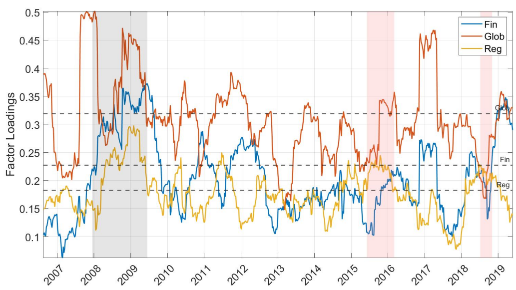
(C) *Western Europe - Half-Year*



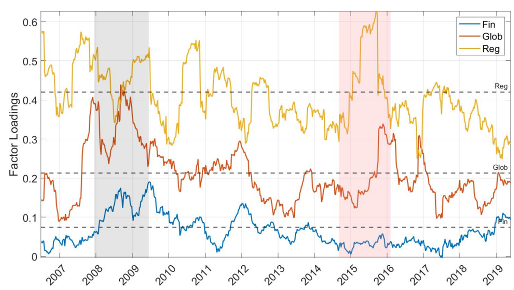
(D) *Emerging Europe - Half-Year*



(E) *Asia-Pacific - Half-Year*



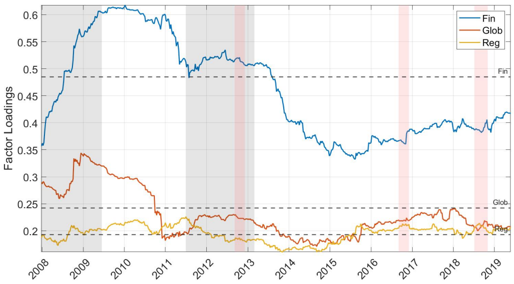
(F) *MEA - Half-Year*



(To be continued)

(Continued)

(G) *North America - Two-Year*

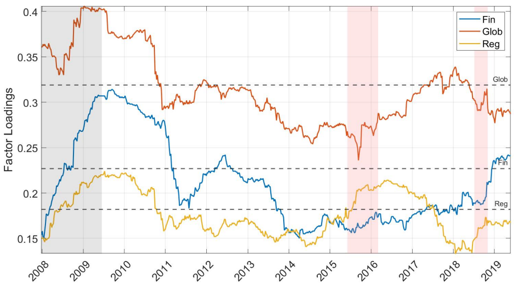


(H) *Latin America - Two-Year*

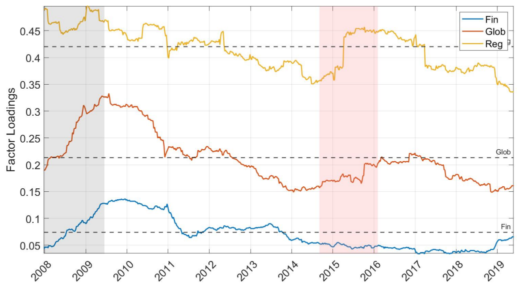
(I) *Western Europe - Two-Year*

(J) *Emerging Europe - Two-Year*

(K) *Asia-Pacific - Two-Year*



(L) *MEA - Two-Year*



$N = 1686$

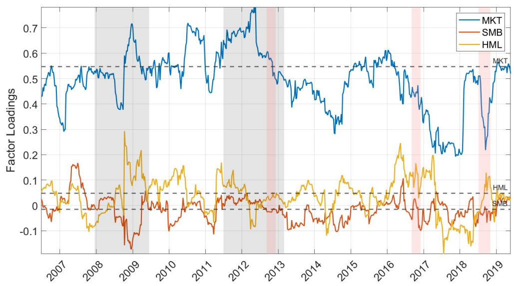
$T = 674$ (half-year, 30th Jun 2006 - 31st May 2019)

$T = 596$ (two-year, 28th Dec 2007 - 31st May 2019)

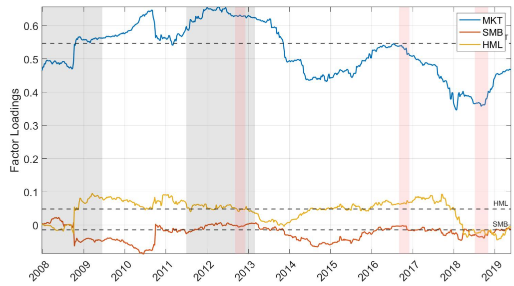
FIGURE 2.11: **Dynamic Loadings - FF3 Model**

The figure shows the estimated rolling factor sensitivities for the FF3 model using a half-year window, left panels, and a two-year window, right panels. For each region, I report the time series evolution of the loadings of the MKT factor, blue line, SMB factor, red line, and HML factor, yellow line. The shaded horizontal lines indicate OLS full-sample estimates of the loadings. The grey bands refer to the GFC of 2007-2009 and the ESDC of 2011-2013, respectively in chronological order. The light red bars isolate the start and end dates corresponding to region-specific financial events. Details on the economic calendar are found in Appendix B.2. To ease the comparison with the Regional and MKT models, I report the estimates for 1232 (out of 1686) tickers with complete price series that remain listed the national equity indeces making up the North America, Asia-Pacific and Western Europe regions. Details on the regional classification of FF are found in Appendix B.3.

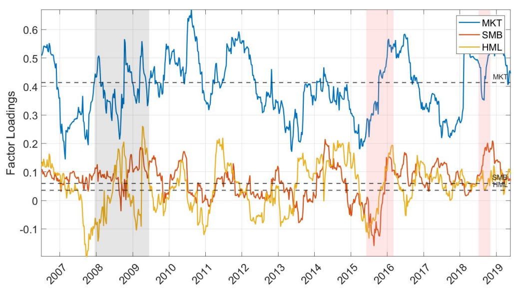
(A) *North America, Half-Year*



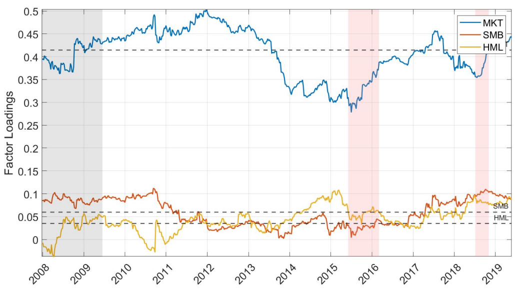
(B) *North America, Two-Year*



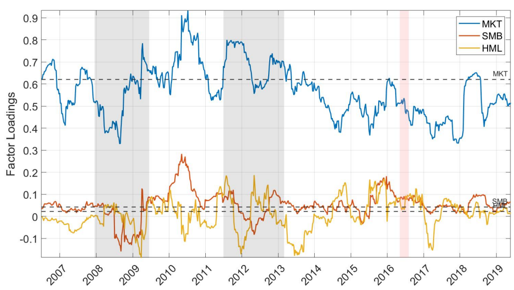
(C) *Asia-Pacific, Half-Year*



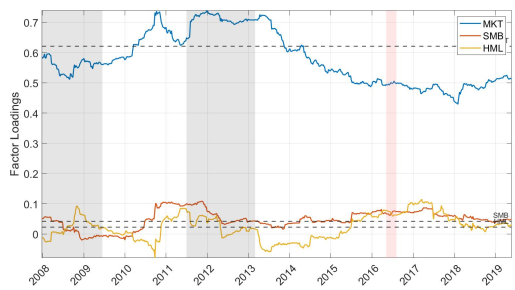
(D) *Asia-Pacific, Two-Year*



(E) *Western Europe, Half-Year*



(F) *Western Europe, Two-Year*



$N = 1232$

$T = 674$ (half-year, 30th Jun 2006 - 31st May 2019)

$T = 596$ (two-year, 28th Dec 2007 - 31st May 2019)

Chapter 3

Factor Investing in the Sovereign Bond Market

Abstract

In this chapter I study the performance of factor-based investment strategies such as momentum, value, and low-risk in the government bond market. My universe comprises bonds issued by nine developed countries in the period January 2010 to end of October 2021, with data available at daily frequency. I examine the cross-country characteristics of the factors, and how these change with respect to the maturity buckets along the issuer-specific curves. My analysis reveals that momentum yields high and statistically significant Sharpe ratios across most countries only when short-dated bonds (less than five years) are included. On the other hand, using value measures based on past returns as in Asness, Moskowitz, and Pedersen (2013) leads to low and insignificant risk-adjusted returns across countries. Although lower in magnitude than momentum, low-risk yields statistically significant returns only for Euro Area bonds. Contrarily to what reported in previous literature, I find no supporting evidence for momentum, value and low-risk when bonds across all countries are considered in a global portfolio.

3.1 Introduction

The literature on factor investing[1] has focused primarily on examining the performance of traditional style factors, such as momentum, value, and low-risk, in explaining future return patterns in the equity and credit markets. Asness, Moskowitz, and Pedersen (2013)[AMP13 thereafter] use a consistent approach across asset classes and draw from the equity literature, e.g. De Bondt and Thaler (1985), Fama and French (1996a), to define momentum and value. Momentum factors benefit from the continuation of short-term trends in the market prices of securities, and are usually constructed using past-return performance measures, while value factors aim to profit from long-term deviations of asset prices relative to their 'fair value', the calculation of which is based on accounting figures in the equity market, or yield curve spreads in the credit market. Frazzini and Pedersen (2014)[FP14 thereafter] also follow a consistent approach across asset classes, and define low-risk[2] as the ability for assets with safer and more consistent return streams, measured by the market beta in stocks or duration in bonds, to deliver higher risk-adjusted performance.

To date, relatively less attention has been given to analyse the characteristics of traditional style factors in the sovereign bond market. AMP13 is one of the first in the literature to do so, and find consistent evidence of return premia for the factors individually, as well as significant diversification benefits when combined in a multi-factor portfolio. They consider a sample of ten countries, with data at monthly frequency, and report their results on aggregate across all countries. FP14 follow an analogous approach but test their low-risk factor only in the US sovereign bond market, for which they report statistically significant risk-adjusted returns. More recently, several contributions have been made to study factor premiums in global government bond markets, eg. Baltussen, Martens, and Penninga (2021), Kunz and Mazzoleni (2018), and Brightman and Shepherd (2016), and I build on this literature by questioning two key features of these studies. Firstly, they report results on aggregate across countries or maturity buckets, which makes it difficult to understand how the factors' performance change in the cross-section (country- or maturity-wise). This is mainly due to the fact that these studies use portfolios as primitive assets, which inevitably leads to information being averaged-out across one dimension (global portfolios for the country dimension, and constant-maturity portfolios for the maturity dimension). Secondly, they all employ data sampled at monthly frequency, which raises questions on the validity of their results at higher frequencies such as daily.

In this chapter I aim to bridge this gap by analysing the cross-country and -maturity differences in the performance of style factors formed on bond-level data at daily frequency, from nine developed countries. I follow AMP13 to construct cross-sectional momentum and

[1]The term 'factor investing' is a used as a synonym for 'style investing', similarly for (style) 'factors' and 'portfolios'.

[2]Also know as 'defensive' or 'quality' factor.

value factors for each country and maturity combination, and test alternative measures based on shorter estimation windows to capture fluctuations in the bonds' prices at daily frequency. Similarly for low-risk, I follow FP14 as baseline and compare maturity- versus duration-based measures. Conditional on a given measure, I compare the factors' performance under different dimensions. I test global factors that allocate capital internationally, i.e. aggregating results across countries as customary in the literature, against the country-specific counterparts. I test factors that invest in bonds with different maturity profiles, i.e. long- vs short-dated, and factors that are build under different investment constraints, i.e. long-only vs long-short.

My results indicate a high degree of variability in the factors' performance across countries or maturity buckets. I find that by using canonical momentum measures as in AMP13, the resulting portfolio returns are decreasing in the maturity of the bonds, with the highest Sharpe ratios found for portfolios formed on short-dated bonds (less than 5-year maturity). For longer-dated securities, momentum does not deliver statistically significant portfolio returns. Comparing across countries, my results reveal that momentum produces consistent statistically significant Sharpe ratios, however this is not true for value and low-risk. Reversal value measures of AMP13 fail to produce portfolios that have positive performance across countries and maturity buckets, while low-risk yields the best results when bonds issued by Euro Area countries are considered for portfolio formation.

I find an unambiguously better performance of factors constructed as long-short over long-only portfolios, with the latter often yielding negative statistically significant Sharpe ratios (in particular for momentum). Similarly, when I allow portfolios to allocate capital internationally, I find that for all factors this leads to a substantially lower performance with respect to the country-specific counterparts. This result is in contrast to what is commonly reported in the literature, and suggests that style factor premia vary substantially across markets when bonds are used as primitive assets. Although this phenomenon is well understood in the equity, e.g. Fama and French (2012), and credit markets, e.g. Bekaert and De Santis (2021), to the best of my knowledge my study is the first to document the performance of local (country-specific) style factors in the sovereign bond market. Based on my findings, I argue that aggregating results across countries, as it is customary in the literature, can lead to biased and inaccurate results as much of the cross-country variability is averaged out.

When I analyse the cross-factor relationships in each country from October 2018 to October 2021, I find limited support for the advocated diversification benefits of momentum and value. Although not negative as reported in AMP13, my results show that their returns are not linearly related in the sample period. However, when I examine duration profile of these portfolios I argue that this result may be spurious, and can arise as a consequence of duration risk being unhedged during portfolio formation. Using the measures of AMP13 and a similar methodology, I show how long-short portfolios can have negative net maturity (or duration) in a given sample period, which indicates that the portfolio is positioned to benefit from rising yields, as opposed to a portfolio with positive net maturity that profits in a declining yields scenario. Brightman and Shepherd (2016) also acknowledges that the procedure of AMP13 can lead to net long or net short positions at any point in time, although they argue that 'those exposures mostly average out over time', which is contrast to what I report. Across countries, value appears to be uncorrelated also with low-risk, suggesting that it is an unlikely duration proxy. Low-risk portfolios tend to have negative net maturity and negative correlation with momentum, and this is true across most countries. The only exceptions are global and Euro Area country portfolios, in which the correlation between momentum and low-risk is high and positive. For the global factor in particular, I find the highest correlation estimate, which suggests that momentum may be considered a duration proxy rather than an anomaly (or premia), due to the fact that low-risk is constructed as a portfolio that invests in bonds with lowest duration (or maturity) in the cross-section at any time.

My study contributes to the recent body of literature on factor investing in the sovereign bond market under many aspects. I expand the approach of Durham (2015), who analyse how cross-sectional momentum change in the duration buckets of a single issuer's curve, to other factors (value and low-risk) and across countries. A key reference for my study is Baltussen, Martens, and Penninga (2021) who examine value, momentum, and low-risk in the sovereign bonds of 16 developed countries, using a sample that spans more than 200 years of data. While I employ similar factor construction techniques, my study extends their analysis to data sampled at higher frequency, and examines global factor premia internationally and across maturities. My contribution also differs from existing studies by considering bonds as base assets, instead of portfolios, which allows me to compare the impact of different constraints during portfolio formation in relation to the returns, risk, and duration of the resulting portfolios.

Overall, based on the evidence in this chapter, I find little support for employing unifying factor measures across countries and maturities in the sovereign bond market. However, the limitations of my study are many and in Section 3.5.1 I propose alternative procedures that expand on this study and can help to better understand global factor premia in sovereign bonds.

Organisation of the chapter. The remainder of this chapter is as follows. Section 3.2 describes the the factor replication procedure using bonds as base assets. Section 3.3 provides the details on the cross-country and -maturity composition of my international bond universe, and Section 3.4 reports the results. In the latter, I firstly study the characteristics of the factors individually, and then collectively for each country. Section 3.5 concludes my study and discuss further research. The chapter is accompanied by Appendix C.

3.2 Methodology

In this section I firstly present the baseline framework to study the cross-section of government bond returns, Section 3.2.1, and subsequently in Section 3.2.2 I detail the factor construction procedure and review alternative approaches from the literature.

3.2.1 Baseline Framework

My aim is to study the cross-section of N bonds issued in the sample period, $t \in [1, T]$, by C national governments, $c \in [1, C]$ and $i \in [1, N]$. I denote with $r_{i,t}$ the end-of-day return from $t-1$ to t on bond's i clean price $P_{i,t}^{(c)}$, $r_{i,t} = (P_{i,t}^{(c)} / P_{i,t-1}^{(c)}) - 1$, excluding any accrued interest from the most-recent coupon payment up to period t[3]. $P_{i,0}^{(c)}$ is bond's i issue price, and $P_{i,T_i}^{(c)}$ its price at maturity. I work in a framework in which $T > N$, although the data matrix $\boldsymbol{r}$ is sparse given the different characteristics (issue date and maturity) of the bonds. This means that the cross-sectional dimension is time-varying, and N_t refers to the number of outstanding bonds at the end of each day in the sample.

$$\underset{(T\times N)}{\boldsymbol{r}} = \begin{bmatrix} r_{1,1} & 0 & 0 & \dots & 0 \\ r_{1,2} & r_{2,1} & 0 & \dots & r_{N,1} \\ \vdots & r_{2,2} & r_{3,1} & \dots & r_{N,2} \\ r_{1,T_1} & \vdots & \vdots & \dots & \vdots \\ 0 & r_{2,T_2} & r_{3,T_3} & \dots & \vdots \\ \vdots & 0 & 0 & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & r_{N,T} \end{bmatrix}$$

I group the bonds N_t into four maturity buckets depending on their term-to-maturity $\Delta T_{i,t} = t - T_i$. My partition follows Brooks and Moskowitz (2017) who examine a set of tradeable bonds covered in the JP Morgan Government Bond Index made of the the most liquid securities across 13 developed markets.[4] Bucket 0 comprise bonds that have less than one year remaining before maturity, bucket 1 from one to five years, bucket 2 from five to ten years, and bucket 3 the bonds with more than 11 years until maturity (up to a maximum of 30 years), $B \in \{0, 1, 2, 3\}$. In total the number of on-the-run bonds at time t is given by $N_t = \sum_c^C \sum_b^B N_{t,c,b}$ where $N_{t,c,b}$ is the bucket- and country-specific total number of securities available.

[3]Similarly to dividend gains in the analysis of single stock returns, I exclude any gains earned from the accumulated interest on each bond, and focus solely on the movements in bond prices related to interest rate changes.

[4]Differently to Brooks and Moskowitz (2017), I also include securities that have less than one year remaining until maturity (bucket 0).

3.2.2 Factor Construction

I construct factor-based investment portfolios following a procedure similar to that used by AMP13, FP14, and Koijen et al. (2018). For any bond $i = 1, ..., N_t$ at time t with signal $S_{i,t}$, I take a position equal to the signal rank minus the cross-sectional average rank of that signal

$$w_{i,t}^{LS} = z_t \left(rank(S_{i,t}) - \frac{N_t + 1}{2} \right) \tag{3.1}$$

with z_t being a scaling factor such that the overall portfolio is scaled to one dollar long and one dollar short. The weights across all bonds at time t sum to zero, representing a dollar-neutral long-short portfolio. In a similar fashion I construct long-only (smart-beta) portfolios based on the signal rank as follows

$$w_{i,t}^{L} = z_t \; rank(S_{i,t}) \tag{3.2}$$

where positions for all assets sum to one at each time t. For the country-style portfolios[5] I use as base assets the bucket-b bonds trading at time t in each market c, $i = 1, ..., N_{t,c,b}$, and for the global portfolio I consider bonds across issuers, $i = 1, ..., \sum_c^C N_{t,c,b}$ for bucket b.

To account for the different maturities of the bonds in the country-bucket portfolios, I adjust the weights $w_{i,t} \in \{w_{i,t}^{LS}, w_{i,t}^{L}\}$ by the term-to-maturity $\Delta T_{i,t}$, delivering another set of weights $\bar{w}_{i,t} = w_{i,t}/\Delta T_{i,t}$. This is equivalent to a strategy that down-weights bonds with a relatively high maturity within each country-bucket portfolio, and up-weights bonds with low maturity, in a similar spirit to low-risk in FP14. Especially for buckets 0, 1, and 2, this adjustment does not lead to substantially different allocations with respect to the signal-based weights $w_{i,t}$ since the bonds can have at maximum a 5-year maturity difference. On the other hand, it is more substantial for bucket 3 bonds that can have a 19-year difference, and in this case it helps me prevent the impact of signals from bonds with extremely high residual maturity on my allocations. The adjusted weights are normalised to respect the dollar-neutrality and fully-investment conditions for long-short and long-only portfolios respectively. ${f_t}^{(c,b)}$ is the country-bucket portfolio return at time t, built as a linear combination of asset asset returns $r_{i,t}$ with $\bar{w}_{i,t}$,

$${f_t}^{(c,b)} = \sum_{i}^{N_{t,c,b}} \bar{w}_{i,t} \; r_{i,t}. \tag{3.3}$$

My approach draws from Durham (2015), who study how momentum patterns change on the duration buckets of a single issuer, and from AMP13, who examine value and momentum for the cross-section of country bond returns (isolating winners and losers across countries, disregarding the maturity dimension). Durham (2015) argues that the allocations in AMP13

[5]In practice, I require a minimum of five bonds to be traded at each time t for the construction of the country-bucket portfolios, $N_{t,c,b} \geq 5$.

expose investors to duration risk, as average duration differs among the markets that make up the global portfolio, and constructs long-only portfolios duration-neutral indeces for each of the six maturity buckets considered (1-3, 3-5, 5-7, 7-10, 10-20, and 20-30 years). Differently to AMP13, the portfolio construction procedure in Durham (2015) does not feature discrete rankings of the individual asset, rather it is the allocation across the six buckets that yields the greatest past return under two constraints. The first is long-only fully-invested portfolios, and the second is that the optimal weights produce portfolios with the same contemporaneous weighted-average duration as the benchmark. The latter is the overall US Treasury index formed by all on-the-run bonds at each time t.

From a methodological perspective, my study differs from Durham (2015) under several aspects: I use bonds as base assets, instead of constant-maturity bond portfolios, I rank bonds based on their cross-sectional signal ranks (issuer- and maturity-wise), and I do not duration-match my portfolios against a benchmark. This last feature is particularly relevant for my long-short portfolios, as they can show negative net duration at any time t in my setup. Although the weights $\bar{w}_{i,t}^{LS}$ sum to zero, with $\sum_i^{N_{t,c,b}} \bar{w}_{i,t}^{LS} \ I(\bar{w}_{i,t}^{LS} > 0) = 1 \ \forall t$ for the long leg and similarly $\sum_i^{N_{t,c,b}} \bar{w}_{i,t}^{LS} \ I(\bar{w}_{i,t}^{LS} < 0) = -1 \ \forall t$ for the short leg, the bonds that make up the latter can have higher average duration than those of the long leg (at any point in time). This results in a portfolio that 'over-hedges' its long exposure by selling higher-duration securities, and thus increases in value when interest rates rise, or in a portfolio with net positive duration, which is positioned to benefit from falling interest rates, with its long leg having higher duration than the short leg.

In my study I work only with traded bond data, which makes it difficult to construct duration-hedged portfolios due to lack of available securities across maturity buckets for each issuer at any point in time, see Section 3.3 for further details on the data used. An alternative approach is to bootstrap the issuer-specific spot curves and price the synthetic coupon bonds for all (continuous) maturities, however, due to the variety in the characteristics of the sovereign bonds issued in the sample period across countries (e.g. day-count, reset and payment frequency, coupon rates, etc.), this requires an accurate pricing engine, a research topic that I leave for future studies, see Section 3.5.1. Although I acknowledge that duration risk is not accounted for in my factor construction procedure, I construct style factors across all available maturity buckets (for each issuer) and examine their relative performance. For long-short portfolios, the factor measures can lead to imbalances in the duration profile of the bond portfolios and I compare against long-only benchmarks (with net positive duration) that feature the same securities for each country-maturity combination, conditional on a given measure.

To tame the effect of passage of time on bond prices, I exclude the securities that mature in less than 30 days for the construction of all portfolios at each time t. This choice is motivated by the work of Brusa, Gu, and Liu (2014), who show that as premium (discount) bonds approach maturity their prices will move lower (higher) over time, and this decrease (increase) will accelerate as they due date gets closer. The relationship between coupon-yield and level of interest rates determine if a bond is trading at premium, discount, or par, and it is time-varying. Based on their results one can argue that sorting bonds based on their deviation to par price at time t can be a candidate risk factor, especially for the securities that are relatively 'close' to maturity, a research question that I leave for further studies, see Section 3.5.1.

In what follows I describe the definitions of the signals $S_{i,t}$ that I use to construct the cross-sectional factors: momentum, value, and low-risk.

Momentum

My definition of momentum follows AMP13, who use the past 12-month cumulative asset's return, skipping the most-recent month, 12-1M. The choice of the window length in momentum strategies has been the subject of numerous studies in the asset pricing literature that analyse stock returns' anomalies. See for instance De Bondt and Thaler (1985), who find that losers (poor past performance) tend to have high future returns and winners (high past performance) low future returns, when portfolios are formed on long-term (3- to 5-year) past returns. Jegadeesh and Titman (1993) study the relative performance of 16 momentum strategies with 3- to 12-month horizons from 1965 to 1985, and find evidence of persistence positive returns when portfolios are formed on short-term (up to a year of) past returns: past losers tend to be future losers, and past winners future winners. Similarly, Fama and French (1996a) confirm the strong continuation of short-term returns in the period 1963 to 1993, and find that average returns for the month after portfolio formation are near-zero for the stocks with the worst short-term past performance (measured from 12 to 2 months before portfolio formation), and about 1.3% annualised for the (decile of the) stocks with highest short-term past return. They also corroborate that average returns tend to reverse when momentum portfolios are formed considering long-term returns from 60 to 13 months prior the date of formation. This is particularly relevant in the 1930-1961 period, which approximately matches the sample of De Bondt and Thaler (1985) from 1932 to 1977, but fails to be consistent in the 1963-1993 period. During this time, long-term reversal (i.e. low past performance leads to high future returns) is observed only when the year prior to portfolio formation is skipped in ranking stocks. When the preceding year is included, short-term continuation (i.e. low past performance leads to low future returns) prevails over long-term reversal, and past losers tend to have lower future returns than past winners for portfolios formed using a long window (up to four years of past data).

Although continuation of momentum strategies is shown to be particularly relevant in the short-term (less than 12-month returns), Fama and French (1996a) always skip the preceding two months prior to the portfolio sorts, arguing that this procedure reduces the bid-ask bounce. This choice is motivated by Jegadeesh (1990), who find statistically significant negative first-order serial correlation in monthly stock returns in the period 1934-1987, and Jegadeesh and Titman (1993), who report that holding period returns for their portfolios are 'slightly higher' when there is a 1-week lag between the formation and holding periods. Both papers attribute this phenomenon to poor liquidity conditions or market microstructure biases, an interpretation that finds widespread support in the literature analysing stock returns. In the same spirit of AMP13, in this study I maintain a simple approach that is consistent across asset classes, and I leave the problem of adapting existing measures commonly used in the equity market to the fixed-income literature for future studies. This includes the calculation of the exclusion window between signal calculation and investment phase, which I keep fixed at one month. Let aside any explanation of this phenomenon in the government bond market, which is surely less liquid compared to equity market in which trading is regulated by a centralised exchanged, skipping a month in forming the portfolios improves the out-of-sample reliability of the signal measure and reduces potential instabilities in short-term (less than one month) returns, a logic that is also shared by Grinblatt and Moskowitz (2004).

Additionally, Brightman and Shepherd (2016) study global momentum strategies using 10-year futures from eight developed markets[6] and the past 12-month return as signal measure. Their momentum portfolios are built as time-series factors in which each contract is compared with its own history, and the strategies invest in bonds with positive past return measure and short those with a negative one. This is different to my setup in which I calculate the time-series momentum signals of each asset considering its own history, but use the relative rank of these signals for the issuer- and bucket-specific bonds at each time t to form the cross-sectional portfolios. Brightman and Shepherd (2016) report an average momentum portfolio return of 2.9% and a Sharpe ratio of 0.7 in the period from January 1989 to June 2016 (monthly frequency), and find evidence of a negative relationship between value and momentum factors (-0.18) as in AMP13, and a positive relationship between momentum and carry (0.2).

Brooks and Moskowitz (2017) study momentum across countries using level, slope, and curvature portfolios. The latter are defined respectively as the portfolios formed by the 10-year bond in each country, the 10-year minus 2-year bonds (duration-adjusted), and the 5-year bond minus an equal-duration weighted average of the 2- and 10-year bonds. They then define momentum signals using the past 12-month return on these portfolios, and find that they deliver statistically significant alphas in the cross-sectional regression of level portfolio (excess) returns on the first three principal components of the yield curve. This results indicate that momentum is not related to the cross-section of country-level returns (i.e. it does not affect any of the PCs calculated from the yield curve), contrarily to value and carry that are captured

[6]Brightman and Shepherd (2016) consider a cross-section of countries that is the same as ours with the exception of Spain which is excluded in their study.

respectively by PC1 and PC2. Since principal components are not sufficient to explain the variation in momentum, Brooks and Moskowitz (2017) add two macroeconomic variables (growth in industrial production and inflation of each country) in their cross-sectional regressions of portfolio excess returns, and find insignificant negative coefficients on the added regressors. Their sample period of runs from December 1971 to March 2016 (monthly frequency), and they use country-bucket portfolios as base assets, for which total returns, duration, average time-to-maturity, and yield-to-maturity are available from the data source.

Similarly to AMP13, Brooks, Palhares, and Richardson (2018) define momentum as each country's own 12-month average excess returns, and use the same dataset of Brooks and Moskowitz (2017) featuring country-maturity portfolios as base assets. Brooks, Palhares, and Richardson (2018) also study the relationship of style bond portfolios (momentum, carry, value, and low-risk) with macroeconomic factors, and conclude that they have less sensitivity to macro shocks than common sovereign bond indeces. They corroborate the evidence of AMP13 on the negative relationship between value and momentum in government bonds, and also find substantial diversification benefits when carry and quality are included in their multifactor portfolios. They report similar figures to those of Brightman and Shepherd (2016), with in-sample correlation between momentum and value of around -0.2.

In Section 3.4.1 I present the results of the replication procedure for the momentum factor defined in AMP13. I use a 12 month window for the signal calculation as in the original definition, 12-1M, as well as shorter 6- and 3-month windows to study shorter-term reversal effects in the daily bond returns. I denote these strategies as 6-1M and 3-1M respectively.

Value

Value is a widely known equity market factor and in short is defined as the tendency for relatively cheap assets to outperform relatively expensive assets, Brooks, Palhares, and Richardson (2018). AMP13 extend its definition from the equity market to government bonds using the 5-year change in the yields of 10-year bonds in each country, a measure similar to the negative of the past 5-year return which is shown to have a high correlation with the book-to-market (value) factor in equity markets, see Fama and French (1996b). As reported earlier, this choice is motivated by De Bondt and Thaler, 1985 who study the phenomenon of long-term reversal in individual stocks and use past return measures to document these patterns. AMP13 find that among US stocks the correlation between the returns formed from the negative of the past 5-year return and the returns from book-to-market sorts is 0.83. This relationship is also documented for the UK, Europe and Japan stocks, as well as for the global stock portfolio. AMP13 also analyse alternative measures such as real bond yield and term spread, the latter is often used as a sorting variable in carry factor definitions. Real yield is calculated as the 10-year bond yield minus the 5-year forecast in inflation[7], while the term

[7]Using investment bank analysts' estimate complied by Consensus Economics.

spread is the bond yield minus the short rate. AMP13 find that the alternative measures produce statistically significant Sharpe ratios of 0.73 and 0.55 respectively (real yield and term spread), and when analysed collectively (average of all three measures) the results are even stronger with a portfolio Sharpe ratio greater than 1.

Brightman and Shepherd (2016) use real bond yield as value measure, and build cross-sectional portfolios by taking an equally-weighted long position in the top third of the futures across countries, and similarly an equally-weighted short position in the bottom third of the contracts. They report a modest return of 0.5% with a 0.27 Sharpe ratio, the lowest figures among the three factors, and find a positive relationship with value strategies in other asset classes (currencies and equities).

Brooks and Moskowitz (2017) and Brooks, Palhares, and Richardson (2018) also employ the real bond yield as value measure using maturity-matched CPI inflation forecasts, and implement portfolios that are long assets with high deviations from fundamental levels, like expected inflation, and short those with low deviations. Brooks and Moskowitz (2017) find that excess returns of level portfolios across countries load positively (and significantly) on the value factor, even when carry and momentum are added as regressors. Value remains significant when principal components are included as well, at the expenses of PC1 that fails to be statistically relevant in the portfolio expected return equation. They study the information content of value and find that it absorbs the pricing information from the first principal component of the yield curve, but also provides additional explanatory power for expected returns due to the level of yields being related to some fundamental anchor (e.g. inflation). This procedure is in contrast with the 5Y reversal strategy of AMP13 that considers only the absolute level of past yields. Finally, Brooks and Moskowitz (2017) corroborate the evidence in AMP13 and Brightman and Shepherd (2016) on the positive relationship of the value factor in level portfolios with value strategies in other asset classes.

Supporting evidence of this finding is also reported in Kunz and Mazzoleni (2018) who use two different measures of value for their bond portfolio sorts, real bond yield as in AMP13 and the term spread. Contrarily to Brooks and Moskowitz (2017) they find that the value factor measured as the real yield delivers negative returns (-0.7% in their sample period 1989 to 2017), while the term spread is the preferred measure with a 3.4% return and a 0.47 Sharpe ratio. Their value measures are constructed using 10-year Treasury bonds and the short-term 3-month cash rate. Kunz and Mazzoleni (2018) further decompose the return of their term spread portfolio into carry and spot returns, with the former defined as in Koijen et al. (2018), and find that roughly half of the term spread's total return (1.4% out of 3.4%) is due to the carry factor and not to long-term reversal, and that the real yield portfolio takes consistent negative carry bets that erode potential returns.

I replicate the value measure of AMP13, 5Y reversal, for the bonds in each country and for the global bond portfolio. Similarly to the momentum factor, I also use an alternative 3-year window to calculate the change in the yields of all bonds within each country-bucket portfolio, 3Y Reversal. Contrarily to AMP13, I calculate these changes for all on-the-run bonds at each time t, and not only for the 10-year securities. The results are reported in Section 3.4.1.

Low-Risk

Frazzini and Pedersen (2014) employ a beta-based strategy (BAB) to construct their low-risk factor on the US Teasury curve. Their hypothesis is that high-risk assets offer lower returns compared to low-risk counterparts when adjusted for risk, and they go long assets with low beta and short the high-beta securities. Their risk measure is the beta from a OLS regression of asset returns on the market factor, which is constructed as the equal-weighted portfolio of all the bonds in the benchmark index over a 36-month period. They find that their BAB portfolio delivers a significant Sharpe ratio of 0.81 in the period 1952 to 2012, which remains almost unchanged when time-varying exposures are included in the regression analysis. Frazzini and Pedersen (2014) argue that BAB bond portfolios are subject to investors funding constraints, as the portfolios maintain market neutrality by targeting a specific risk target. As reported earlier, in this study I do not target any ex-ante duration, maturity, or volatility value, and I construct my cross-sectional factors considering bonds across maturity buckets for each issuer. However, contrarily to momentum and value which are normally constructed as cross-sectional country factors, low-risk is a cross-sectional maturity (or duration) factor and as such I calculate one single factor for each country.

Brooks, Palhares, and Richardson (2018) employ effective duration as their measure and construct the low-risk factor as a global bond portfolio that bets on shorter duration assets against longer duration ones, across all countries. I follow an analogous approach for my global low-risk factors, however I use both duration and term-to-maturity as signal measures. A crucial difference between Brooks, Palhares, and Richardson, 2018 approach and ours is that I use maturity rather than duration-adjusted weights in the portfolio sorts, a feature which does not take into account the different coupons of the bonds that are trading at each time t in each country (local) or across countries (global portfolio)[8]. I leave this topic for future research, see Section 3.5, but analyse the differences in performance of low-risk for the two measures. I take as given the series of modified duration from the data source which cover 55% of the total bond universe, see Section 3.3 for further details on the data. The long-short defensive portfolio of Brooks, Palhares, and Richardson (2018) produces a significant alpha when its excess returns are regressed on traditional market risk premia, and exhibits a low positive

[8]Long-dated bonds that were issued at the beginning of my sample tend to have consistently higher coupon rates with respect to those that were issued more recently. Consider for instance a situation in which a 30-year bond issued in 2012 with a relatively high coupon has the same time-to-maturity of a shorter dated bond that may be issued recently with a lower coupon. The former might have lower duration that the latter due to the role of coupons in discounting future cash flows in the duration calculation.

correlation with value and carry (0.21 and 0.15), whilst it is negatively correlated with momentum (-0.22). Thus, the low-risk portfolio can provide diversification benefits when combined with style (country-selection) factors such as value and momentum in a multifactor framework.

I replicate the low-risk factor of Brooks, Palhares, and Richardson (2018) using duration, Dur, and I compare the results against a benchmark measure that uses the bond's time-to-maturity, strategy TTM. The results are reported in Section 3.4.1.

3.3 Data and Summary Statistics

In this section I describe the data used in my analysis, Section 3.3.1, and discuss the summary statistics and stylised facts of my bond universe, Section 3.3.2.

3.3.1 Data Description

The sample period of my analysis runs from the beginning of January 2010 to the end of October 2021, and most of my data is available at the end of each trading day in the respective national calendars. I consider nine developed countries: Australia, Canada, Japan, the United Kingdom, the United States, and for the Euro Area I include France, Germany, Italy, and Spain. The main source is Bloomberg, from which I download CUSIP-level data for the bonds issued in the sample period by the nine national governments. I include bonds with maturity greater than 2 and up to 30 years, option-free, bullet and zero-coupon bonds. I exclude green bonds, international bonds, and in general all bonds with non-standard characteristics (e.g. retail, exchange-traded, when-issued bonds etc.). The total number of securities with complete series of end-of-day clean prices is $N = 972$, this figure drops to 788 if I consider the series of dirty prices[9]. Further details on my dataset can be found in AppendixB.3.

[Figure 3.1 about here.]

In figure 3.1 I show the time-varying composition of the bond universe by issuer and maturity. At the end of each month I count the number of outstanding securities with available price data, which are less than 100 at the beginning of the sample (2010-2011) and, as new bonds are issued, grow to about 700 by the end of the sample. On average, the number of outstanding securities each month is 362 across countries. Issuer-wise, United States Treasuries make up about 40% of the total number of bonds in the sample period, followed by Euro Area bonds with 28%, Japanese bonds with 19%, and Canadian bonds with 6%. Maturity-wise, the share of bonds maturing in 10 years from their issue date is 23% of the universe, 16% for the 5-year, 14% for the 7-year, and 13% for the 30-year bonds.

3.3.2 Summary Statistics

Using historical price series and CUSIP-level information, I report summary statistics on different aspects of the bonds such as issue price, issue volume, and daily returns. Table 3.1 reports country-specific summary statistics, and figure 3.2 shows the time-series of generic yields for each issuer considering four key points in the yield curve (2-, 5-, 10-, and 30-year maturities), as well as the relevant policy rates.

[Table 3.1 about here.]

[Figure 3.2 about here.]

[9] In total the number of bonds with complete series of end-of-day yields-to-maturity (based on the clean prices) is 717, 706 for yields on the dirty prices, 540 bonds with historical dollar duration series, and 288 bonds with complete Z-spread daily values.

Issue Price

One of the key results in the fixed income literature is that the price of a bond approaches its face value as the bond matures, which calls for a distinction between premium and discount bonds. I make this distinction at the beginning of the bond's lifetime (on the issue date), although the direction and speed of convergence of the day-t price to the bond's face value vary with the bond's age, see Brusa, Gu, and Liu (2014) for further details. I use the coupon-yield ratio $= (c_i/y_{i,t}) - 1$ from their paper to determine if bonds are trading at premium (ratio > 0), discount (< 0) or par ($= 0$). I count 624 discount bonds (65% of the universe), 295 premium bonds (30%), and only 14 par bonds (1%)[10] . All countries with the exception of Japan and Germany issued most of their bonds at premium in the sample period, which is expected given the persistent negative rates environment in the two countries from 2016 up to the end of the sample for all maturities below 10 years in Japan, and from 2015 to the end of the sample for the maturities below 5 years in Germany (and from 2019 for the 10 year point as well). See figure 3.2 for the details. In table 3.1 I also report the N-average coupon (annualised, in %) of each sovereign issuer. Japan and Germany have the lowest figures across all countries with about 85bp and 65bp per annum, while the coupon earned on the Australian debt is the highest on average with 2.81%, followed by Italy with 2.13%, and Spain with 2.09%.

Issue Volume

I report summary statistics on the volume of debt sold by the issuers in the sample period. In particular, column *Sold* in table 3.1 shows the N-average amount of debt sold (in billions of local currency) to market participants through syndications, auctions, and tap issues. The sale process of newly created debt securities can take place via syndication or actions. In their annual sovereign borrowing report[11], the OECD finds that the principal issuing procedure in use among countries is auctions, with 88% of the sovereigns (28 countries) issuing short-term debt, and 84% (27 countries) longer-term debt. Syndicated bond offerings are also found to be common in 23 OECD countries, they are mostly used for the first-time issuance of new instruments (in countries such as Australia, France, and Germany), and for longer-dated bonds (e.g. Australia, Italy and France). Although there are country-specific characteristics that influence the choice on the issuance process of new debt, syndications are generally used for ad-hoc procedures[12] , whilst auctions are the most common way for governments to raise debt on a consistent basis through regular auction calendars.

[10]Missing information on the issue price for 39 out of 972 bonds, which are unclassified.

[11]Source: OECD Sovereign Borrowing Outlook 2016

[12]Syndications are likely to lead to higher placing certainty in difficult market conditions. If the investors' demand for bonds is not substantial to fill the available supply, only part of the bonds on offer attract bids, which results in a failed auction. Although this phenomenon has been rare in the recent years, by the nature of the syndication model the issuers benefit from the intermediation of the dealer banks, that place orders on behalf of their clients (investors) to buy the newly issued debt. In the 2021 ECB Advisory Report on Debt Issuance, large European public issuers are found to pay between 7bp and 25bp of the total issuance amount to the syndicate banks, according to a standard fees schedule (the longer the maturity of the bond, the higher the fee).

For each security I also gather data on the cumulative amount of debt issued from its pricing date up to the end of the sample, this amount includes taps and re-openings. Column *Issued* reports N-average of these figures. Due to lack of bond-specific information on the dates of the re-openings, I use the difference between the amount of debt sold and issued (cumulative for that line up as of October 2021) to gauge weather the bonds were part of a new issuance[13], or reopening. In particular, I classify bonds as being part of new issuances if the difference *Sold - Issued* is within $\pm 10\%$ of the cumulative amount of debt issued for that line (*Issued*), including re-openings. On the other hand, all bonds whose amount sold is a fraction of the amount of debt issued as of end-of October 2021 are classified as tap issues (ie. bonds issued at the prevailing market prices using the terms of a past issue). Among the 972 securities considered in the sample period, 209 bonds (22%) are part of new debt offerings, the vast majority being 2-year Japanese Government Securities (15), and 2- (18), 3- (39), 5- (46), and 7-year (77) US Treasuries. The remaining 763 bonds belong to tap issues.

[Figure 3.3 about here.]

In figure 3.3 I report country-specific statistics on the yearly issue volume from 2010 to 2021. At the end of each year I calculate the amount of debt sold using bond-level data for each issuer, this includes bonds that were newly issued and those part of a re-opening (tap). The figures are calculated considering the cross-section of 972 bonds for which I have complete price series, and thus are only a fraction of the total amount of debt issued in the sample period by the national governments. For instance, when I compare my results with official data sources for the United States[14] I find that the trend in the gross amount of debt issued differs from what I report. The increase in the amount of debt issued in 2018 is 20%, versus the estimated 60% in panel 3.3a, which is lower than the 33% in 2020 officially reported, and comparable to the 2012 figure of 29%. My estimates are 14% and -32% respectively for 2020 and 2012. Overall, my results confirm that the pace of issuance in 2020 increase considerably across countries in response to the COVID-19 pandemic, this phenomenon is apparent when I look at the aggregate figures for the countries in the Euro Area that almost double their issue volumes from 2019 to 2020 (with the exception of France). I also find that with the exception of 2011 and 2012, the amount of debt sold through market participants by the US Department of the Treasury is always increasing on a yearly basis. The same holds for Japan from 2016 up to the end of the sample.

The summary statistics help me better understand the macro events that drove an increase in the supply of newly issued bonds, which are included in my universe. For instance, I find that 25% of Spanish bonds were issued in the period 2012-2014 following an injection of cash

[13]New issuances can be either syndications or auctions. With my data I are not able to uniformly differentiate the two across countries, mostly due to the different definitions and uses of each issuer. For a detailed survey refer to OECD Sovereign Borrowing Outlook 2016.

[14]The Securities Industry and Financial Markets Association (SIFMA) provides detailed statistics on the US Treasury market issuance, and I look at the YoY change in total gross issues of US Treasury Securities considering notes (2-, 3-, 5-, 7-, 10-year), and bonds (20, and 30-year).

from the ECB for bank recapitalisation, and similarly in 2020-2021 to tackle the COVID-19 pandemic (20% of the bonds). In July 2012 the Eurogroup agreed a financial assistance programme of up to 100 €B (for the following 18 months) aimed at stabilising the Spanish banking system, which came under pressure due to a combination of low growth, high unemployment, and inflated property prices. From the official sources[15], in 2012 the Spanish government issued a total of 97€B, and 74€B in 2013, which are the highest figures in the sample (2010-2021), together with the 110€B in 2020. This pattern is consistent with my data, panel 3.3a shows the highest figures in the period 2013-2014 and in 2020-2021.

Daily Returns

In table 3.1 I report summary statistics on the bonds' simple daily returns, calculated using the time series of clean prices. Column *Pearson* shows the N-average pair-wise correlation of the bonds in each country (across all maturities, considering the time-series of returns), which gives me a snapshot of the intra-group dependence. For instance, I find that Japanese bonds tend to behave less uniformly, and across all groups are the ones that show the lowest correlation estimates, 0.69. Contrarily, the bonds in countries such as the United Kingdom and Italy tend to have similar price trajectories from a time-series perspective, 0.84 and 0.87 respectively. I also report the minimum daily return recorded, column Min, the maximum daily return in the sample period, Max, and N-averages (in %, annualised[16]) of the bonds' standard deviations and premia, calculated using the available time-series (from the issue date to maturity, or up to the end of the sample for the bonds that are on-the-run as of October 2021). The bonds that carry the highest premium on average are found in Italy, 1.15% estimated on a sample of 98 bonds with an average maturity of 7 years, and Spain, 0.78% considering 59 securities with a similar maturity profile. Australia (6.19%, 32 securities with average maturity of 12 years), the United Kingdom (5.82%, 35 bonds with maturity of 12), and Italy (5.66%) dominate the statistics on the average risk profile of the bonds, whilst Canada (1.99%, 62 bonds with 4.6 maturity), Germany (2.31%, 72 bonds with 6 years maturity), and Japan, (3.14%, 184 bonds with the highest average maturity of 15.7 years) have the lowest figures.

To better understand the statistics on market-driven changes in bond prices, for each issuer and maturity bucket I calculate the N-average return and standard deviation (of the returns) using the series of clean prices. I consider the lifespan of each bond from issuance up to maturity for off-the-run bonds as of October 2021, and up to the end of the sample for on-the-run bonds. These figures are equivalent to the average returns earned by investing in equal portion in the bonds of each issuer during the sample period (ex-post), by maturity bucket. I report the results in figure 3.4.

[15] Source: 2021 Funding Strategy, amount net of redemption.

[16] Let $\hat{r}$ be the in-sample daily average return of an asset (or portfolio) and $\hat{std}(r)$ its standard deviation, I annualise the measures as $Avg = (1+\hat{r})^T - 1$ and $Std = \hat{std}(r)\,\sqrt{T}$ where T is the number of daily time-series observations. This applies to all summary tables in this chapter unless noted otherwise.

[Figure 3.4 about here.]

I find that for short-dated securities (buckets 1 and 2), Italian and Spanish bonds yield on average the highest return across countries, with the highest volatility figures. On the other hand, when I consider longer-dated securities US Treasury bonds dominate the risk-return statistics. For most countries, the return on bonds with maturity of less than 11 years (buckets 1 and 2) is on average negative, while it is positive for the bonds in bucket 3. The only exceptions are Germany and the United Kingdom, whose bonds yield negative or near-zero returns across all maturity buckets. My calculations give a snapshot of the term-premium for each issuer, the amount by which the yield on a long-term bond is greater than the yield on a shorter-term bond. From a maturity-selection perspective, for the same issuer the returns earned by an investor who bears higher risk in the form of higher sensitivity to interest-rate changes (duration) should be higher than the returns earned by investing into low-duration bonds. My data confirms that for each issuer the average return and standard deviation are increasing in the buckets (average maturity of the portfolios)[17]. For instance, the return differential for US Treasuries in the first and last buckets is about 3% annualised, and the standard deviation increases from 1.4% in bucket 1 to 14.68% in bucket 3. In the scatter plot of panel 3.4a I also report the risk-return profile of Japan versus US -issued bonds by maturity bucket. JP1 is the point that corresponds to the $1/N$ long-only portfolio of Japanese bonds with 1 to 5 year-maturity, JP2 considers the bonds in bucket 2, and JP3 those in bucket 3. Similarly for US Treasury bonds, US1, US2, and US3. A cross-issuer and -maturity analysis of US and Japanese bonds reveals that the risk profile of JP2 is similar to US1, with both yielding near-zero average returns. Japanese bonds have the lowest in-sample (average) standard deviation of returns across issuers and within each bucket.

[17] With the exception of the buckets for which I have a low number of bonds.

3.4 Results

In this section I report the results on the performance of momentum, value, and low-risk individually, Section 3.4.1, and examine the factors collectively by country and maturity bucket, Section 3.4.2.

3.4.1 Factor Replication

Following the methodology outlined in Section 3.4.1, I construct my cross-sectional bond factors using the standard definitions from the literature, and compare their performance against alternative measures. I now analyse the results for each factor individually.

Momentum

Table 3.2 reports summary statistics on the estimated momentum factors, considering the issuer-specific bonds for the country portfolios, and all bonds for the global portfolio. I use the canonical 12-1M definition of AMP13, panels 3.2a-3.2b, and alternative measures with shorter estimation windows of 6 and 3-month past returns, panels 3.2c-3.2d for the 6-1M measure and 3.2e-3.2f for 3-1M. For each measure I report the results of the long-only and long-short portfolios. I require a minimum of five bonds to be traded at each time t per maturity bucket, and a minimum of three years of data. Frequency is daily. Column *Start* indicates the start date of the time-series of factor returns, *Avg* is the estimated average return per annum in %, *Std* the annualised standard deviation, *SR* is the Sharpe ratio[18], *t-Stat* is its t-statistic[19], and ΔT is the time-series average of the portfolio term-to-maturity in the relevant sample, expressed in years. I indicate with an asterisk if at any time in the sample period the portfolio net term-to-maturity turns negative (only for long-short portfolios). Finally T refers to the number of time-series observations (in days), and N to the number of bonds considered in the whole sample for the country-bucket portfolios. All factors are estimated up to October 29th 2021, unless noted otherwise.

[Table 3.2 about here.]

I begin by analysing my results considering the canonical 12-1M long-short measure as baseline, and compare the performance of the factors across countries (in local currency terms), and across maturities. For all countries, momentum strategies deliver high statistically significant Sharpe ratios when short-term bonds are included, up to 5-year maturity. When bucket 3 bonds are considered, 11 to 30-year maturity, the effectiveness of these strategies decreases considerably and they fail to produce statistical significant Sharpe ratios. The global portfolio delivers low and statistically insignificant Sharpe ratios since 2011 for all maturity buckets, while Euro area countries dominate the rankings, with France's and Spain's momentum

[18] Assuming a zero risk-free rate.

[19] I measure the t-statistic as $t\text{-}Stat = \sqrt{T}\ SR$ following Lo (2002) where T is the length of the portfolio return series in years, and SR its annualised Sharpe ratio.

portfolios yielding about 4.0 annualised risk-adjusted performance from 2012/2013 to 2021. On the other end of the spectrum, Canadian and Japanese bond portfolios have the lowest figures with about 1.3 and 1.6.

Compared to long-only portfolios, average returns on the long-short counterparts are consistently higher, which does not come at the cost of increased volatility. However, for virtually all long-short momentum strategies, the portfolio net duration always turns negative in the sample period, which carries additional costs related to open and maintain short positions in government bonds. One of the main limitations of my study is that I do not incorporate such costs in my performance calculations, since I do not rebalance my portfolios to match a desired duration target. This implies that the performance measures of my long-short portfolios are less conservative and require further investigation, see Section 3.5.1 for the details. Considering the 12-1M measure and bucket 1 bonds, I find that long-only portfolio sorts in countries such as France, Germany, and Japan deliver negative statistically significant Sharpe ratios from 2013, 2012, and 2014 respectively. For all other countries, the Sharpe ratios are near zero.

Overall, I find that the performance of the long-short momentum portfolios is robust to the choice of the estimation window, with the longer 12-month window providing slightly better results. Using the alternative measures, panels 3.2c-3.2d for 6-1M and panels 3.2c-3.2d for 3-1M, I corroborate the evidence that long-short portfolios deliver consistently superior performances than the long-only counterparts, across countries and maturity buckets, and achieve higher returns when shorter-term bonds are included.

The figures for the global momentum factor in panel 3.2b are in line with AMP13 who report an average return of 1.0% and standard deviation of 5.8% between 1982 to 2011. My bucket 1 portfolio statistics are 1.7% and 5.8% respectively, between 2011 and 2021. The results remain consistent when I employ the 6-1M and 3-1M measures, the average return of the 3-1M and 6-1M bucket 1 portfolios are 1.3% and 0.98%, and the respective standard deviations 5.09% and 5.04%. The average term to maturity of these portfolios range from 6.4 (3-1M) to 7.8 (12-1M) years.

Value

Similarly to before, table 3.3 reports summary statistics on the estimated value factors, considering the issuer-specific bonds for the country portfolios, and all bonds for the global portfolio. Panels 3.3a-3.3b show the results for the canonical 5Y Reversal of AMP13, and panels 3.3c-3.3d for the alternative 3Y Reversal with a shorter look-back window. As usual, I require a minimum of five bonds to be traded at each time t per maturity bucket, and a minimum of three years of data. With respect to the momentum factor, the number of countries for which I have complete series of value factor returns is lower, due to the longer estimation window needed to produce the signals.

[Table 3.3 about here.]

I begin by analysing my results for the 5Y Reversal long-short portfolios. From July 2018 up to October 2021, the US value factor returned an average 1.5% per annum with a Sharpe ratio of 0.5. Its average net duration is slightly negative at -0.7 years, indicating that it was positioned to benefit from the rising interest rates environment up to early 2019, but suffered from the consecutive interest rate cuts in 2020 and 2021 in response to the COVID-19 pandemic. Comparing across buckets and measures, bucket 1 with the 3Y Reversal and bucket 3 with the 5Y Reversal, I find that the risk-adjusted return on the value factor tends to be higher when longer-dated bonds are included. This is true in particular for Japanese bucket 3 bonds from 2016, in which the value strategy delivers the highest Sharpe ratio across groups, about 0.8. When I analyse the results of the global portfolio for bucket 3 bonds, I find that its return is negative and the average term-to-maturity is -12.8 years. Similarly to momentum, I find less support for global rather than local value factors. Compared to long-only portfolios, average returns on the long-short counterparts are consistently higher and less volatile. This is true in particular for Japan in which the returns of long-only sorts are negative across all buckets, and then become positive when short selling is allowed.

When I use a shorter look-back window of three years, the Sharpe ratios of the long-short portfolio of France and Germany (bucket 1) are the only ones that become statistically significant across countries, contrarily to the 5Y Reversal portfolios that all show evidence of statistically insignificant Sharpe ratios. The return on the 3Y Reversal German portfolio from September 2016 is 1.9% annualised, with a 1.3 Sharpe ratio, and the return on France's portfolio is 2.3% from June 2018, with a 1.2 Sharpe ratio. Similarly to 5Y Reversal, the return of the global portfolio is negative and its Sharpe ratio insignificant for all maturity buckets. Perhaps interestingly, I find supporting evidence for a reversal in Germany's bond returns when I employ long-only 3Y Reversal strategies: -1.53% is the average return of the long-only portfolio with a maturity of 6 years, versus the long-short one with 1.9% and maturity of 3 years.

Compared to AMP13, who report an average return of 0.5%, standard deviation of 6.4% and Sharpe of 0.07 from 1982 to 2011, the average returns of my global value factors in panel 3.3b are considerably higher. When I employ the 3Y Reversal measure my figures are more in line with AMP13. The average return on bucket 1 global portfolio is 0.6% from November 2013, standard deviation of 6.93% and Sharpe of 0.07, and similarly the bucket 2 portfolio with 0.86%, 5.55% and 0.15 respectively The average term to maturity of these portfolios range from 2.5 (3Y Reversal long-short global bucket 1) to 5.5 (5Y Reversal long-short global bucket 2) years.

Low-risk

Table 3.4 reports the results on the construction of the low-risk factor, I use duration, panels 3.3a-3.3b, and the term-to-maturity, panels 3.3c-3.3d.

[Table 3.4 about here.]

FP14 report a Sharpe ratio of 0.81 and a 2.43% volatility for their low-risk factor on US Treasury bonds from 1952 to 2012, which is line with my estimates of 0.35 and 2.49% respectively using the bonds' term-to-maturity from February 2014, see panel 3.4d. My US low-risk portfolio has an average ΔT of 1.3 years which is always positive in the sample period. This is an exception across countries and measures, with all the other low-risk factors constructed as long-short portfolios, panels 3.3b and 3.4d, having net negative duration at some point from 2010 to 2021. When I consider ΔT as baseline, I find that the highest Sharpe ratios across countries are the ones on France and Germany's bonds with about 1.2 to 1.3 for the long-short portfolios, and 0.78 and 0.45 for the long-only counterparts. When I use duration as measure, which takes into account the role of coupons in discounting future cash flows, my results remain consistent and the highest Sharpe ratios are the ones of the Euro Area countries.

Overall, the differences in performance of the low-risk portfolios using the bonds' term-to-maturity or duration are numerically relevant, contrarily to what FP14 reports. For the US, the TTM long-short portfolio has a Sharpe ratio of 0.35 from 2010, versus 0.13 for the duration portfolio from 2014. This difference is less pronounced for the Euro Area countries and the UK, and reverses for Japan, in which the duration portfolio's Sharpe ratio is 0.58 from 2014, versus 0.19 for the TTM portfolio from 2010. Comparing long-short versus long-only weights, I corroborate the findings of momentum and value factors also for low-risk. Ceteris paribus, the performance of long-only portfolios is inferior to the long-short counterparts, although I do not account for rebalancing and duration matching in my performance calculation.

3.4.2 Multi-Factor Results

In this section I analyse the factors collectively for each country and maturity bucket. To facilitate the comparison, I focus only on bucket 1 (2- to 5-year) and bucket 3 (11- to 30-year) bonds and exclude countries with only one available factor. I also fix the measures for factor construction, 12-1M for momentum, 5Y and 3Y Reversal for value (bucket 3 and 1 respectively), and duration for low-risk. As highlighted in the previous section, the differences in the characteristics of factors constructed under long-only or long-short constraints are relevant, and I present my results separately. This choice is also justified by the constraints that each investor faces, e.g. duration-matching in long-only, and -hedging in long-short portfolios, which results in portfolios with very different risk-return (and duration) profiles. Section 3.4.2 reports the results considering all available data for each factor, country, and bucket combination, and Section 3.4.2 reports the results on a restricted time span, from October 2018 to October 2021.

Unrestricted Sample

Table 3.5 presents the multi-factor results considering all available data for each factor, country, and bucket combination.

[Table 3.5 about here.]

For a long-only investor, panel 3.5a, the benefits of factor investing in the sovereign bond market appear to be slim, with most factors returning negative or low Sharpe ratios across countries and buckets. For bucket 1 US bonds, the momentum factor has a Sharpe ratio of -0.22 and the value factor of 0.18, with the the duration of the latter being considerably higher than the one of momentum. Value has a duration profile similar to that of the low-risk, with about 16 years. For Germany, momentum and value factors show evidence of statistically significant Sharpe ratios, both of which are negative. Similarly, for France and Japan bucket 1 bonds momentum yields near -1 Sharpe ratios. The highest Sharpe ratios are the ones of momentum on bucket 3 Japanese bonds, 0.54 from 2011, and value on US bucket 3 bonds, 0.51 from 2018. The low-risk factors show evidence of statistically insignificant Sharpe ratios across all countries.

For a long-short investor, panel 3.5b, I find that all three factors have positive and statistically significant Sharpe ratios in France and Germany (bucket 1), and similarly in Italy and Spain for momentum and low-risk. For bucket 3 bonds, none of the factors have statistically significant Sharpe ratios across countries.

Restricted Sample

To further assess the relative performance of the factors across countries, in the this section I calculate the summary statistics on a restricted time span, from October 2018 to October 2021. Table 3.6 reports the results.

[Table 3.6 about here.]

For a long-only investor I corroborate the evidence on the poor performance of the momentum factor across countries and buckets, this is true in particular Japan and Germany (bucket 1) that have negative statistically significant Sharpe ratios of -1.33[20]. The only other factor with statistically significant risk-adjusted performance is low-risk for Italy, with 1.19. The highest figures across countries are the ones for bucket 3 US bonds, in which the momentum factor yields a 0.63 Sharpe, value 0.77 and low-risk 0.48. When I consider bonds from all countries for my portfolio construction procedure, I find that none of the factors have statistical significant Sharpe ratios. However, with respect to most of the country-specific portfolios, the global one tends to deliver higher risk-adjusted returns for momentum and value (bucket 1).

[20]For Japan only, the value factor is available from October 2019 to April 2021.

For a long-short bond investor, I find evidence of positive and statistical significant returns on the momentum and low-risk factors only for certain country-factor combinations. For momentum, this is true only for bucket 1 bonds, with factors constructed using longer-dated securities delivering insignificant albeit positive Sharpe ratios. For euro area countries, the momentum factors performs well in the sample period with Sharpe ratios ranging from 3.62 for Germany to 11.46 in Spain, and similarly for low-risk with 0.94 and 1.89 respectively. Value appears to have a statistically significant positive risk premia in Germany alone, with a Sharpe ratio of 1.48.

To better understand the dynamics of the three factors in each country, in figure 3.5 I report time-series of cumulative returns (left panels), together with the relevant summary statistics (right panels), from October 2018 to October 2021. I report the results only for the countries in which all three factors are available, and focus on bucket 1 long-short portfolios for value and momentum.

[Figure 3.5 about here.]

I start by analysis my results for the US, the issuer with the highest number of securities in my universe. Although not negative as reported in AMP13, I find a low correlation of 0.04 between my momentum (12-1M) and value (3Y Reversal) factors. This may be due to the fact that the value portfolio's net term-to-maturity is negative, -8 years, which indicates that is positioned to deteriorate in a declining-yield scenario (usually associated with economic contraction). Although the anti-cyclical behavior of the value factor can be important for multi-factor investing, my findings suggest that this may be due to the way in which long-short signal-based portfolios are constructed. As in AMP13, I do not rebalance the portfolios to match a desired duration target, which exposes investors to duration risk, and in general does not take account of the costs associated to maintain open short positions. This implies that the returns on the value factor can be even lower than what reported, which casts doubt on its potential in the US sovereign bond market (and internationally too). On the other hand, I find that low-risk is negatively correlated to momentum, -0.23, and uncorrelated to the value factor. Its Sharpe ratio is 0.11 and its maturity -1.4 years, which can partly offset the net positive maturity of 0.9 years on momentum. For US Treasury securities, the low-risk factor appears to be a better candidate for inclusion in a multi-factor portfolio than value. My figures are in line with Brooks, Palhares, and Richardson (2018) who find a negative correlation of about -0.22 between momentum and low-risk, although their results are presented for a global portfolio of 13 countries (from 1996 through 2017).

When I consider factors that allocate capital internationally, panel 3.5c, I find that the information content of momentum and low-risk is similar, with their in-sample correlation being 0.91. This suggests that momentum may be considered a duration proxy rather than an anomaly, since low-risk is constructed by investing in bonds with lowest dollar duration in the cross-section at any time t. Perhaps interestingly, I find that momentum and value have a negative near-zero correlation in the sample period, -0.09, only for the global factor. AMP13 also use global portfolios for their factors and estimate a -0.18 correlation from 1989 to 2016 at monthly frequency.

For Japan, I find evidence evidence of near-zero or negative cross-factor relationships in the sample period. The correlation between momentum and value is 0.05, -0.13 between momentum and low-risk, and -0.05 between value and low-risk. In line with my expectations, I find that all factors enjoyed a positive return in the sample period, although only momentum produces statistically significant Sharpe ratios. From October 2018 to the beginning of 2021 the value factor returned approximately 3.5%, however its net term-to-maturity is the highest across factors with about 6.4 years, which results in higher volatility. This can be seen from the time-series plot in panel 3.5a, that shows how during the period from August to end-of-September 2019 the factor almost doubled in value (from 2.5% to about 5% cumulative return) before erasing all the gains towards the end of the year. At its 30 July meeting, board members of the Bank of Japan decided to keep its monetary policy unchanged, with the shot-term policy rate charged to financial institutions' deposits at the Bank at -0.1%. During the same period the FED decided to lower the benchmark rate by 25bp, citing 'implications of global developments for the economic outlook as well as muted inflation pressures'[21]. In this study I do not establish contagion effects across country factors, a topic that I leave for future research, however this particular period offers a precedent for the analysis due to the fact that also the ECB decided to keep its rates unchanged[22]. During the same period (August to end-of-September 2019), I observe substantial volatility for most of the factors across countries, see for instance low-risk in Germany or Italy that profited the most from the events by the end of the summer, from 0.5% to 2.5% and from 4% to 8% cumulative return respectively. Overall, given the low (or negative) correlation estimates and positive risk-adjusted returns, I find more support for the three factors in Japan rather than in the United States.

Moving on to Euro-area countries, panels 3.5d - 3.5f, I notice that the cross-factor correlations remain somewhat stable across countries. In fact, the value factor is uncorrelated with both low-risk and momentum, while the latter has a numerically relevant correlation with low-risk, ranging from 0.37 in France to 0.76 in Italy. For Italy, this is also reflected in the estimated Sharpe ratios on the two factors, which are statistically significant and numerically high. From the time-series plots I can see how cumulative returns on momentum follow an upward steady path in the sample period, in line with the decline in interest rates (the yield on Italian 10-year bonds went from 3% at the beginning of 2019 to 1% on October 2019, see figure 3.2). For peripheral countries such as Italy and Spain, the momentum factor fares particularly well with the highest risk-adjusted returns across all countries, followed by the returns on momentum

[21]Source: FED Press Release July 2019.
[22]Source: ECB Press Release July 2019.

factors of semi-core countries such as France, and Germany (core). This result is in line with the relative level of yields across countries, and thus their perceived risk by market participants (default risk). The dynamics on momentum are somewhat different to low-risk, with the latter showing higher volatility in the sample period due to its duration being always greater than the one of momentum. For the value factor I can see that it follows an anti-cyclical behavior for France and Italy, due to its negative net duration in the sample period. This is in contrast with the results for Germany that show how, despite being uncorrelated with momentum, value has a net positive duration and its Sharpe ratio is statistically significant.

3.5 Conclusion and Further Research

In this chapter I studied factor premiums in global government bond markets using CUSIP-level data at daily frequency from the beginning of January 2010 to the end of October 2021. I constructed cross-sectional style factors (momentum, value, and low-risk) as country-maturity portfolios that allocate capital proportional to relative rank of each asset's signal in the cross-section. I used the definitions of AMP13 for momentum and value, and FP14 for low-risk, and for each measure I constructed long-only and long-short portfolios to examine the relative performance of the factors across various dimensions, countries, maturity buckets, and investment constraints.

My analysis reveals a number of findings that do not support the view of consistent factor premia across countries, or maturity buckets. Using 12-1M as momentum measure, I find that risk-adjusted returns are decreasing in the maturity of the bonds, with the highest Sharpe ratios found for portfolios formed on short-maturity bonds (lower than 5 years). For longer-dated securities, momentum does not deliver statistically significant portfolio returns. When analysed across countries, my results reveal that momentum produces consistent statistically significant Sharpe ratios, however this is not true for value and low-risk. Comparing across factors, I find that standard reversal measures of value as in AMP13, the 3- and 5-year change in bond yields, produce portfolios with statistically insignificant risk-adjusted performance. The only exception is Germany, for which the value factor has a positive and significant Sharpe ratio from October 2019 to October 2021. Low-risk yields statistically relevant results only in Euro Area countries.

I find an unambiguously better performance of long-short rather than long-only portfolios, with the latter often yielding negative statistically significant Sharpe ratios (in particular for momentum). Similarly, when I allow portfolios to allocate capital internationally, I find that for all factors this leads to a substantially lower performance with respect to the country-specific counterparts. This result is in contrast to what is commonly reported in the literature, and suggests that premia on style factors in the sovereign bond market vary substantially across countries.

Using data from October 2018 to October 2021, I analyse the cross-factor relationships in each country, which show how standard results in the literature for global style factors fail to be consistent locally. For the United States, I find that the returns on value and momentum are not linearly related, which may be due to the fact that value portfolio's net term-to-maturity is negative (contrarily to momentum). From a methodological perspective, I corroborate the evidence that the approach of AMP13 expose investors duration risk, which may help explain why they report a negative correlation between momentum and value. Using their definitions and a similar methodology, I show how long-short portfolios can have negative net maturity (or duration) in a given sample period, which indicates that the portfolio is positioned to benefit in periods of economic expansion (usually associated with rising yield), as opposed to a portfolio with positive net maturity that profits in a declining yield scenario (usually associated with declining yields). For the United Sates, value is also uncorrelated with low-risk, suggesting that it is an unlikely duration proxy, and the latter has negative net maturity and shows a negative correlation with momentum. This pattern is somehow stable across countries, with the exception of global and Euro Area countries portfolios in which the correlation between momentum and value is high and positive. For the global factor in particular, I find the highest correlation estimate, which suggests that momentum may be considered a duration proxy rather than an anomaly (or premia), due to the fact that low-risk is constructed as the portfolio that invests in bonds with lowest duration (or maturity) in the cross-section at any time t.

3.5.1 Further Research

In this chapter I make a series of assumptions and methodological choices that help study global factor premia using bond-level data at daily frequency. In this section I address the main limitations that arise from making such choices and propose further research.

From a methodological perspective, the main limitation of my study is to rely on predetermined measures and standard factor construction techniques that are adapted from the equity literature. As discussed in Section 3.2, I leave the problem of finding new measures or modifying existing ones to study the cross-section of sovereign bond returns for future research. Based on my findings this is particularly relevant for value, which fails to generate statistically significant risk-adjusted returns using past return measures, i.e. in the same spirit of momentum factors, which are constructed based on each asset's own history. In the same section, I saw how recent literature tends to depart from AMP13's definitions of value by considering deviations from a fundamental anchor, e,g. inflation expectations as in Brooks and Moskowitz (2017) and Brooks, Palhares, and Richardson (2018), or carry-based strategies, see Koijen et al. (2018). In Appendix C.2 I report the details on the implementation of the carry factor extraction procedure in Koijen et al. (2018) for my framework.
I also briefly comment on reconciling the results from studies using portfolios as base assets with my setup that features individual securities.

The key limitation of the methodology in my study is that I do not rebalance the portfolios to match a desired duration target. While this choice allows me to compare style factors across maturity buckets, it exposes investors to duration risk. The procedure above can be also be used to ensure that the signal-based weights are adjusted to match a desired duration target, a topic which I briefly comment in Appendix C.2 and leave for future studies .

Another extension of my study is be to test the theory of Brusa, Gu, and Liu (2014), who argues that time decay has an impact on bonds' price trajectories proportional to the coupon-yield ratio $= (c_i/y_{i,t}) - 1$. Based on their findings, one can test the performance of a strategy that goes long bonds trading at discount that are close to maturity and shorts those trading at premium. Since the speed of convergence to the bond's face value is proportional the coupon-yield ratio, one can use the latter as signal for portfolio sorts. Note that this requires taking into account the bond's accrued interest, and not only the clean price as I do in my study.

Moreover, in this chapter I focused solely on momentum, value, and low-risk, and exclude other sources of price variation, such as liquidity, which have been shown to be important determinants to explain contemporaneous return variation. Although I consider only bonds issued by developed countries, which are more liquid than those issued by emerging market economies, studying the supply of bonds issued by the national government is of crucial importance. In this regard, I propose to use data on the Federal Reserve System Open Market Account (SOMA), which are available from my data source (only for the United States). This is particularly relevant in light of the FED's declining balance sheet, which means that more bonds that are held at the Bank will likely be available in the secondary market, thus increasing the supply. The data is available at CUSIP-level, which allow me to assess also which portion of the curve is more affected, ie. short-, medium-, or long-dated securities.

Finally, in this study I not analyse the performance of my factors in explaining contemporaneous returns or forecast future ones, a topic which I leave for future research. Although I report high and statistically significant Sharpe ratios for some of the factors, I am not able to accurately assess their pricing performance and how it changes in relation to macroeconomic and financial events.

TABLE 3.1: **Summary Statistics**

The table reports country-specific summary statistics for the 972 bonds issued from January 15th 2010 to October 29th 2021, for which I have complete price series information. N refers to the total number of securities in each country during the sample period. Column # *Disc* reports the number of bonds issued at a discount price, similarly # *Prem* the premium bonds, and # *Par* the par bonds. *Cpn (%)* shows the N-average annualised coupon in %. I group the bonds into three maturity buckets: # *1-5* are the bonds issued with a maturity of one to five years, # *6-10* those with a maturity of six to ten years, and # *11-30* are longer-dated bonds with maturities ranging from 11 up to 30 years. The columns *Mat (yr)* reports the N-average maturity (in years) of the bonds issued in each country during the sample period. I also show average volume information on the funding techniques adopted by each issuer in the sample period. In particular, column *Sold (B)* reports the average amount of debt sold (in billions of local currency) to market participants through syndications, auctions, and re-opening of existing bond lines (taps). Column $Issued(B)$ reports issuer-specific averages of the cumulative amount of debt issued using bond-specific time-series information. In particular, for each bond in the universe I obtain the cumulative amount of debt sold (including taps) up to the end of the sample, and I take N-averages of these values. Finally, I report summary statistics on the bonds' simple daily returns, calculated using the time series of clean prices. *Pearson* is the average pair-wise correlation of the bonds in the relevant country, estimated in the full sample, and $Min(\%)$ is the minimum daily return recorded (in %), analogously $Max(\%)$ is the maximum daily return in the sample period. I also report N-averages of the bond-specific estimated standard deviations (in %, annualised), column $Std(\%)$, and N-averages of the bonds' premia calculated using the available time-series, $Avg(\%)$, from the issue date to maturity, or up to the end of the sample for the bonds that are still on-the-run as of October 2021.

		Issue Price				Maturity Date				Issue Size			Daily Returns				
Issuer	N	# Disc	# Prem	# Par	Cpn (%)	# 1-5	# 6-10	# 11-30	Mat (yr)	Cny	Sold (B)	Issued (B)	Pearson	Min (%)	Max (%)	Std (%)	Avg (%)
Australia	32	16	1	0	2.81	7	4	21	12	AUD	6.5	26.9	0.77	-6.1	7.05	6.19	0.21
Canada	62	61	1	0	1.3	49	4	9	4.6	CAD	3.8	18.7	0.77	-2.37	1.87	1.99	-1.32
Japan	184	66	118	0	0.85	38	53	93	15.7	JPY	1775.8	4294.6	0.69	-4.99	5.67	3.14	0.63
United Kingdom	35	17	13	0	1.32	4	20	11	12	GBP	4.1	31.2	0.82	-6.53	5.67	5.82	-0.52
United States	385	329	45	11	1.85	170	162	53	8.9	USD	33.4	48.2	0.7	-8.79	10.11	4.22	0.23
Euro Area																	
France	45	33	10	1	1.2	15	20	10	8.9	EUR	4.6	37.2	0.76	-3.62	2.45	3.32	0.52
Germany	72	22	48	1	0.65	40	30	2	6.6	EUR	5	20.3	0.72	-1.98	2.24	2.31	-0.34
Italy	98	52	45	1	2.13	46	36	16	7.7	EUR	4.9	18.2	0.82	-7.43	7.37	5.66	1.05
Spain	59	28	14	0	2.09	27	23	9	7.9	EUR	5.5	20.6	0.75	-4.1	7.53	4.14	0.78
Total	972	698	328	35	1.6	424	418	298	9								

$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

TABLE 3.2: **Momentum**

The table reports the results on the construction of the momentum factor, considering the issuer-specific bonds for the country portfolios, and all bonds for the global portfolio. I use the canonical 12-1M definition of Asness, Moskowitz, and Pedersen (2013), panels 3.2a-3.2b, and alternative measures with shorter estimation windows of 6 and 3-month past returns, panels 3.2c-3.2d for the 6-1M measure and 3.2e-3.2f for 3-1M. For each measure I report the results of the long-only and long-short portfolios. I require a minimum of five bonds to be traded at each time t per maturity bucket, and a minimum of three years of data. Frequency is daily. Column *Start* indicate the start date of the time-series of factor returns, *Avg (%)* is the estimated average return per annum in %, *Std (%)* the annualised standard deviation, *SR* is the Sharpe ratio, *t-Stat* is its t-statistic, and $\Delta T(yr)$ is the time-series average of the portfolio term-to-maturity in the relevant sample, expressed in years. I indicate with asterisk if at any time in the sample period the portfolio net term-to-maturity turns negative (only for long-short portfolios). Finally T refers to the number of time-series observations (in days), and N to the number of bonds considered in the whole sample for the country-bucket portfolios. All factors are estimated up to October 29th 2021.

(A) *12-1M Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	Bucket 2	24-Apr-2017	0.52	4.81	0.11	0.23	11.4	1140	13
	Bucket 3	01-Apr-2015	1.82	10.32	0.18	0.44	24.1	1525	18
Canada									
	Bucket 1	08-May-2012	0.27	1.38	0.2	0.53	3.8	1811	43
	Bucket 2	02-Aug-2016	-0.24	4.92	-0.05	-0.1	11.2	1028	10
Japan									
	Bucket 1	01-Aug-2014	-0.66	0.58	-1.14	-2.97	4.5	1759	56
	Bucket 2	03-Oct-2011	0.55	1.95	0.28	0.87	10.7	2528	66
	Bucket 3	01-Jul-2011	2.73	5.05	0.54	1.7	27.8	2582	87
United Kingdom									
	Bucket 1	16-Sep-2015	-0.93	1.58	-0.59	-1.23	4.5	1135	18
United States									
	Bucket 0	02-Oct-2015	-0.63	0.18	-3.5	-8.37	0.8	1488	152
	Bucket 1	02-Feb-2012	-0.39	1.76	-0.22	-0.68	4.5	2449	232
	Bucket 2	27-May-2011	1.05	4.5	0.23	0.73	9.2	2624	146
	Bucket 3	29-Feb-2012	3.19	13.91	0.23	0.7	37.8	2435	45
France									
	Bucket 1	08-Aug-2013	-0.82	0.95	-0.86	-2.42	4.1	2052	29
	Bucket 2	24-Oct-2013	1.73	3.68	0.47	1.3	10.7	2000	22
Germany									
	Bucket 1	08-May-2012	-1.01	1.03	-0.98	-2.96	4	2365	52
	Bucket 2	10-Sep-2012	1	3.76	0.27	0.8	10.6	2278	27
Italy									
	Bucket 1	01-May-2012	-0.16	3.51	-0.05	-0.15	3.9	2321	66
	Bucket 2	18-Mar-2013	3.25	6.96	0.47	1.35	10.3	2154	39
	Bucket 3	07-Feb-2018	4.71	12.93	0.36	0.64	21.8	828	10
Spain									
	Bucket 1	18-Sep-2012	-0.5	1.48	-0.34	-0.97	3.7	2099	38
	Bucket 2	04-Jun-2014	1.79	4.64	0.39	1.04	10.7	1848	24
Global									
	Bucket 0	22-Dec-2014	1.11	3.99	0.28	0.72	9.3	1718	213
	Bucket 1	27-May-2011	0.37	3.98	0.09	0.28	10.2	2591	390
	Bucket 2	26-May-2011	0.73	2.49	0.29	0.91	10.3	2580	221
	Bucket 3	01-Jul-2011	2.59	4.19	0.62	1.95	22.1	2572	199

(To be continued)

(Continued)

(B) *12-1M Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	Bucket 2	24-Apr-2017	1.21	2.02	0.6	1.26	2.6	1140	13
	Bucket 3	01-Apr-2015	1.35	3.08	0.44	1.07	5.3*	1525	18
Canada									
	Bucket 1	08-May-2012	0.79	0.57	1.39	3.67	0.3*	1811	43
	Bucket 2	02-Aug-2016	0.23	2.57	0.09	0.18	4.7	1028	10
Japan									
	Bucket 1	01-Aug-2014	0.65	0.41	1.59	4.14	1.2*	1759	56
	Bucket 2	03-Oct-2011	1.13	0.84	1.35	4.21	2.3*	2528	66
	Bucket 3	01-Jul-2011	1.33	2.15	0.62	1.95	6.9*	2582	87
United Kingdom									
	Bucket 1	16-Sep-2015	2.63	0.79	3.33	6.96	1.3*	1135	18
United States									
	Bucket 0	02-Oct-2015	0.62	0.12	5.17	12.37	0.3*	1488	152
	Bucket 1	02-Feb-2012	0.97	0.68	1.43	4.39	0.7*	2449	232
	Bucket 2	27-May-2011	0.52	0.58	0.9	2.86	0.5*	2624	146
	Bucket 3	29-Feb-2012	1.38	2.72	0.51	1.56	5.5	2435	45
France									
	Bucket 1	08-Aug-2013	2.8	0.68	4.12	11.57	1.1*	2052	29
	Bucket 2	24-Oct-2013	2.9	1.68	1.73	4.8	2.4	2000	22
Germany									
	Bucket 1	08-May-2012	1.81	0.67	2.7	8.14	1.1*	2365	52
	Bucket 2	10-Sep-2012	1.97	1.95	1.01	2.99	3.3	2278	27
Italy									
	Bucket 1	01-May-2012	2.4	0.79	3.04	9.08	0.5*	2321	66
	Bucket 2	18-Mar-2013	2.03	1.34	1.51	4.35	1.4*	2154	39
	Bucket 3	07-Feb-2018	2.02	3.12	0.65	1.16	6.8	828	10
Spain									
	Bucket 1	18-Sep-2012	3.41	0.83	4.11	11.68	0.1*	2099	38
	Bucket 2	04-Jun-2014	3.22	1.72	1.87	4.99	2	1848	24
Global									
	Bucket 0	22-Dec-2014	2.03	5.38	0.38	0.98	6.4*	1718	213
	Bucket 1	27-May-2011	1.74	5.78	0.3	0.95	7.8	2591	390
	Bucket 2	26-May-2011	0.74	3.21	0.23	0.72	1.1*	2580	221
	Bucket 3	01-Jul-2011	0.72	2.96	0.24	0.75	5.7*	2572	199

(To be continued)

(Continued)

(C) *6-1M Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	Bucket 1	23-Nov-2015	-0.84	1.33	-0.63	-1.12	4.3	829	14
	Bucket 2	24-Apr-2017	0.42	4.86	0.09	0.19	11.4	1140	14
	Bucket 3	24-Sep-2014	3.74	10.36	0.36	0.94	23.2	1785	19
Canada									
	Bucket 1	31-Oct-2011	0.06	1.39	0.04	0.12	3.8	2253	50
	Bucket 2	03-Jun-2016	-0.65	4.95	-0.13	-0.3	11.4	1360	11
Japan									
	Bucket 1	27-Jan-2014	-0.55	0.57	-0.96	-2.58	4.5	1885	68
	Bucket 2	29-Mar-2011	0.87	2.02	0.43	1.37	11.2	2655	68
	Bucket 3	27-Dec-2010	2.77	5.17	0.54	1.74	28.5	2712	91
United Kingdom									
	Bucket 1	09-Sep-2015	-0.68	1.66	-0.41	-0.86	4.7	1157	19
	Bucket 2	04-Jun-2014	2.55	4.87	0.52	1.3	10.8	1632	21
United States									
	Bucket 0	02-Oct-2015	-0.62	0.18	-3.44	-8.23	0.8	1487	152
	Bucket 1	02-Sep-2011	-0.35	1.78	-0.2	-0.63	4.5	2552	273
	Bucket 2	19-Nov-2010	1.15	4.84	0.24	0.78	9.6	2753	154
	Bucket 3	23-Aug-2011	3.89	14.6	0.27	0.85	38.3	2565	49
France									
	Bucket 1	27-Aug-2012	-0.46	1.13	-0.41	-1.19	4.4	2203	31
	Bucket 2	22-Apr-2013	1.57	3.97	0.4	1.14	11.1	2130	25
Germany									
	Bucket 1	31-Oct-2011	-0.47	1.27	-0.37	-1.15	4.3	2494	56
	Bucket 2	06-Mar-2012	1.54	4.24	0.36	1.1	11.2	2408	28
Italy									
	Bucket 1	24-Oct-2011	1.05	4.53	0.23	0.71	4.2	2501	69
	Bucket 2	06-Sep-2012	4.23	7.29	0.58	1.72	10.8	2282	41
	Bucket 3	02-Aug-2017	5.55	12.54	0.44	0.89	22.4	1055	12
Spain									
	Bucket 1	14-Mar-2012	-0.1	2.93	-0.03	-0.09	4	2401	40
	Bucket 2	25-Nov-2013	3.22	4.96	0.65	1.79	11.2	1978	27
Global									
	Bucket 0	16-Apr-2014	0.74	3.62	0.2	0.54	9.5	1880	267
	Bucket 1	05-Dec-2011	0.61	4.91	0.12	0.37	13.1	2456	471
	Bucket 2	29-Mar-2011	1.14	3.2	0.36	1.15	11.8	2646	328
	Bucket 3	27-Dec-2010	2.11	3.81	0.55	1.77	21.3	2703	189

(To be continued)

(Continued)

(D) *6-1M Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	Bucket 1	23-Nov-2015	2.55	1.18	2.16	3.86	0.7*	829	14
	Bucket 2	24-Apr-2017	0.9	2.01	0.45	0.94	2.5	1140	14
	Bucket 3	24-Sep-2014	1.65	2.23	0.74	1.94	3.3*	1785	19
Canada									
	Bucket 1	31-Oct-2011	0.8	0.55	1.45	4.27	0*	2253	50
	Bucket 2	03-Jun-2016	0.6	2.52	0.24	0.55	4.7	1360	11
Japan									
	Bucket 1	27-Jan-2014	0.72	0.42	1.71	4.6	1.2*	1885	68
	Bucket 2	29-Mar-2011	1.23	0.89	1.38	4.41	2.7*	2655	68
	Bucket 3	27-Dec-2010	1.39	2.29	0.61	1.97	7.3*	2712	91
United Kingdom									
	Bucket 1	09-Sep-2015	2.7	0.89	3.03	6.39	1.5*	1157	19
	Bucket 2	04-Jun-2014	2.91	2.15	1.35	3.38	3.3*	1632	21
United States									
	Bucket 0	02-Oct-2015	0.58	0.12	4.83	11.55	0.3*	1487	152
	Bucket 1	02-Sep-2011	0.99	0.76	1.3	4.07	0.8*	2552	273
	Bucket 2	19-Nov-2010	0.68	0.77	0.88	2.86	0.9*	2753	154
	Bucket 3	23-Aug-2011	1.26	2.81	0.45	1.41	5.6	2565	49
France									
	Bucket 1	27-Aug-2012	3.02	0.89	3.39	9.87	1.5*	2203	31
	Bucket 2	22-Apr-2013	2.98	1.89	1.58	4.52	2.8	2130	25
Germany									
	Bucket 1	31-Oct-2011	2.04	0.84	2.43	7.53	1.1*	2494	56
	Bucket 2	06-Mar-2012	2.22	2.11	1.05	3.2	3.7*	2408	28
Italy									
	Bucket 1	24-Oct-2011	3.07	1.21	2.54	7.88	0.9*	2501	69
	Bucket 2	06-Sep-2012	2.05	1.5	1.37	4.06	1.7*	2282	41
	Bucket 3	02-Aug-2017	2.16	2.98	0.72	1.45	7.2*	1055	12
Spain									
	Bucket 1	14-Mar-2012	3.23	1.19	2.71	8.24	0.7*	2401	40
	Bucket 2	25-Nov-2013	3.6	2.09	1.72	4.74	2.7	1978	27
Global									
	Bucket 0	16-Apr-2014	-1.14	4.34	-0.26	-0.7	4.3*	1880	267
	Bucket 1	05-Dec-2011	0.98	5.04	0.19	0.58	7.2*	2456	471
	Bucket 2	29-Mar-2011	0.64	4.43	0.14	0.45	3.1*	2646	328
	Bucket 3	27-Dec-2010	-0.25	3.07	-0.08	-0.26	7.9*	2703	189

(To be continued)

(Continued)

(E) *3-1M Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	Bucket 1	23-Nov-2015	-0.69	1.35	-0.51	-0.91	4.4	830	14
	Bucket 2	24-Apr-2017	0.48	4.81	0.1	0.21	11.4	1140	15
	Bucket 3	24-Mar-2014	3.78	10.26	0.37	0.99	23	1863	20
Canada									
	Bucket 1	23-Aug-2011	0.08	1.47	0.05	0.15	3.8	2401	51
	Bucket 2	11-Feb-2016	-1.14	4.89	-0.23	-0.53	11.3	1373	22
Japan									
	Bucket 1	08-Nov-2013	-0.54	0.56	-0.96	-2.62	4.6	1932	71
	Bucket 2	12-Jan-2011	0.91	2.07	0.44	1.42	11.4	2705	69
	Bucket 3	12-Oct-2010	2.36	5.27	0.45	1.47	28.8	2763	92
United Kingdom									
	Bucket 1	09-Sep-2015	-0.82	1.66	-0.49	-1.03	4.6	1157	20
	Bucket 2	20-Mar-2014	1.77	4.78	0.37	0.97	10.7	1788	21
United States									
	Bucket 0	02-Oct-2015	-0.62	0.18	-3.44	-8.23	0.8	1487	152
	Bucket 1	21-Jun-2011	-0.09	1.86	-0.05	-0.16	4.6	2603	281
	Bucket 2	08-Sep-2010	1.1	4.97	0.22	0.72	9.7	2805	158
	Bucket 3	09-Jun-2011	5.19	14.89	0.35	1.11	38.6	2617	51
France									
	Bucket 1	13-Jun-2012	-0.37	1.23	-0.3	-0.89	4.5	2304	31
	Bucket 2	05-Feb-2013	1.96	4.04	0.49	1.42	11.1	2182	31
Germany									
	Bucket 1	17-Aug-2011	-0.3	1.49	-0.2	-0.63	4.4	2546	58
	Bucket 2	19-Dec-2011	1.71	4.39	0.39	1.2	11.4	2460	30
Italy									
	Bucket 1	10-Aug-2011	0.98	4.64	0.21	0.66	4.3	2553	69
	Bucket 2	04-Sep-2012	4.59	7.33	0.63	1.87	10.8	2284	44
	Bucket 3	21-Nov-2016	5.67	12.44	0.46	0.95	22.1	1116	21
Spain									
	Bucket 1	28-Dec-2011	0.3	3.09	0.1	0.31	4.2	2453	40
	Bucket 2	29-Oct-2013	3.3	5.04	0.65	1.8	11.3	1996	29
Global									
	Bucket 0	21-Nov-2012	0.49	3.07	0.16	0.47	8.8	2211	306
	Bucket 1	13-Apr-2011	1.9	4.11	0.46	1.46	11.1	2616	560
	Bucket 2	29-Oct-2010	1.23	3.64	0.34	1.1	11.5	2733	374
	Bucket 3	12-Oct-2010	1.82	3.75	0.49	1.59	20.6	2753	186

(To be continued)

(Continued)

(F) *3-1M Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	Bucket 1	23-Nov-2015	2.67	1.2	2.23	3.98	0.9*	830	14
	Bucket 2	24-Apr-2017	1.25	1.89	0.66	1.38	2.4	1140	15
	Bucket 3	24-Mar-2014	1.54	2.04	0.75	2.01	2.7*	1863	20
Canada									
	Bucket 1	23-Aug-2011	0.85	0.6	1.42	4.32	0.1*	2401	51
	Bucket 2	11-Feb-2016	0.5	2.36	0.21	0.48	4.1*	1373	22
Japan									
	Bucket 1	08-Nov-2013	0.7	0.44	1.59	4.33	1.3*	1932	71
	Bucket 2	12-Jan-2011	1.28	0.9	1.42	4.58	2.9*	2705	69
	Bucket 3	12-Oct-2010	1.32	2.37	0.56	1.83	7.4*	2763	92
United Kingdom									
	Bucket 1	09-Sep-2015	2.83	0.84	3.37	7.11	1.4*	1157	20
	Bucket 2	20-Mar-2014	2.5	1.86	1.34	3.51	2.7*	1788	21
United States									
	Bucket 0	02-Oct-2015	0.59	0.12	4.92	11.77	0.3*	1487	152
	Bucket 1	21-Jun-2011	1.28	0.92	1.39	4.4	0.9*	2603	281
	Bucket 2	08-Sep-2010	0.69	0.88	0.78	2.56	1.1*	2805	158
	Bucket 3	09-Jun-2011	1.15	2.84	0.4	1.27	5.7	2617	51
France									
	Bucket 1	13-Jun-2012	3.07	0.95	3.23	9.62	1.5*	2304	31
	Bucket 2	05-Feb-2013	2.81	1.93	1.46	4.23	2.7*	2182	31
Germany									
	Bucket 1	17-Aug-2011	2.27	0.93	2.44	7.64	1.2*	2546	58
	Bucket 2	19-Dec-2011	2.32	2.19	1.06	3.26	3.8*	2460	30
Italy									
	Bucket 1	10-Aug-2011	2.94	1.2	2.45	7.68	1.1*	2553	69
	Bucket 2	04-Sep-2012	2.03	1.54	1.32	3.91	1.8*	2284	44
	Bucket 3	21-Nov-2016	0.98	2.77	0.35	0.73	6.1*	1116	21
Spain									
	Bucket 1	28-Dec-2011	3.35	1.18	2.84	8.72	0.9*	2453	40
	Bucket 2	29-Oct-2013	3.52	2.14	1.64	4.54	2.8*	1996	29
Global									
	Bucket 0	21-Nov-2012	-0.53	3.9	-0.14	-0.41	5.3*	2211	306
	Bucket 1	13-Apr-2011	1.31	5.09	0.26	0.82	6.4	2616	560
	Bucket 2	29-Oct-2010	0.35	4.85	0.07	0.23	1.7*	2733	374
	Bucket 3	12-Oct-2010	-0.36	4.74	-0.08	-0.26	9.4*	2753	186

$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

TABLE 3.3: **Value**

The table reports the results on the construction of the value factor, considering the issuer-specific bonds for the country portfolios, and all bonds for the global portfolio. I use the canonical 5Y Reversal definition of Asness, Moskowitz, and Pedersen (2013) (i.e. the negative of the 5-year change in the bond yields), panels 3.3a-3.3b, and an alternative measure with a shorter look-back window of 3 years, panels 3.3c-3.3d. For each measure I report the results of the long-only and long-short portfolios. I require a minimum of five bonds to be traded at each time t per maturity bucket, and a minimum of three years of data. Frequency is daily. Column *Start* indicate the start date of the time-series of factor returns, *Avg (%)* is the estimated average return per annum in %, *Std (%)* the annualised standard deviation, *SR* is the Sharpe ratio, *t-Stat* is its t-statistic, and $\Delta T(yr)$ is the time-series average of the portfolio term-to-maturity in the relevant sample, expressed in years. I indicate with an asterisk if at any time in the sample period the portfolio net term-to-maturity turns negative (only for the long-short portfolios). Finally T refers to the number of time-series observations (in days), and N to the number of bonds considered in the whole sample for the country-bucket portfolios. All factors are estimated up to October 29th 2021.

(A) *5Y Reversal Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Japan									
	Bucket 1	30-Jan-2017	-0.28	2.53	-0.11	-0.2	18.9	896	16
	Bucket 3	20-May-2016	-1.23	3.52	-0.35	-0.8	23.4	1346	28
United States									
	Bucket 3	13-Jul-2018	3.59	7.05	0.51	0.91	17.3	832	15
Global									
	Bucket 1	20-May-2016	-0.45	3.7	-0.12	-0.27	14.2	1348	75
	Bucket 2	14-Jun-2017	0.21	4.85	0.04	0.08	13.3	1007	53
	Bucket 3	02-Sep-2015	-0.85	2.15	-0.4	-0.97	10.5	1524	101

(B) *5Y Reversal Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Japan									
	Bucket 1	30-Jan-2017	0.81	1.13	0.72	1.34	1.8*	896	16
	Bucket 3	20-May-2016	0.78	1	0.78	1.77	2.8*	1346	28
United States									
	Bucket 3	13-Jul-2018	1.5	3.02	0.5	0.89	-0.7*	832	15
Global									
	Bucket 1	20-May-2016	1.65	5.2	0.32	0.73	0.2*	1348	75
	Bucket 2	14-Jun-2017	4.31	7.19	0.6	1.18	5.5*	1007	53
	Bucket 3	02-Sep-2015	-1.57	4.64	-0.34	-0.82	-12.8*	1524	101

(To be continued)

(Continued)

(C) *3Y Reversal Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Japan									
	Bucket 1	23-Sep-2015	0.71	3.85	0.18	0.42	21.9	1387	21
	Bucket 3	29-Apr-2014	1.65	4.76	0.35	0.93	25.1	1855	34
United States									
	Bucket 0	05-Jul-2016	1.69	5.07	0.33	0.63	11.2	945	53
	Bucket 1	20-Sep-2013	1.52	5.96	0.26	0.73	16.3	2039	76
	Bucket 3	23-Jun-2015	1.58	5.81	0.27	0.67	15.6	1600	25
France									
	Bucket 1	30-May-2018	-0.04	3.05	-0.01	-0.02	8.9	823	10
Germany									
	Bucket 1	07-Sep-2016	-1.53	1.8	-0.85	-1.87	6.1	1264	19
Italy									
	Bucket 1	03-Nov-2017	1.5	6.46	0.23	0.45	8.5	975	17
Global									
	Bucket 0	04-Sep-2015	0.13	6.11	0.02	0.04	17.8	1057	72
	Bucket 1	05-Nov-2013	1.76	5.02	0.35	0.96	17	1966	162
	Bucket 2	03-Oct-2013	0.05	4.07	0.01	0.03	12.6	2004	126
	Bucket 3	06-Aug-2013	0.56	2.61	0.21	0.59	13.3	2036	130

(D) *3Y Reversal Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Japan									
	Bucket 1	23-Sep-2015	1.06	1.55	0.68	1.57	4.4*	1387	21
	Bucket 3	29-Apr-2014	0.36	1.12	0.32	0.85	2.4*	1855	34
United States									
	Bucket 0	05-Jul-2016	0.03	5.75	0.01	0.02	-7.3*	945	53
	Bucket 1	20-Sep-2013	-0.24	4.21	-0.06	-0.17	-5*	2039	76
	Bucket 3	23-Jun-2015	1.61	4.05	0.4	0.99	2.1*	1600	25
France									
	Bucket 1	30-May-2018	2.34	2	1.17	2.08	0.2*	823	10
Germany									
	Bucket 1	07-Sep-2016	1.9	1.44	1.32	2.91	3*	1264	19
Italy									
	Bucket 1	03-Nov-2017	1.17	3.14	0.37	0.72	-1.8*	975	17
Global									
	Bucket 0	04-Sep-2015	1.85	7.04	0.26	0.52	1.3*	1057	72
	Bucket 1	05-Nov-2013	0.6	6.93	0.09	0.25	2.5*	1966	162
	Bucket 2	03-Oct-2013	0.86	5.55	0.15	0.42	4.4*	2004	126
	Bucket 3	06-Aug-2013	-1.82	4.66	-0.39	-1.09	-15.8*	2036	130

$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

TABLE 3.4: **Low-risk**

The table reports the results on the construction of the low-risk factor, considering the issuer-specific bonds for the country portfolios, and all bonds for the global portfolio. I use duration (Dur) as in Frazzini and Pedersen (2014), panels 3.3a-3.3b, and the bonds' term-to-maturity (TTM), panels 3.3c-3.3d. For each measure I report the results of the long-only and long-short portfolios. I require a minimum of five bonds to be traded at each time t, and a minimum of three years of data. Frequency is daily. Column *Start* indicate the start date of the time-series of factor returns, *Avg (%)* is the estimated average return per annum in %, *Std (%)* the annualised standard deviation, *SR* is the Sharpe ratio, *t-Stat* is its t-statistic, and $\Delta T(yr)$ is the time-series average of the portfolio term-to-maturity in the relevant sample, expressed in years. I indicate with an asterisk if at any time in the sample period the portfolio net term-to-maturity turns negative (only for the long-short portfolios). Finally T refers to the number of time-series observations (in days), and N to the number of bonds considered in the whole sample for the country-bucket portfolios. All factors are estimated up to October 29th 2021.

(A) *Dur Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Canada									
	All Buckets	22-Jul-2015	0.95	3.23	0.29	0.71	7.2	1543	35
Japan									
	All Buckets	21-Mar-2014	2.09	5.12	0.41	1.09	25.2	1839	61
United Kingdom									
	All Buckets	04-Aug-2016	2.37	7.5	0.32	0.71	17.1	1280	27
United States									
	All Buckets	19-Feb-2014	2.38	6.39	0.37	0.97	15.7	1795	231
France									
	All Buckets	10-Feb-2015	1.74	4.56	0.38	0.96	12.6	1652	27
Germany									
	All Buckets	20-Jan-2015	0.54	3.34	0.16	0.4	8.5	1649	45
Italy									
	All Buckets	17-Jun-2014	3.84	7.44	0.52	1.36	11.5	1781	61
Spain									
	All Buckets	05-Mar-2015	2.41	5.4	0.45	1.13	12.5	1626	34
Global									
	All Buckets	17-Oct-2013	2.13	5.6	0.38	0.94	15.3	1601	539

(To be continued)

(Continued)

(B) *Dur Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Canada									
	All Buckets	22-Jul-2015	1.64	2.53	0.65	1.58	2.1*	1543	35
Japan									
	All Buckets	21-Mar-2014	1.01	1.73	0.58	1.54	2.9*	1839	61
United Kingdom									
	All Buckets	04-Aug-2016	3.54	5.66	0.63	1.4	10.5*	1280	27
United States									
	All Buckets	19-Feb-2014	0.24	1.91	0.13	0.34	-2.2*	1795	231
France									
	All Buckets	10-Feb-2015	2.28	2.2	1.04	2.62	3.3*	1652	27
Germany									
	All Buckets	20-Jan-2015	1.08	1.19	0.91	2.29	0.1*	1649	45
Italy									
	All Buckets	17-Jun-2014	2.78	2.42	1.15	3.01	2.9*	1781	61
Spain									
	All Buckets	05-Mar-2015	2.48	2.47	1	2.5	4*	1626	34
Global									
	All Buckets	17-Oct-2013	-0.48	5.76	-0.08	-0.2	1.1*	1601	539

(C) *TTM Long-only*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	All Buckets	28-Oct-2011	3.58	8.21	0.44	1.36	16.9	2491	32
Canada									
	All Buckets	14-Dec-2010	0.81	2.64	0.31	1	6	2680	62
Japan									
	All Buckets	28-Apr-2010	1.67	3.9	0.43	1.4	21.4	2747	184
United Kingdom									
	All Buckets	09-Mar-2012	2.24	6.2	0.36	1.09	14.1	2378	35
United States									
	All Buckets	16-Mar-2010	2.04	6.13	0.33	1.06	13.3	2706	385
France									
	All Buckets	26-Jan-2011	3.32	4.26	0.78	2.49	11.2	2647	45
Germany									
	All Buckets	27-Sep-2010	1.67	3.75	0.45	1.45	8.7	2703	72
Italy									
	All Buckets	30-Sep-2010	2.46	7.32	0.34	1.09	9.9	2677	98
Spain									
	All Buckets	10-Nov-2010	3.13	6.52	0.48	1.54	10.7	2683	59
Global									
	All Buckets	17-Feb-2010	1.53	5.06	0.3	0.88	12.4	2258	972

(To be continued)

(Continued)

(D) *TTM Long-Short*

		Start	Avg (%)	Std (%)	SR	t-Stat	ΔT (yr)	T	N
Australia									
	All Buckets	28-Oct-2011	3.9	4.87	0.8	2.48	7.8*	2491	32
Canada									
	All Buckets	14-Dec-2010	1.25	1.6	0.78	2.5	1.5*	2680	62
Japan									
	All Buckets	28-Apr-2010	0.35	1.84	0.19	0.62	4.3*	2747	184
United Kingdom									
	All Buckets	09-Mar-2012	2.75	4.23	0.65	1.97	1*	2378	35
United States									
	All Buckets	16-Mar-2010	0.88	2.49	0.35	1.13	1.3	2706	385
France									
	All Buckets	26-Jan-2011	3.14	2.48	1.27	4.05	4.1*	2647	45
Germany									
	All Buckets	27-Sep-2010	2.74	2	1.37	4.42	3.6*	2703	72
Italy									
	All Buckets	30-Sep-2010	3.38	2.84	1.19	3.82	3.3*	2677	98
Spain									
	All Buckets	10-Nov-2010	3.53	2.51	1.41	4.53	3.4*	2683	59
Global									
	All Buckets	17-Feb-2010	-0.12	5.45	-0.02	-0.06	-2.5*	2258	972

$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

TABLE 3.5: **Multi-Factor Results - Unrestricted Sample**

The table reports the multi-factor results considering all available data for each factor, country, and bucket combination (unrestricted sample). The information content of this table is equivalent to tables 3.2, 3.3, and 3.4 combined, considering only bucket 1 and bucket 3 bonds for value and momentum. Panel 3.5a shows the results for long-only portfolios, and panel 3.5b for long-short ones. Column *Start* indicate the start date of the time-series of factor returns, *SR* is the Sharpe ratio, and ΔT is the time-series average of the portfolio term-to-maturity in the relevant sample, expressed in years. I indicate an with asterisk if at any time in the sample period the portfolio net term-to-maturity turns negative (only for the long-short portfolios). Sharpe ratios in bold have a t-stat greater than 1.96, indicating significance at 5% level. All factors are estimated up to October 29th 2021.

(A) *Long-only*

		Momentum				Value				Low-risk			
		Start	Measure	SR	ΔT	Start	Measure	SR	ΔT	Start	Measure	SR	ΔT
Canada													
	Bucket 1	08-May-2012	12-1M	0.2	3.8								
	All Buckets									22-Jul-2015	Dur	0.29	7.2
Japan													
	Bucket 1	01-Aug-2014	12-1M	**-1.14**	4.5	23-Sep-2015	3Y Rev	0.18	21.9				
	Bucket 3	01-Jul-2011	12-1M	0.54	27.8	20-May-2016	5Y Rev	-0.35	23.4				
	All Buckets									21-Mar-2014	Dur	0.41	25.2
United Kingdom													
	Bucket 1	16-Sep-2015	12-1M	-0.59	4.5								
	All Buckets									04-Aug-2016	Dur	0.32	17.1
United States													
	Bucket 1	02-Feb-2012	12-1M	-0.22	4.5	20-Sep-2013	3Y Rev	0.26	16.3				
	Bucket 3	29-Feb-2012	12-1M	0.23	37.8	13-Jul-2018	5Y Rev	0.51	17.3				
	All Buckets									19-Feb-2014	Dur	0.37	15.7
France													
	Bucket 1	08-Aug-2013	12-1M	**-0.86**	4.1	30-May-2018	3Y Rev	-0.01	8.9				
	All Buckets									10-Feb-2015	Dur	0.38	12.6
Germany													
	Bucket 1	08-May-2012	12-1M	**-0.98**	4	07-Sep-2016	3Y Rev	**-0.85**	6.1				
	All Buckets									20-Jan-2015	Dur	0.16	8.5
Italy													
	Bucket 1	01-May-2012	12-1M	-0.05	3.9	03-Nov-2017	3Y Rev	0.23	8.5				
	Bucket 3	07-Feb-2018	12-1M	0.36	21.8								
	All Buckets									17-Jun-2014	Dur	0.52	11.5
Spain													
	Bucket 1	18-Sep-2012	12-1M	-0.34	3.7								
	All Buckets									05-Mar-2015	Dur	0.45	12.5
Global													
	Bucket 1	27-May-2011	12-1M	0.09	10.2	05-Nov-2013	3Y Rev	0.35	17				
	Bucket 3	01-Jul-2011	12-1M	0.62	22.1	02-Sep-2015	5Y Rev	-0.4	10.5				
	All Buckets									17-Oct-2013	Dur	0.38	15.3

(To be continued)

(Continued)

(B) *Long-Short*

		Momentum				Value				Low-risk			
		Start	Measure	SR	ΔT	Start	Measure	SR	ΔT	Start	Measure	SR	ΔT
Canada													
	Bucket 1	08-May-2012	12-1M	**1.39**									
	All Buckets									22-Jul-2015	Dur	0.65	2.1*
Japan													
	Bucket 1	01-Aug-2014	12-1M	**1.59**	1.2*	23-Sep-2015	3Y Rev	0.68	4.4*				
	Bucket 3	01-Jul-2011	12-1M	0.62	6.9*	20-May-2016	5Y Rev	0.78	2.8*				
	All Buckets									21-Mar-2014	Dur	0.58	2.9*
United Kingdom													
	Bucket 1	16-Sep-2015	12-1M	**3.33**	1.3*								
	All Buckets									04-Aug-2016	Dur	0.63	10.5*
United States													
	Bucket 1	02-Feb-2012	12-1M	**1.43**	0.7*	20-Sep-2013	3Y Rev	-0.06	-5*				
	Bucket 3	29-Feb-2012	12-1M	0.51	5.5	13-Jul-2018	5Y Rev	0.5	-0.7*				
	All Buckets									19-Feb-2014	Dur	0.13	-2.2*
France													
	Bucket 1	08-Aug-2013	12-1M	**4.12**	1.1*	30-May-2018	3Y Rev	**1.17**	0.2*				
	All Buckets									10-Feb-2015	Dur	**1.04**	3.3*
Germany													
	Bucket 1	08-May-2012	12-1M	**2.7**	1.1*	07-Sep-2016	3Y Rev	**1.32**	3*				
	All Buckets									20-Jan-2015	Dur	**0.91**	0.1*
Italy													
	Bucket 1	01-May-2012	12-1M	**3.04**	0.5*	03-Nov-2017	3Y Rev	0.37	-1.8*				
	Bucket 3	07-Feb-2018	12-1M	0.65	6.8								
	All Buckets									17-Jun-2014	Dur	**1.15**	2.9*
Spain													
	Bucket 1	18-Sep-2012	12-1M	**4.11**	0.1*								
	All Buckets									05-Mar-2015	Dur	**1**	4*
Global													
	Bucket 1	27-May-2011	12-1M	0.3	7.8	05-Nov-2013	3Y Rev	0.09	2.5*				
	Bucket 3	01-Jul-2011	12-1M	0.24	5.7*	02-Sep-2015	5Y Rev	-0.34	-12.8*				
	All Buckets									17-Oct-2013	Dur	-0.08	1.1*

$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

TABLE 3.6: **Multi-Factor Results - Restricted Sample**

The table reports the multi-factor results from October 2018 to October 2021 for each factor, country, and bucket combination. The number of time series observations is $T = 677$ days. Panel 3.6a shows the results for long-only portfolios, and panel 3.6b for long-short ones. Column *SR* is the Sharpe ratio, and ΔT is the time-series average of the portfolio term-to-maturity in the relevant sample, expressed in years. I indicate an with asterisk if at any time in the sample period the portfolio net term-to-maturity turns negative (only for the long-short portfolios). Sharpe ratios in bold have a t-stat greater than 1.96, indicating significance at 5% level.

(A) *Long-only*

		Momentum			Value			Low-risk		
		Measure	SR	ΔT	Measure	SR	ΔT	Measure	SR	ΔT
Canada										
	Bucket 1	12-1M	0.4	4.1						
	All Buckets							Dur	0.85	6.8
Japan										
	Bucket 1	12-1M	**-1.33**	4.3	3Y Rev	-0.09	23			
	Bucket 3	12-1M	0.12	27	5Y Rev	-0.01	22.9			
	All Buckets							Dur	0.17	20.3
United Kingdom										
	Bucket 1	12-1M	-0.41	4.5						
	All Buckets							Dur	0.56	19.5
United States										
	Bucket 1	12-1M	0.8	4	3Y Rev	0.9	13.7			
	Bucket 3	12-1M	0.63	36.1	5Y Rev	0.77	16.7			
	All Buckets							Dur	0.48	11.9
France										
	Bucket 1	12-1M	-1.04	4.4	3Y Rev	0.04	8.7			
	All Buckets							Dur	0.44	11.1
Germany										
	Bucket 1	12-1M	**-1.33**	4.2	3Y Rev	-0.65	6.2			
	All Buckets							Dur	0.11	7.9
Italy										
	Bucket 1	12-1M	0.51	4.1	3Y Rev	1.03	8			
	Bucket 3	12-1M	0.81	22						
	All Buckets							Dur	**1.19**	10.6
Spain										
	Bucket 1	12-1M	-0.97	3.9						
	All Buckets							Dur	1.11	12.7
Global										
	Bucket 1	12-1M	0.6	8.1	3Y Rev	0.65	14.8			
	Bucket 3	12-1M	0.38	18	5Y Rev	-0.2	9.6			
	All Buckets							Dur	0.33	11.9

(To be continued)

(Continued)

(B) *Long-Short*

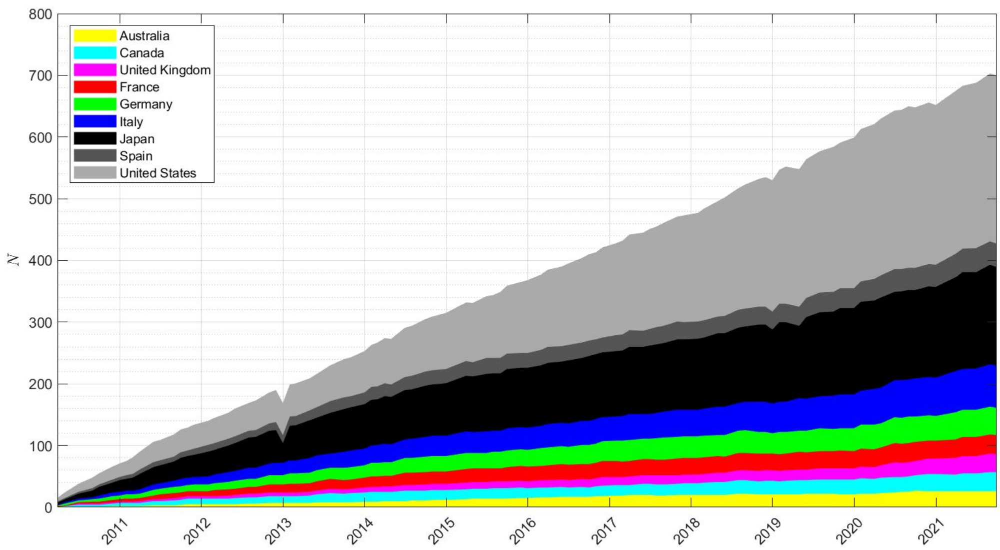
		Momentum			Value			Low-risk		
		Measure	SR	ΔT	Measure	SR	ΔT	Measure	SR	ΔT
Canada										
	Bucket 1	12-1M	0.47	0*						
	All Buckets							Dur	1.18	0.6*
Japan										
	Bucket 1	12-1M	**1.79**	2.5	3Y Rev	0.78	6.4*			
	Bucket 3	12-1M	0.85	10.6	5Y Rev	1.03	4.5*			
	All Buckets							Dur	0.71	-0.8*
United Kingdom										
	Bucket 1	12-1M	**3.09**	1.7*						
	All Buckets							Dur	0.79	11.9*
United States										
	Bucket 1	12-1M	**1.22**	0.9	3Y Rev	-0.05	-8*			
	Bucket 3	12-1M	0.69	8.5	5Y Rev	0.84	-2.1*			
	All Buckets							Dur	0.11	-1.4*
France										
	Bucket 1	12-1M	**4.58**	1.7*	3Y Rev	1.1	-0.4*			
	All Buckets							Dur	**1.36**	2.5*
Germany										
	Bucket 1	12-1M	**3.62**	1.7	3Y Rev	**1.48**	3.5			
	All Buckets							Dur	0.94	1.8*
Italy										
	Bucket 1	12-1M	**5.22**	0.5*	3Y Rev	0.38	-2.8*			
	Bucket 3	12-1M	0.9	6.6*						
	All Buckets							Dur	**1.59**	2.7*
Spain										
	Bucket 1	12-1M	**11.46**	-0.1*						
	All Buckets							Dur	**1.89**	5.5*
Global										
	Bucket 1	12-1M	1.07	3.3	3Y Rev	1.04	2.5*			
	Bucket 3	12-1M	0.74	-9*	5Y Rev	0.53	-9.2*			
	All Buckets							Dur	-0.02	-0.4*

$T = 677$ (31st Oct 2018 - 29th Oct 2021)

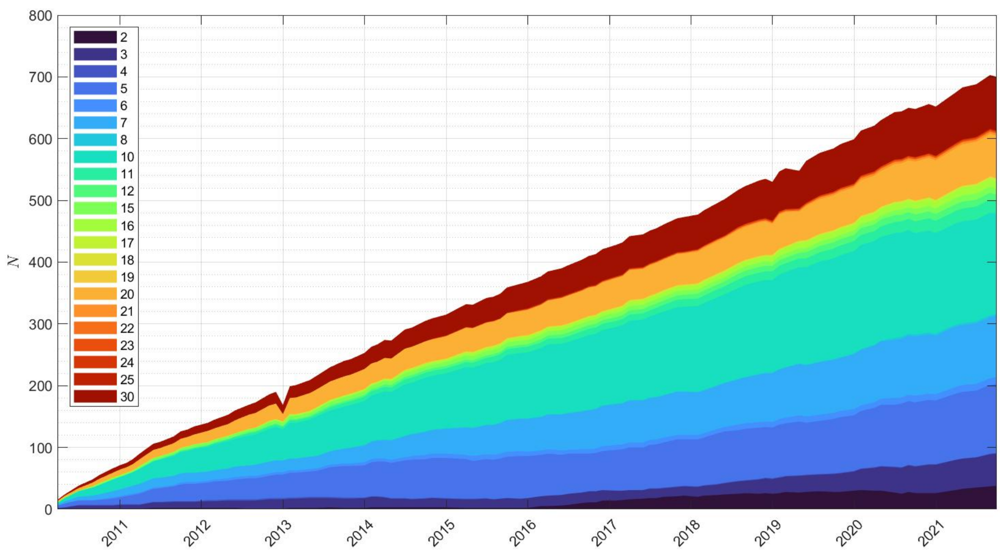
FIGURE 3.1: **Bond Universe**

The figure shows the time-varying composition of the bond universe by country, panel 3.1a, and by maturity (in years), panel 3.1b. At the end of each month I count the number of bonds with available clean prices and group the figures by issuer, or maturity at issuance. The total number of bonds issued in the sample period is 972.

(A) *Issuers*



(B) *Maturities*

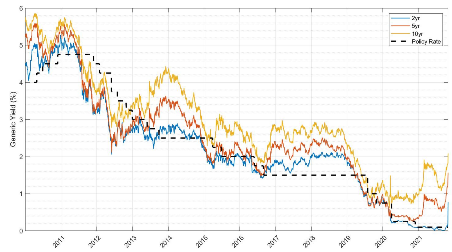


$T = 98$ (monthly, 31st Mar 2010 - 30th Sep 2021)

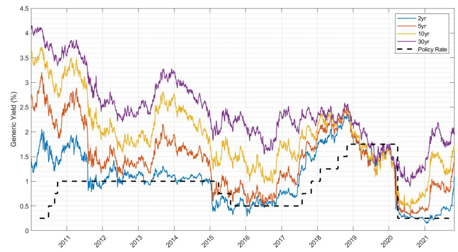
FIGURE 3.2: **Market Data**

The figure shows the time series of the constant-maturity generic (benchmark) yields for the 2 year (blue), 5 year (red), 10 year (yellow) and 30 year (purple) points for each issuer in the sample period, together with EUR swap rates. The dashed line is the relevant monthly policy rate for each country.

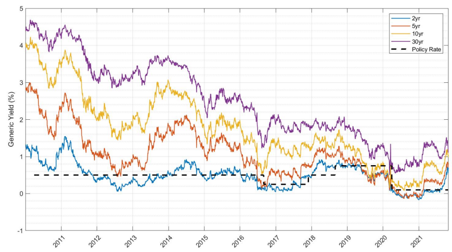
(A) *Australia*



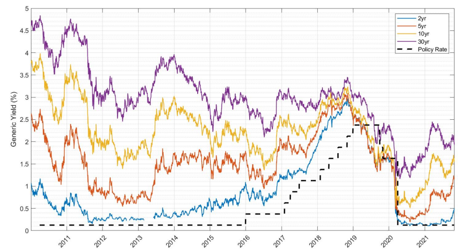
(B) *Canada*



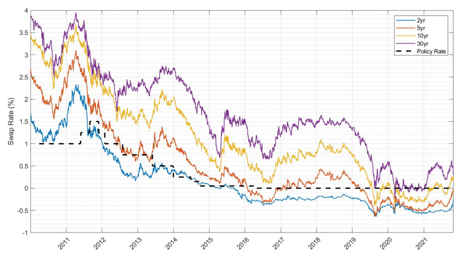
(C) *United Kingdom*



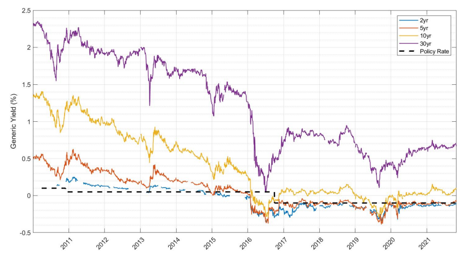
(D) *United States*



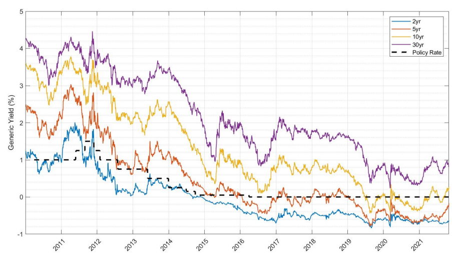
(E) *Euro Area*



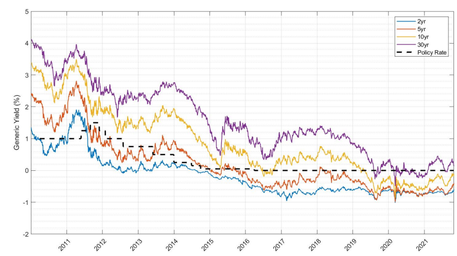
(F) *Japan*



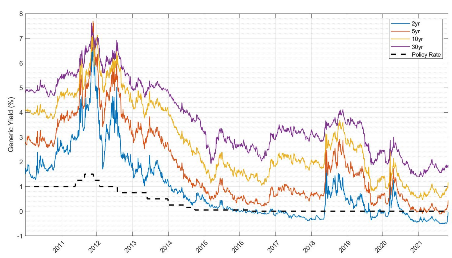
(G) *France*



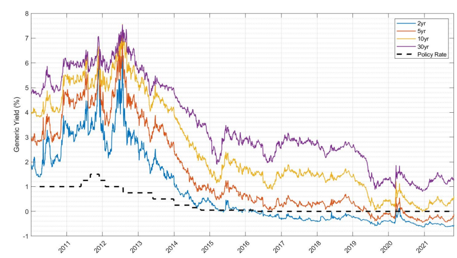
(H) *Germany*



(I) *Italy*



(J) *Spain*

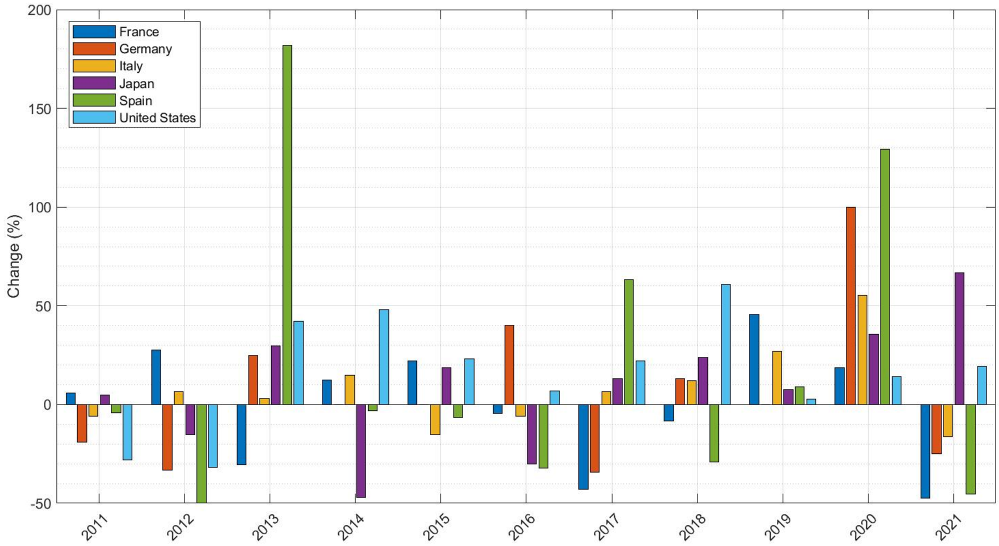


$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

FIGURE 3.3: **Issue Volume**

The figure shows country-specific summary statistics on the yearly issue size from 2010 to 2021, using bond-level data. Panel 3.3a reports the year-on-year change (in %) in the amount of debt sold via syndications, issuances and tap issues, considering the bonds of each issuer in the sample period. In panel 3.3b I show the time-series of the total amount of debt issued (in local currency). I report the results for all issuers in panel 3.3b and I exclude Canada and Australia in panel 3.3a.

(A) *Yearly Change*



(B) *Total Amount*

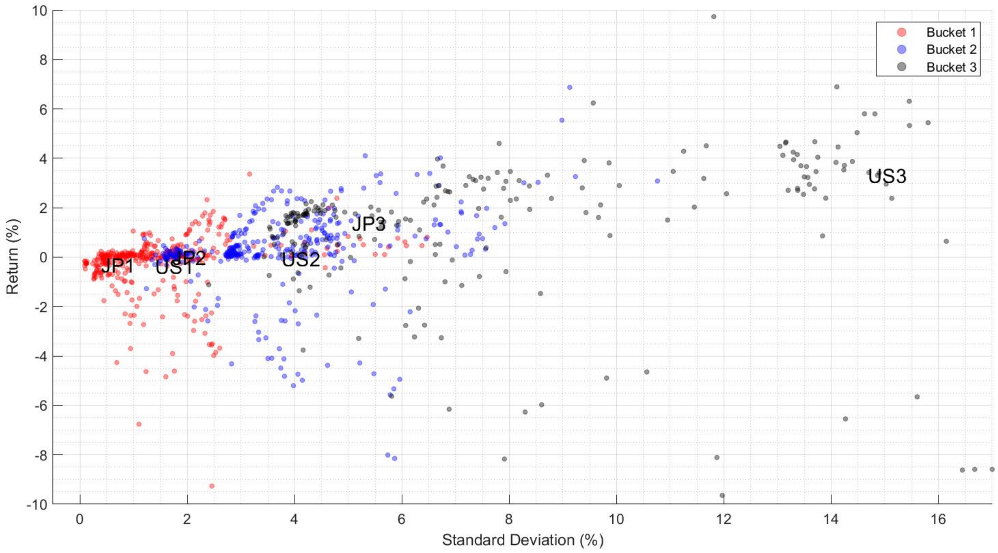
Year	FR	GE (€B)	IT	SP	AU (A$B)	CA (C$B)	JP (¥T)	UK (£B)	US ($B)
2010	17	37	33	23	2	22	31	12	627
2011	18	30	31	22	5	13	32	7	451
2012	23	20	33	11	4	9	27	7	308
2013	16	25	34	31	11	15	35	9	438
2014	18	25	39	30	16	9	19	7	648
2015	22	25	33	28	10	6	22	3	797
2016	21	35	31	19	19	22	15	8	851
2017	12	23	33	31	24	22	17	5	1038
2018	11	26	37	22	16	22	22	8	1670
2019	16	26	47	24	3	17	23	6	1716
2020	19	52	73	55	99	44	32	47	1959
2021	10	39	61	30		26	53	26	2337

$T = 11$ (yearly, 15th Jan 2010 - 29th Oct 2021)

FIGURE 3.4: **Risk-Return Profile of the Bonds**

The figure shows the risk-return profile of the 972 bonds in the sample period by maturity bucket. In panel 3.4a I report the ex-post average return (y-axis) and standard deviation (x-axis) of each security, colored by maturity bucket as issuance. Bucket 1 (red) includes bonds with 1 to 5 years of maturity, bucket 2 (blue) 6-10 years, and bucket 3 (grey) 11-30 years. Results are in % and annualised. Panel 3.4b reports the summary statistics of the equally-weighted country- and maturity-style portfolios. In panel 3.4a I also include the risk-return profile of the bonds for the two biggest contributors to the universe, Japan (JP) and the United States (US), by maturity bucket. JP1 is the point that corresponds to the risk-return profile of the $1/N$ long-only portfolio of Japanese bonds with 1 to 5 year-maturity, JP2 considers the bonds in bucket 2, and JP3 those in bucket 3. Similarly for US Treasury bonds: US1, US2, and US3.

(A) *Risk-Return*



(B) *Summary Statistics*

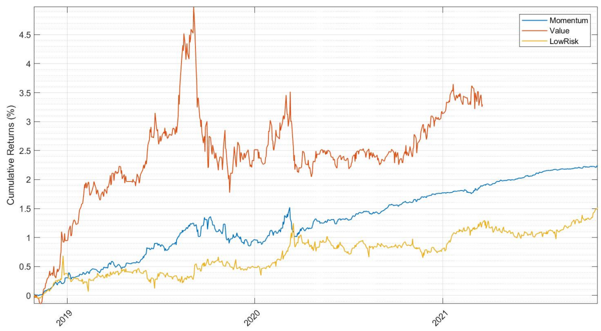
	Bucket 1			Bucket 2			Bucket 3		
Issuer	Avg (%)	Std (%)	N	Avg (%)	Std (%)	N	Avg (%)	Std (%)	N
Australia	-0.44	1.73	7	-1.06	2.9	4	0.71	8.3	21
Canada	-1.41	1.28	49	-2.36	4	4	0.56	4.95	9
Japan	-0.28	0.39	38	0.01	1.71	53	1.4	5.07	93
United Kingdom	-1.66	1.41	4	-0.34	3.49	20	-0.22	11.67	11
United States	-0.35	1.4	170	-0.03	3.76	162	3.32	14.68	53
Euro Area									
France	0.02	1.41	15	0.51	3.43	20	1.33	5.99	10
Germany	-0.51	1.07	40	0	3.71	30	-1.81	6.24	2
Italy	0.31	3.24	46	1.18	6.62	36	3	10.44	16
Spain	0.09	2.33	27	0.9	4.87	23	2.57	7.7	9

$T = 2696$ (15th Jan 2010 - 29th Oct 2021)

FIGURE 3.5: **Long-Short Bucket 1 Factors by Country**

The figure shows the time-series of cumulative returns on the three factors (left panels), together with the relevant summary statistics (right panels) in the sample period, October 2018 to October 2021 (daily). I report the results only for the countries in which all three factors are available, focusing on bucket 1 long-short sorts. In the time-series plots, blue refers to momentum, red to value, and yellow to low-risk. On the right panels, I report the estimated factor correlation matrix, the Sharpe ratio, row *SR*, and the average portfolio term-to-maturity, row ΔT, expressed in years. Sharpe ratios in bold have a t-stat greater than 1.96, indicating significance at 5% level. If at any time in the sample period the portfolio net term-to-maturity turns negative, I report the figure with an asterisk. All factors are estimated up to October 29th 2021, unless noted otherwise.

(A) *Japan*

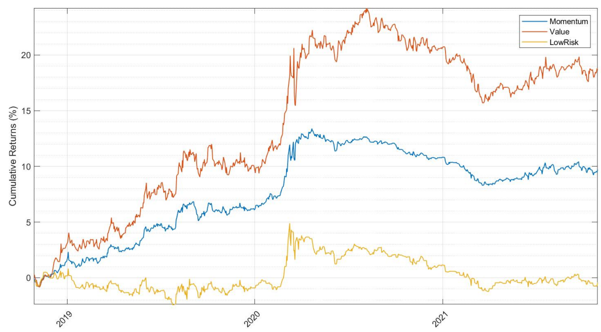


	Mom	Val	Low
Mom	1		
Val	0.05	1	
Low	-0.13	-0.05	1
SR	**1.79**	0.78	0.71
ΔT	2.5	6.4*	-0.8*

(B) *United States*

	Mom	Val	Low
Mom	1		
Val	0.04	1	
Low	-0.23	0	1
SR	**1.22**	-0.05	0.11
ΔT	0.9	-8*	-1.4*

(C) *Global*



	Mom	Val	Low
Mom	1		
Val	-0.09	1	
Low	0.91	-0.16	1
SR	1.07	1.04	-0.02
ΔT	3.3	2.5*	-0.4*

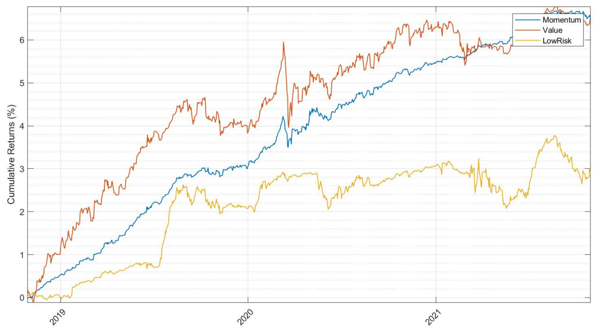
(To be continued)

(Continued)

(D) *France*

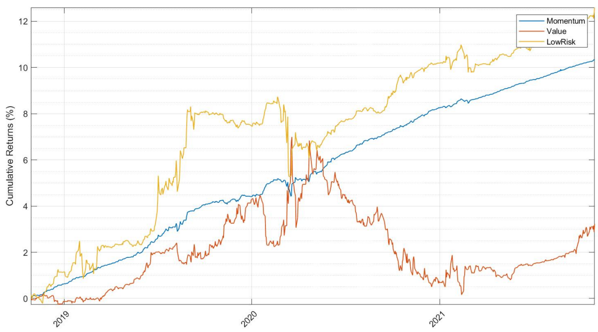
	Mom	Val	Low
Mom	1		
Val	0.1	1	
Low	0.37	-0.01	1
SR	**4.58**	1.1	1.36
ΔT	1.7*	-0.4*	2.5*

(E) *Germany*



	Mom	Val	Low
Mom	1		
Val	0.01	1	
Low	0.46	0.01	1
SR	**3.62**	**1.48**	0.94
ΔT	1.7*	3.5	1.8*

(F) *Italy*



	Mom	Val	Low
Mom	1		
Val	0.02	1	
Low	0.76	0.04	1
SR	**5.22**	0.38	**1.59**
ΔT	0.5*	-2.8*	2.7*

$T = 677$ (31st Oct 2018 - 29th Oct 2021)

General Conclusion

In this section I make the concluding remarks of this thesis and summarise further research questions that expand on the limitations of my analysis.

In this thesis I studied international stock returns via factor models featuring time-varying factor sensitivities, Chapters 1 and 2, and international factor premia in government bond returns for momentum, value, and low-risk factors, Chapter 3. The findings of my analysis are based on large panels of individual securities from different countries with historical data from 2010 to 2021. The data is sampled at a relatively higher frequency with respect to most contributions in international asset pricing, which provides a further perspective to interpret existing results in the literature and document new patterns in the higher-frequency returns of asset prices.

The main results of this thesis can be summarised as follows. Firstly, I compare the in-sample performance of the maximum-likelihood estimator of the time-varying betas in Borghi et al. (2018) against the rolling least square estimator, which suffers from severe mis-specification issues when the window size is chosen arbitrarily. Comparing dynamic-loadings models versus static counterparts, I find that the latter dominates the former when the variance of the rolling betas is pronounced, which coincides with a relatively short estimation window. Although the maximum-likelihood estimator provides the best in-sample performance from a statistical perspective, my analysis needs to be validated out-of-sample, a research question which remains open.

Secondly, I document how the explanatory and predictive power of time-varying sensitivities estimated via rolling least squares in linear asset pricing models of stock returns changes solely with respect to the window size. I find that the ability of the models to forecast future returns can improve by about 10%, in terms of out-of-sample R^2 by keeping the sampling frequency fixed, when the window contains about two years of recent information. This is because there appears to be a trade-off between statistical accuracy, which calls for a larger window size, and economic relevance of the estimates, which decreases with the window size. However, my results are not consistent when I employ standard mean-squared or -absolute error functions, which are minimised when a substantially larger window size is used for estimation. This suggests that a more comprehensive review of alternative objective functions is needed to gauge the models performance for varying window sizes, a research question that remains open. Moreover, I do not analyse how the models performance change for varying sampling frequency and look-ahead windows, keeping the estimation window (rolling) fixed. Although I briefly explain how the sampling frequency interacts with the distribution of time-varying estimates for a single stock and factor, this topic remains relatively unexplored in the literature and as Robertson (2018) reports, the behavior of rolling betas depends on the interplay between window size and sampling frequency.

Thirdly, I examine international risk premia in sovereign bonds by replicating characteristics-based portfolios such as momentum, value, and low-risk from asset-level data of nine developed countries. My analysis reveals a substantial variation in the factor premia in the cross sections, country- (issuer) and maturity-wise, which does not support the evidence of Asness, Moskowitz, and Pedersen (2013) and Frazzini and Pedersen (2014) on their unifying pricing ability across countries and asset classes. Contrarily to what reported in the literature, see e.g. Baltussen, Martens, and Penninga (2021), I find no supporting evidence for the existence of statistically significant positive risk premia for characteristics-based global portfolios that include bonds from all countries. Additionally, I find that momentum produces consistent statistically significant Sharpe ratios only for short-dated securities, while the performance of value and low-risk is mixed, and seem to be relevant only for bonds issued by specific countries. My analysis of risk premia on bond returns is only a first step to bridge the gap in the literature with the studies that use factor models for the analysis of stock returns (locally and internationally), which are the vast majority. In particular, I believe that further research is needed to understand how characteristic-based factors can help explain cross-sectional return patterns, as well as the potential of unobserved factors. These topics remain relatively unexplored in the context of international sovereign bonds, and in this thesis I paid tribute to the complexity of such question by focusing on the construction of three observed pricing factors and on their realised performance in the cross sections.

Bibliography

Andersen, Torben G et al. (2006). "Realized beta: Persistence and predictability". In: *Advances in Econometrics* 20.2, pp. 1–39.

Ang, Andrew, Joseph Chen, and Yuhang Xing (Mar. 2006). "Downside Risk". In: *The Review of Financial Studies* 19.4, pp. 1191–1239.

Armstrong, Christopher S, Snehal Banerjee, and Carlos Corona (2013). "Factor-loading uncertainty and expected returns". In: *The Review of Financial Studies* 26.1, pp. 158–207.

Asness, Clifford S, Tobias J Moskowitz, and Lasse Heje Pedersen (2013). "Value and momentum everywhere". In: *The Journal of Finance* 68.3, pp. 929–985.

Baltussen, Guido, Martin Martens, and Olaf Penninga (2021). "Factor Investing in Sovereign Bond Markets: Deep Sample Evidence". In: *The Journal of Portfolio Management*.

Baltussen, Guido, Laurens Swinkels, and Pim Van Vliet (2021). "Global factor premiums". In: *Journal of Financial Economics* 142.3, pp. 1128–1154.

Barigozzi, Matteo, Marc Hallin, and Stefano Soccorsi (2019). "Identification of global and local shocks in international financial markets via general dynamic factor models". In: *Journal of Financial Econometrics* 17.3, pp. 462–494.

Bates, Brandon J et al. (2013). "Consistent factor estimation in dynamic factor models with structural instability". In: *Journal of Econometrics* 177.2, pp. 289–304.

Bekaert, Geert and Roberto A De Santis (2021). "Risk and return in international corporate bond markets". In: *Journal of International Financial Markets, Institutions and Money* 72, p. 101338.

Bekaert, Geert, Campbell R Harvey, and Christian Lundblad (2007). "Liquidity and expected returns: Lessons from emerging markets". In: *The Review of Financial Studies* 20.6, pp. 1783–1831.

Bekaert, Geert, Robert J Hodrick, and Xiaoyan Zhang (2009). "International stock return comovements". In: *The Journal of Finance* 64.6, pp. 2591–2626.

Bekaert, Geert et al. (2014). "The global crisis and equity market contagion". In: *The Journal of Finance* 69.6, pp. 2597–2649.

Borghi, Riccardo et al. (2018). *The dynamics of factor loadings in the cross-section of returns*. CREATES Research Papers. Department of Economics and Business Economics, Aarhus University. URL: `https://EconPapers.repec.org/RePEc:aah:create:2018-38`.

Breitung, Jörg and Sandra Eickmeier (2014). *Analyzing business and financial cycles using multi-level factor models*. Available at SSRN. Bundesbank Discussion Paper No. 11/2014. URL: `https://dx.doi.org/10.2139/ssrn.2796962`.

Breitung, Jörg and Sandra Eickmeier (2015). "Analyzing business cycle asymmetries in a multi-level factor model". In: *Economics Letters* 127, pp. 31–34.

Breitung, Jörg and Jörn Tenhofen (2011). "GLS estimation of dynamic factor models". In: *Journal of the American Statistical Association* 106.495, pp. 1150–1166.

Brightman, Chris and Shane Shepherd (2016). *Systematic global macro*. Research Paper. Research Affiliates. URL: `https://www.researchaffiliates.com/publications/articles/563-systematic-global-macro`.

Brooks, Jordan and Tobias J Moskowitz (2017). *Yield curve premia*. Available at SSRN. URL: `https://dx.doi.org/10.2139/ssrn.2956411`.

Brooks, Jordan, Diogo Palhares, and Scott Richardson (2018). "Style investing in fixed income". In: *The Journal of Portfolio Management* 44.4, pp. 127–139.

Brusa, Jorge, Jenny Gu, and Grace Yaru Liu (2014). "The Time Decay of Bond Premium and Discount—An Analysis of the Time Passage Effect on Bond Prices". In: *Theoretical Economics Letters* 2014.

Bry, Gerhard and Charlotte Boschan (1971). "Front matter to" Cyclical Analysis of Time Series: Selected Procedures and Computer Programs"". In: *Cyclical analysis of time series: Selected procedures and computer programs*. NBER, pp. 13–2.

Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of finance* 52.1, pp. 57–82.

Chamberlain, Gary (1983). "Funds, factors, and diversification in arbitrage pricing models". In: *Econometrica* 51.5, pp. 1305–1323.

Connor, Gregory and Robert A Korajczyk (1986). "Performance measurement with the arbitrage pricing theory: A new framework for analysis". In: *Journal of Financial Economics* 15.3.

De Bondt, Werner FM and Richard Thaler (1985). "Does the stock market overreact?" In: *The Journal of Finance* 40.3, pp. 793–805.

Durham, J Benson (2015). "Can long-only investors use momentum to beat the US treasury market?" In: *Financial Analysts Journal* 71.5, pp. 57–74.

Fama, Eugene F and Kenneth R French (1996a). "Multifactor explanations of asset pricing anomalies". In: *The journal of finance* 51.1, pp. 55–84.

— (1996b). "The CAPM is wanted, dead or alive". In: *The Journal of Finance* 51.5, pp. 1947–1958.

— (1997). "Industry costs of equity". In: *Journal of Financial Economics* 43.2, pp. 153–193.

— (1998). "Value versus growth: The international evidence". In: *The Journal of Finance* 53.6, pp. 1975–1999.

— (2012). "Size, value, and momentum in international stock returns". In: *Journal of Financial Economics* 105.3, pp. 457–472.

— (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22.

— (2017). "International tests of a five-factor asset pricing model". In: *Journal of Financial Economics* 123.3, pp. 441–463.

Fama, Eugene F and James D MacBeth (1973). "Risk, return, and equilibrium: Empirical tests". In: *Journal of Political Economy* 81.3, pp. 607–636.

Ferson, Wayne E and Campbell R Harvey (1991). "The variation of economic risk premiums". In: *Journal of Political Economy* 99.2, pp. 385–415.

Frazzini, Andrea and Lasse Heje Pedersen (2014). "Betting against beta". In: *Journal of Financial Economics* 111.1, pp. 1–25.

Gagliardini, Patrick and Hao Ma (2019). *Extracting statistical factors when betas are time-varying*. Available at SSRN. Swiss Finance Institute Research Paper No. 19-65. URL: `https://dx.doi.org/10.2139/ssrn.3426534`.

Grinblatt, Mark and Tobias J Moskowitz (2004). "Predicting stock price movements from past returns: The role of consistency and tax-loss selling". In: *Journal of Financial Economics* 71.3, pp. 541–579.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (Feb. 2020). "Empirical Asset Pricing via Machine Learning". In: *The Review of Financial Studies* 33.5, pp. 2223–2273.

Guo, Bin et al. (2017). "The five-factor asset pricing model tests for the Chinese stock market". In: *Pacific-Basin Finance Journal* 43, pp. 84–106.

Harding, Don and Adrian Pagan (2002). "Dissecting the cycle: a methodological investigation". In: *Journal of Monetary Economics* 49.2, pp. 365–381.

Harvey, Campbell R (1991). "The world price of covariance risk". In: *The Journal of Finance* 46.1, pp. 111–157.

Hou, Kewei, Chen Xue, and Lu Zhang (Sept. 2014). "Digesting Anomalies: An Investment Approach". In: *The Review of Financial Studies* 28.3, pp. 650–705.

Inoue, Atsushi, Lu Jin, and Barbara Rossi (2017). "Rolling window selection for out-of-sample forecasting with time-varying parameters". In: *Journal of Econometrics* 196.1, pp. 55–67.

Jegadeesh, Narasimhan (1990). "Evidence of predictable behavior of security returns". In: *The Journal of Finance* 45.3, pp. 881–898.

Jegadeesh, Narasimhan and Sheridan Titman (1993). "Returns to buying winners and selling losers: Implications for stock market efficiency". In: *The Journal of Finance* 48.1, pp. 65–91.

Kelly, Bryan T, Tobias J Moskowitz, and Seth Pruitt (2021). "Understanding momentum and reversal". In: *Journal of Financial Economics* 140.3, pp. 726–743.

Kelly, Bryan T, Diogo Palhares, and Seth Pruitt (2021). *Modeling corporate bond returns*. Available at SSRN. URL: `https://dx.doi.org/10.2139/ssrn.3720789`.

Koijen, Ralph SJ et al. (2018). "Carry". In: *Journal of Financial Economics* 127.2, pp. 197–225.

Kubota, Keiichi and Hitoshi Takehara (2018). "Does the Fama and French five-factor model work well in Japan?" In: *International Review of Finance* 18.1, pp. 137–146.

Kunz, Brandon and Michele Mazzoleni (2018). *When value goes global*. Research Paper. Research Affiliates. URL: `https://www.researchaffiliates.com/publications/articles/655-when-value-goes-global`.

Lesmond, David A, Joseph P Ogden, and Charles A Trzcinka (1999). "A new estimate of transaction costs". In: *The Review of Financial Studies* 12.5, pp. 1113–1141.

Lettau, Martin and Markus Pelger (2020a). "Estimating latent asset-pricing factors". In: *Journal of Econometrics*.

Lettau, Martin and Markus Pelger (2020b). “Factors that fit the time series and cross-section of stock returns”. In: *The Review of Financial Studies* 33.5, pp. 2274–2325.

Lewellen, Jonathan and Stefan Nagel (2006). “The conditional CAPM does not explain asset-pricing anomalies”. In: *Journal of Financial Economics* 82.2, pp. 289–314.

Lo, Andrew W (2002). “The statistics of Sharpe ratios”. In: *Financial Analysts Journal* 58.4, pp. 36–52.

Mikkelsen, Jakob Guldbæk (2017). *Testing for time-varying loadings in dynamic factor models*. CREATES Research Papers. Department of Economics and Business Economics, Aarhus University. URL: `https://EconPapers.repec.org/RePEc:aah:create:2017-22`.

Mikkelsen, Jakob Guldbæk, Eric Hillebrand, and Giovanni Urga (2019). “Consistent estimation of time-varying loadings in high-dimensional factor models”. In: *Journal of Econometrics* 208.2, pp. 535–562.

Pesaran, M Hashem and Allan Timmermann (2007). “Selection of estimation window in the presence of breaks”. In: *Journal of Econometrics* 137.1, pp. 134–161.

Petkova, Ralitsa and Lu Zhang (2005). “Is value riskier than growth?” In: *Journal of Financial Economics* 78.1, pp. 187–202.

Robertson, Donald (2018). *Estimating β*. Technical Report. Ofgem. URL: `https://www.ofgem.gov.uk/system/files/docs/2018/12/ofgem_dr_dec_2018.pdf`.

Stock, James H and Mark W Watson (2002). “Forecasting using principal components from a large number of predictors”. In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179.

Wang, Peng (2008). *Large dimensional factor models with a multi-level factor structure: identification, estimation and inference*. Research Paper. Department of Economics, New York University.

Appendix A

Chapter 1

This appendix contains two sections. Section A.1 analyse the problem of reconciling the rolling out-of-sample forecasting methodology in Inoue, Jin, and Rossi, 2017 with the contemporaneous-equation framework of Borghi et al. (2018). Section A.2 studies the relationship between expected returns and beta parameters (such as variance, persistence, and magnitude), a topic which is discussed also in Armstrong, Banerjee, and Corona (2013) and Borghi et al. (2018).

A.1 Predictive Framework

In this appendix I analyse the problem of reconciling the rolling out-of-sample forecasting methodology in Inoue, Jin, and Rossi, 2017 with the contemporaneous-equation framework of Borghi et al. (2018). Assumption 4 of Inoue, Jin, and Rossi, 2017 provides the condition to approximate the parameter functions in (t/T), equation (1.11), via Taylor expansions up to the second order, which allows to model stock returns as an explicit function of the selected estimation window W. For simplicity, consider a predictive framework as in model (1.11) that features K factors, without disentangling global and regional risk drivers.

$$X_{i,t+h} = F_t \; \Lambda_i(t/T)^\top + e_{i,t+h}. \tag{A.1}$$

I can approximate the function $\Lambda_{i,W}(t/T)$, with W indicating the window size, by

$$\Lambda_{i,W}(t/T) = \lambda_i(1) + \lambda_i^{(1)}(1)\left(\frac{t-T}{T}\right) + \frac{\lambda_i^{(2)}(c)}{2!}\left(\frac{t-T}{T}\right)^2 \tag{A.2}$$

where $c = \phi\frac{t}{T} + (1-\phi)\frac{T}{T}$ for $\phi \in (0,1)$, $\lambda_i^{(j)}(.)$ denotes the jth derivative of $\Lambda_{i,W}(.)$. By substituting equation (A.2) into (A.1), I obtain

$$\begin{aligned} X_{i,t+h} &= F_t \; \lambda_i(1)^\top + F_t \; \lambda_i^{(1)}(1)^\top \left(\frac{t-T}{T}\right) + F_t \; \lambda_i^{(2)}(c)^\top \left(\frac{t-T}{T}\right)^2 + u_{i,t+h} \\ &= F_t \; \lambda_i(1)^\top + F_t \; \lambda_i^{(1)}(1)^\top \left(\frac{t-T}{T}\right) + \epsilon_{i,t+h} \end{aligned} \tag{A.3}$$

where $\epsilon_{i,t}$ is a composite error term made of $e_{i,t+h}$ and the second order terms in equation (A.2).

Inoue, Jin, and Rossi, 2017 suggest replacing the unknown functions in $\lambda_i(1)$ (together with their respective first-order derivatives) via OLS estimates using the most recent W_0 data, where W_0 is a given pilot window size for the local linear regression (from **timmerman**). Theorem 2 in Inoue, Jin, and Rossi, 2017 provides the conditions for the asymptotic optimality of their MSFE criterion, so that the error introduced by replacing $\lambda_i(1)$ with the sample counterpart $\hat{\lambda}_i(1)$ is negligible. This allows to quantify the bias induced by the rolling OLS estimator using the most recent W observations, which should be considered carefully. The rolling OLS estimator for the unknown functions $\lambda_i(1)$ and $\lambda_i(1)^{(1)}$ in (A.3) is given by

$$\begin{bmatrix} \hat{\lambda}_i(1) \\ \hat{\lambda}_i(1)^{(1)} \end{bmatrix} = \begin{bmatrix} \sum F_t^\top F_t & \sum F_t^\top F_t \left(\frac{t-T}{T}\right) \\ \sum F_t^\top F_t \left(\frac{t-T}{T}\right) & \sum F_t^\top F_t \left(\frac{t-T}{T}\right)^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum F_t^\top X_{i,t+h} \\ \sum F_t^\top X_{i,t+h} \left(\frac{t-T}{T}\right) \end{bmatrix} \tag{A.4}$$

with the summation $\sum$ going from $t = T - W_0 + 1$ up to $T - h$.

The MSFE criterion in Inoue, Jin, and Rossi, 2017 can also be adapted to contemporaneous (explanatory) regressions, and in this case I am selecting the optimal window size $\hat{W}$ based on the minimisation of the MSE at the end of the sample T, with $X_{i,T}$ describing the return of stock i with factors and loadings estimated up to T. When I consider a a contemporaneous regression framework, the population MSE at the end of the sample is defined as

$$E_T[(X_{i,T} - F_T\ \Lambda_i(1)^\top)^2] \tag{A.5}$$

with $E_T(.)$ being the conditional expectation based on the information set at T. The feasible MSE replaces $\Lambda_i(1)$ with the estimated parameter $\hat{\Lambda}_{i,W}(1)$ based on the last W observations in the sample. I then choose the window size W that minimises

$$\begin{aligned} E_T[(X_{i,T} - F_T\ \hat{\Lambda}_{i,W}(1)^\top)^2] &= E_T[(F_T\ \lambda_i(1)^\top + F_T\ \lambda_i^{(1)}(1)^\top \left(\frac{T-T}{T}\right) + e_{i,T} - F_T\ \hat{\Lambda}_{i,W}(1)^\top)^2] \\ &= E_T[(F_T\ \lambda_i(1)^\top + e_{i,T} - F_T\ \hat{\Lambda}_{i,W}(1)^\top)^2] \\ &= E_T[e_{i,T}^2] - 2E_T[F_T\ (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))^\top\ e_{i,T}] + ... \\ &\quad ... + E_T[F_T\ (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))^\top (\lambda_i(1) - \hat{\Lambda}_{i,W}(1)) F_T^\top]. \end{aligned} \tag{A.6}$$

Minimising the above expression is equivalent to minimising

$$\begin{aligned} E_T[(X_{i,T} - F_T\ \hat{\Lambda}_{i,W}(1)^\top)^2] &= E_T[F_T\ (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))^\top (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))\ F_T^\top] + ... \\ &\quad ... - 2E_T[F_T\ (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))^\top\ e_{i,T}] \end{aligned} \tag{A.7}$$

since the variance of the idiosyncratic error $E_T[e_{i,T}^2]$ is independent of W. However, equation (A.7), developed in a contemporaneous-equation framework, is not feasible because it depends on the unknown error $e_{i,T}$. In Inoue, Jin, and Rossi, 2017, the term $2E_T[F_T\ (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))^\top\ e_{i,T}]$ does not enter the equation due to the predictive feature of their model.

When I consider a predictive framework, with factors and loadings estimated up to time T, I aim to minimise the MSFE at the end of the sample, and equation (A.5) reads

$$E_T[(X_{i,T+h} - F_T\ \Lambda_i(1)^\top)^2]. \tag{A.8}$$

Due to the properties of conditional expectations, the true but unknown error of the predictive equation for the returns at horizon $T + h$, $e_{i,T+h}$, falls outside the information set at T and in this case the minimisation problem resolves to

$$E_T[(X_{i,T+h} - F_T\ \Lambda_i(1)^\top)^2] = E_T[F_T\ (\lambda_i(1) - \hat{\Lambda}_{i,W}(1))^\top(\lambda_i(1) - \hat{\Lambda}_{i,W}(1))\ F_T^\top]. \tag{A.9}$$

Inoue, Jin, and Rossi, 2017 suggests replacing the unknown $\lambda_i(1)$ with the local linear estimate $\hat{\lambda}_i(1)$ computed on a pilot window that considers the most recent W_0 observations. In their framework, the minimisation problem does not depend on the unknown parameter $\lambda_i(1)$, replaced by the sample counterpart $\hat{\lambda}_i(1)$, and on the unknown error term $e_{i,T+h}$, which does not enter the equation due to the property of conditional expectations. The optimal window size W thus minimises

$$\hat{W} = \min\{E_T[F_T\ (\hat{\lambda}_i(1) - \hat{\Lambda}_{i,W}(1))^\top(\hat{\lambda}_i(1) - \hat{\Lambda}_{i,W}(1))F_T^\top]\}. \tag{A.10}$$

A.2 Relationship with Expected Returns

In this appendix I test the claim of Armstrong, Banerjee, and Corona (2013) that firm-specific uncertainty of factor sensitivities negatively affects expected returns. They employ a CAPM model for US stocks with time-varying factor sensitivities that are estimated via rolling OLS regressions, and Borghi et al. (2018) extend their analysis internationally to global and regional factors, with time-varying loadings estimated via MLE. The results of Borghi et al. (2018) indicate a premium for holding stocks with highly volatile exposure to global systematic risk (financial and global factors), implying an additional source of priced risk in the cross-section of returns. Their results are at odds with Armstrong, Banerjee, and Corona (2013) who report a negative relationship between variance of the loadings and expected returns. To expand on their work, I repeat the analysis of Borghi et al. (2018) considering an extended sample from January 2006 to May 2019, and I fail to confirm their evidence in an 'out-of-sample' context. In fact, I do not find evidence of a monotonic relationship between expected returns and beta parameters (magnitude, variance and persistence), considering all three factors, My results point to an irrelevance of the beta parameters in the cross-section of stock returns, in contrast to both Borghi et al. (2018) and Armstrong, Banerjee, and Corona (2013).

To document my findings I compute the T-average arithmetic return over the full sample for each stock (cum-dividend data, no standardised), obtaining a vector of N (weekly) expected returns. I then estimate the empirical quantiles for three beta parameters (variance, persistence, magnitude) and plot the median within-quantile values of the variable in question (expected returns) against the quantile number, I do so for each beta parameter. Financial stocks are excluded, together with the equities with non-identifiable time-varying betas. Figure A.1 reports the results.

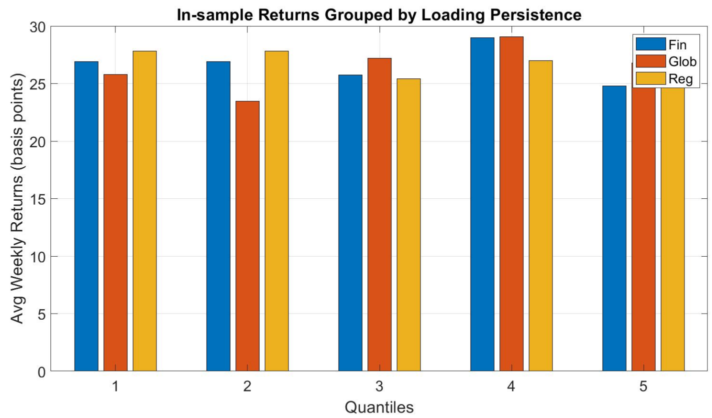
[Figure A.1 about here.]

FIGURE A.1: **Relationship with Expected Returns**

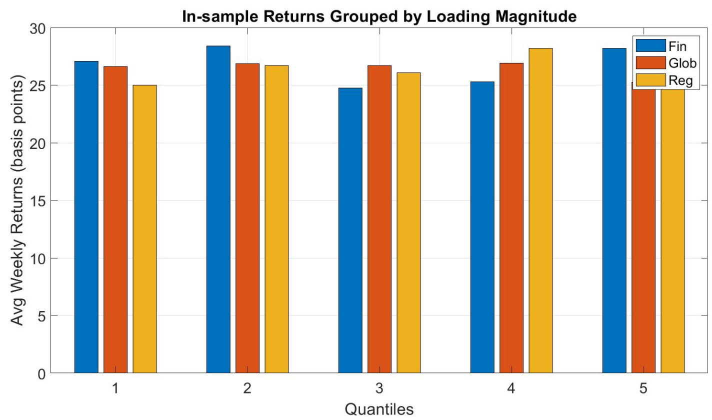
The figure reports the average weekly arithmetic returns from January 2006 to May 2019 expressed in basis points, ordered by factor loadings variance (Panel A.1a), persistence (Panel A.1b) and magnitude (Panel A.1c). At the end of the sample, stocks are sorted in quantiles of either loading variance, persistence, or magnitude. Quantile five contains the largest value. Then, for each quantile I calculate the median return, and I plot it against the quantile number. Financial stocks are excluded, as well as those stocks with loadings that vary so little that I consider them constant.

(A) *Variance*

(B) *Persistence*



(C) *Magnitude*



Appendix B

Chapter 2

This appendix contains three sections. Section B.1 reports an extension of the benchmark model described in Section 2.2 that allows to analyse the contribution of time-varying betas with respect to the estimation of co-movements. Section B.2 reports further details on the financial events that I consider for the economic interpretation of the time-varying factor betas in Section 2.5.2. Finally, Section B.3 provides further details on the data cleaning procedure, and on the regional composition of the FF factors.

B.1 Model-Implied Covariance Structure

In this appendix I report an extension of the benchmark model described in Section 2.2 that allows me to analyse the contribution of time-varying betas in shaping the co-movements structure implied by the factor model in (2.2). Similarly to the case of expected returns, I assume that factors have a static covariance structure that is estimated considering the information on the full sample T, which allows to study in isolation the role of time-varying betas in explaining stock return co-movements.

Under the assumption of weak exogeneity between factors and error terms[1], the covariance matrix of stock returns can be decomposed into a systematic and idiosyncratic component. Let $\boldsymbol{r}_{1:T}$ denote the $(T \times N)$ matrix of excess asset returns with observations up to period T, then

[1]Standard factor models assumption include:

- Stationarity of factor realisations $\boldsymbol{f}_{1:T}$ with unconditional moments given by

$$E[\boldsymbol{f}_{1:T}] = \boldsymbol{\lambda}$$
$$cov(\boldsymbol{f}_{1:T}) = E[\boldsymbol{f}_{1:T}^\top \boldsymbol{f}_{1:T}] - \boldsymbol{\lambda}\, \boldsymbol{\lambda}^\top.$$

- Weak exogeneity between the factors and error terms

$$cov(f_{k,t},\ \epsilon_{i,t}) = 0, \quad \text{for all } i, k, \text{and } t.$$

- Error terms are serially uncorrelated and contemporaneously uncorrelated across assets

$$E[\boldsymbol{\epsilon}_{1:T}^\top\, \boldsymbol{\epsilon}_{1:T} \mid \boldsymbol{f}_{1:T}] = diag(\sigma_1^2, ..., \sigma_N^2).$$

the associated covariance decomposition of model (2.2) reads

$$cov_T(\boldsymbol{r}_{1:T+h}) = \boldsymbol{\beta}_T \; cov(\boldsymbol{f}_{1:T}) \; \boldsymbol{\beta}_T^\top + cov(\boldsymbol{\epsilon}_{1:T+h}) \tag{B.1}$$

where $cov_T(\boldsymbol{r}_{1:T+h})$ is the $(N \times N)$ variance-covariance matrix of future-period excess returns conditional on the T-period factor loadings $\boldsymbol{\beta}_T$, $cov(\boldsymbol{f}_{1:T}) = E[\boldsymbol{f}_{1:T}^\top \boldsymbol{f}_{1:T}] - \boldsymbol{\lambda} \; \boldsymbol{\lambda}^\top$ is the $(K \times K)$ unconditional variance-covariance matrix of factor returns, and $cov(\boldsymbol{\epsilon}_{1:T+h}) = diag(\sigma_i^2, ..., \sigma_N^2)$ is the $(N \times N)$ diagonal matrix of idiosyncratic error covariances.

I can then evaluate the out-of-sample model's ability to predict future (expected) covariances, given the conditional beta estimates in the sub-sample $t^* = 1, ..., T_W$ (each of length W), as

$$cov_{t^*}(\boldsymbol{r}_{t*+h}) = \boldsymbol{\beta}_{t^*} \; cov(\boldsymbol{f}_{1:T}) \; \boldsymbol{\beta}_{t^*}^\top + cov(\boldsymbol{\epsilon}_{t^*+h}), \quad t^* = 1, ..., T_W \tag{B.2}$$

where $cov_{t^*}(\boldsymbol{r}_{t*+h})$ is the $(N \times N)$ variance-covariance matrix of future excess returns considering information in period $t*$, and $cov(\boldsymbol{\epsilon}_{t^*+h})$ is the $(K \times K)$ unconditional covariance matrix of factor realisations. In a similar spirit as in Kelly, Palhares, and Pruitt (2021), I can evaluate the performance of the time-varying beta estimator in predicting future covariances by considering a variety of error measures. For instance, equation (2.17) can be adapted to

$$\text{OOS error}_{t^*} = \hat{\boldsymbol{S}}_{t^*+h} - (\hat{\boldsymbol{\beta}}_{t^*} \; \widehat{cov(\boldsymbol{f}_{1:T})} \; \hat{\boldsymbol{\beta}}_{t^*}^\top), \quad t^* = 1, ..., T_W \tag{B.3}$$

where $\hat{\boldsymbol{S}}_{t^*+h}$ can be the sample covariance matrix of stock returns estimated in sub-sample t^* that also includes the h-period ahead observations, and $\widehat{cov(\boldsymbol{f}_{1:T})}$ the estimated factor covariance matrix on the full sample. Note that this approach is similar to Bekaert, Hodrick, and Zhang (2009) who use a mean squared error criterion defined as the time-series mean of a weighted average of squared errors (from individual stock covariances). I leave this extension for future studies.

B.2 Economic Calendar

In this appendix I report further details on the financial events that I consider for the economic interpretation of the time-varying factor betas in Section 2.5.2. I differentiate between events that have a world-wide impact, and events that are relevant to the equity markets of specific world regions.

The former set includes two periods of declining economic activity across the major world economies.

1. **Global Financial Crisis (GFC).** I follow the the NBER Business Cycle Dating Committee

[2] that identifies the dates of peaks and troughs that define the economic recessions and expansions in the me economy. The emphasis on the NBER's definition of recession is on the severity of the decline in economic activity that is spread across the whole economy, and on the duration of this contraction. The determination of the months of peaks and troughs is based on a range of monthly measures of aggregate real economic activity published by the federal statistical agencies, the peak month of the GFC is estimated to be on December 2007 and the through on June 2009. A recession is the period between a peak of economic activity and its subsequent trough, or lowest point. On the other hand, the economy is in an expansion between trough and peak dates. According to the NBER chronology, after the through in June 2009 the US economy enjoyed an expansionary phase up until the most recent peak, which occurred in February 2020. The most recent trough occurred in April 2020. The analysis in this chapter is based on a 13-year sample from January 2006 to May 2019, and as such it includes only one recessionary phase, the GFC, as per NBER's definition. In Section 2.6 I highlight that further research based on an extended time frame is needed to assess the impact of the COVID-19 outbreak on the dynamics of factor sensitivities. Based on the ample evidence of spill-over effects from the US economy to other world regions during the GFC, I assume that this event has an impact on the equity prices of the firms listed in all six regions considered, thus I highlight the peak and through dates in most of the time-series charts in this chapter to ease the economic interpretation of my estimates.

2. **European Sovereign Debt Crisis (ESDC).** I consider the peak and through dates of economic recessions in the EU-wide area published by The Conference Board [3] which builds on the NBER's convention for the US economy. The Conference Board uses a business cycle dating algorithm based on Bry and Boschan (1971), and Harding and Pagan (2002) and analyses the turning points in the economic activity of four major European economies (France, Germany, Spain and the UK) as well as for the Euro Area. For countries like Spain, the estimates indicate a peak date on June 2010, while for the Euro Area the peak date is shifted almost one year forward to July 2011. In fact, in 2010 the level of Spain public debt relative to GDP was only 60%, more than 20% less than Germany, France or the US, and more than 60% less than Italy or Greece, effectively being one of the lowest among advanced economies prior to the crisis. The thorough for the Euro Area is in line with the through dates of France and Spain, while it is assumed that no recession occurred in Germany during the same period, and only partly in the UK (from August 2010 to December 2011). The ESDC originated in Europe and the detailed causes of the crisis vary from country to country. Throughout this chapter I assume that the crisis has a major impact on the equities listed in the Western and Emerging Europe, as well as those in the North America regions. When I present the aggregate results for these regions, I highlight the peak and through dates of the ESDC in the time-series charts.

[2]Source: NBER Busines Cycle Dating.
[3]Source: The Conference Board Business Cycle Indicators.

During the 13-year period considered, I also analyse the financial events that I believe have a significant impact on the equity markets of specific world regions[4], but do not necessarily spread to the whole economy, or to multiple countries/regions. Some of the events have a very concise time span, like the United States presidential elections, while others span multiple years and included sub-periods of significant market turmoil, like the 2014-2016 oil crash, or the 2015-2016 Chinese stock market crash. The dates that I consider are based on a variety of sources including major news outlets as well as Bloomberg.

1. **2012 US Presidential Election.** On November 6 2012 President Obama is officially re-elected with 51% of the popular vote. Early voting began at the end of September 2012 in 12 states, followed by October which featured three presidential debates roughly one week apart from each other. Given the forward-looking nature of financial markets, I analyse the performance of factor sensitivities starting from September 2012 and ending on November 2012 (end-of-month). I follow the same approach for the subsequent presidential election in 2018, and decide not to include the 2008 elections given the peculiar regime in which the economy was at that time. In the 10 subsequent days following election day, the $S\&P500$ dropped -2.8% overall, with a peak of -5% (daily). The region affected by the presidential elections is North America (which includes Canada, the US, and Mexico).

2. **2014-2016 Oil Crash.** From September 2014 to January 2016 the Brent price of crude oil experienced a 53% decline that rivalled only the 73% drop occurred during the GCF, from July 2008 to the end of the year. Up until 2014, oil prices were primarily driven by the increase in demand of China and other emerging economies, following years of significant sustained growth. Adding on the demand side, by that time also hydraulic fracturing in the US and oil production in Canada rose, which ultimately contributed to an increase in global oil production. In the middle of 2014 however, price started declining due to increase supply, and simultaneously slowing demand in the emerging countries. During 2014-2015, OPEC members consistently exceeded their production ceiling, at the same time US oil production increased further which led to a plunge in US oil import requirements and a high volume of oil inventories in storage across the world, causing the oil price to collapse. The oil glut, spurred a sharp downward spiral in the price of the commodity that reached its peak at the end of January 2016, when oil was below the 30$ mark. During this period, oil and gas companies faced high risks of bankruptcy worlwide due to low output costs. Following my regional classification, I assume that the oil crash has a significant impact on emerging markets in Europe, and in the Middle East & Africa regions.

3. **2015-2016 Chinese Stock Market Crash.** From the beginning of June 2015 to early February 2016, the Chinese stock market experienced a major decline which led to a third of

[4]Throughout this chapter, the emphasis is on the regional classification of stocks based on the company's domicile. Although I acknowledge that certain events may be also confined to specific sectors of the economy, I present most of the cross-sectional results region-wise.

the value of local-currency-denominated shares on the Shanghai Stock Exchange being erased within one month of the event. Major aftershocks occurred around July 27th (-7%) and August 24th 'Black Monday' (-8.9%). By the beginning of July 2015, the stock market fell more than 30% over the three preceding weeks, with almost $1,500$ companies (more than half of those listed) filing for trading halts in an attempt to prevent further losses. The Chinese government intervened by imposing a series of bans that prevented investors owning more than 5% of a company's stock to short-sell the shares. Given the peculiar composition of the Chinese stock market, the Government also provided cash to brokers to buy shares, backed by central-bank cash, in an attempt to boost demand for the securities. For this event, I assume that only the Asia-Pacific region is affected.

4. **2016 US Presidential Election**. On November 8 2016 the American citizen were asked to vote for the new President after President Obama's second mandate. During the preceding month, candidate Hilary Clinton for the Democratics and Donald Trump for the Repulicans engaged in a series of presidential debates which received substantial media coverage. By 2.45 m Eastern time on election day, Donald Trump appears to be the projected winner of election, effectively becoming President-elect. In the 10 subsequent days following election day, the S&P500 entered a steady bull phase topping $+2.96\%$ overall, with modest peaks ($+1\%$) at the end of each day trading session. Similarly for the 2012 presidential election, I study the performance of factor sensitivities of North American stocks from September 2016 up to the end of November 2016.

5. **2016 Brexit Referendum.** On June 23rd 2016 the UK asked the electorate whether the country should remain a member of, or leave, the European Union. The referendum resulted in a slim victory for the 'leave' side (51.9% of the votes cast), and although the referendum was legally non-binding, the government of the time promised to implement the measures. Share prices of the five largest British banks fell an average of 21% in the morning session after the referendum. All of the major credit rating agencies reacted negatively to the vote. Standard & Poor's cut the British credit rating from AAA to AA, Fitch Group cut from AA+ to AA, and Moody's cut the UK's outlook to negative. On the morning of June 24th, the pound sterling fell to its lowest level against the US dollar since 1985. The drop over the day was 8%, making it the biggest one-day fall in the currency since the introduction of floating exchange rates at the end of the Bretton Woods in 1971. The region affected by the referendum is Western Europe, which includes the major EU economies as well as the UK, and I analyse the dynamics of factor sensitivities from the beginning of May to the end of July 2016.

6. **2018-2019 Trade War.** During the course of 2018 under the Presidency of Donald Trump, the US and China engaged in a series of import restrictions and increased tariffs which shaped the so-called 'US-China Trade War'. During the first half of 2018, the increase in tariff was moderate. In the months of July thorough September 2018 both sides increased tariffs: US average tariffs increased from 3.8% to 12%, and China's ones from 7.2% to

18.3%[5]. During the same period the S&P500 increased by 7.2 percentage points. After an 8-month period (September 2018 to June 2019) of little change in tariffs, June to September 2019 saw another set of trade tariffs increases being implemented. The dataset for my analysis ends on May 2019, and I am not able to analyse the effects of stage five of the trade deal on the dynamics of factor loadings. I leave this for future research, see Section 2.6. In the same months that saw the trade war escalating, the US stock market experienced periods of sustained market volatility due to geopolitical tensions with North Korea, as well as uncertainty on the monetary policy, and I acknowledge that it can be difficult to accurately isolate the impact of single events on the factor sensitivities for this period. I focus on the months of July through October 2018 (included) which saw the steepest increase in trade tariffs during the trade escalations. The regions affected by these shocks are North America and Asia-Pacific.

Table B.1 reports a summary of the relevant dates for the events listed above.

[Table B.1 about here.]

B.3 Data Cleaning

In this appendix I provide further details on the data cleaning procedure, as well as on the regional composition of the local FF factors.

Data Cleaning. In table 2.2 I report the summary statistics on the country-region composition of my universe, and based on the figures in column #Full I end up excluding from my analysis some of the countries due to data quality issues. Before estimation, I apply a conservative screen to delete the tickers with no more than 12 consecutive missing observations and at least one year of data, and I also delete the dually-listed tickers. To facilitate the comparison of my results across window sizes, I fix the cross-sectional dimension to $N = 1686$, which comprise the tickers that remain listed in the national equity indeces throughout the entire sample. For some of the countries there are no equities that remain listed for the entirety of the $T = 700$ weeks, and as such are excluded from the analysis. These countries are India, Korea, Russia, and Ukraine. Weekends are non-trading days as per the NYSE calendar are excluded.

FF Factors. I download the FF3[6] and FF5[7] factor returns at daily frequency from Kenneth French's web site at Dartmouth. I firstly synchronise the daily series to my weekly (end-of-Friday) benchmark, and similarly to the other observed factors I winsorise the data at 99% level. The factors are region specific and B.2 reports the details on the regional classification employed by FF. They analyse the returns for the companies listed in 23 countries and partition their

[5]Source: Peterson Institute for International Economics.
[6]Source: Kenneth R. French Data Library.
[7]Source: Kenneth R. French Data Library.

universe into four regions: Europe, which closely matches my classification for the Western Europe region, North America, which includes only the me and Canada, and Asia Pacific ex Japan, which includes Australia, Hong Kong, New Zealand and Singapore. In table B.2 I denote with ✓the countries in FF that are also included in my universe. Only three countries are analysed in FF but not in my study, Greece, Italy, and Singapore. On the other hand, I include Mexico in the North America region, as well as Thailand, Indonesia and China in the Asia-Pacific region. Japan is treated as a separate region in FF while I include it in Asia-Pacific. Overall the partition in FF closely matches the one in Bekaert et al. (2014) which is the reference in my study.

[Table B.2 about here.]

TABLE B.1: **Economic Calendar**

The table reports the relevant dates of the financial events that I consider for the economic interpretation of the time-varying factor betas in Section 2.5.2. Panel B.1a shows the peak and through dates of the two major crises that I believe have a wide-spread impact on the major global equity markets, while panel B.1b reports the region-specific events. NA stands for North America, LA for Latin America, AP for Asia-Pacific, WE for Western Europe, EE for Eastern Europe, and MEA for Middle East & Africa.

(A) *Global Crises*

Event	Peak	Through	Regions Affected
Great Financial Crisis	Dec 2007	June 2009	All
European Sovereign Debt Crisis	July 2011	Feb 2013	WE, EE, NA

(B) *Region-Specific Events*

Event	Start	End	Regions Affected
2012 me Presidential Election	Sep 2012	Nov 2012	NA
Oil Crash	Sep 2014	Jan 2016	LA, EE, MEA
Chinese Stock Market Crash	June 2015	Feb 2016	AP
2016 me Presidential Election	Sep 2016	Nov 2016	NA
2016 Brexit Referendum	May 2016	July 2016	WE, EE
2018-2019 Trade War	July 2018	Oct 2018	NA, AP

TABLE B.2: **Fama-French Regional Classification**

The table reports details on the regional classification employed by FF for the construction of their factors. Column *FF* indicates if the country is also considered part of the corresponding FF regions (North America, Europe, and Asia Pacific ex Japan).

Index	Country	Region	FF	#Full
SPTSX60	Canada	North America	✓	64
OEX	US	North America	✓	129
MEXBOL	Mexico	North America	x	38
TPXL70	Japan	Asia-Pacific	x	104
SSE50	China	Asia-Pacific	x	90
HSCEI	HongKong	Asia-Pacific	✓	56
LQ45	Indonesia	Asia-Pacific	x	70
SET50	Thailand	Asia-Pacific	x	64
NZSE50FG	NewZealand	Asia-Pacific	✓	45
AS31	Australia	Asia-Pacific	✓	55
ATX	Austria	Western Europe	✓	23
BEL20	Belgium	Western Europe	✓	26
KFX	Denmark	Western Europe	✓	26
HEX25	Finland	Western Europe	✓	26
CAC	France	Western Europe	✓	48
DAX	Germany	Western Europe	✓	37
ISEQ	Ireland	Western Europe	✓	22
AEX	Netherlands	Western Europe	✓	29
OBX	Norway	Western Europe	✓	28
PSI20	Portugal	Western Europe	✓	21
IBEX	Spain	Western Europe	✓	33
OMX	Sweden	Western Europe	✓	36
SMI	Switzerland	Western Europe	✓	43
UKX	UK	Western Europe	✓	119
Total				1232

Appendix C

Chapter 3

This appendix contains two sections. Section C.1 reports further details on CUSIP-level data and describes the data cleaning procedure. Section C.2 propose an adaptation of the carry factor construction procedure of Koijen et al. (2018) to my framework.

C.1 Data Cleaning

The sample period of our study runs from the beginning of January 2010 to the end of October 2021, and most of the data is available at the end of each trading day in the respective national business calendars. To syncrhonise the data across countries, I take the perspective of a US investor and select the observations that match the dates of the official New York Stock Exchange calendar[1]. In the United States, the number of trading days in a year averages 260. This procedure inevitably discards observations in countries other than the US due to non-overlapping holidays, however it is necessary to produce a data set with no look-ahead bias at the end of each day across markets[2].

Once the data is syncrhonised, I apply the following screens that help identify entry errors or bad data in general. This approach is inspired by Baltussen, Swinkels, and Van Vliet (2021) who apply a series of data quality checks to produce a high-quality historical dataset for four asset classes (equities, bonds, commodities, and FX) spanning more than 100 years. I build upon their framework based on end-of-month observations and modify it to mydaily frequency. The adjustments that I perform are the following (in order).

1. I drop the observations recorded as integers (including negative values). Data on the country-specific deposit rates is available with 4 decimal places (in %), swap rates, generic yields are recorded in % with 3 decimals, and real yields with 2 decimals (also in %). Bond prices (clean and dirty) are recorded with 3 decimal places and are expressed with respect to a par price of 100 (in local currency), and bond yields are recorded with 3 decimal places (in %).

2. I drop consecutive repeated (identical) observations.

[1]Source : NYSE Calendar.

[2]I assume that the effects of overlapping trading hours across markets within a given day are negligible.

3. Finally, I drop the observations that belong to a month in which less than 75% of data is available. This filter is inspired by the 'zero return screen' in Baltussen, Swinkels, and Van Vliet (2021) which leaves out observations with more than one zero or missing spot in the past 12 months. In their work, the threshold for the portion of missing data within a year (monthly observations) is set to 8.3% (1/12). In my case (daily observations), I allow a maximum of 5 non-consecutive missing observations within a month, which corresponds to about 23% of missing data in a business year (60/260). This adjustment is particularly relevant for the bonds' daily price series since it can be taught of as a screen for reduced liquidity. The idea of using the incidence of zero returns (same consecutive price) in the sample as a proxy of liquidity dates back to Lesmond, Ogden, and Trzcinka (1999), who develop a model for transaction costs based solely on equity price data, and was leveraged by Bekaert, Harvey, and Lundblad (2007) on emerging markets data.

The screens introduce missing data points in the series, and for each country I select the data that has no more than *maxdays* consecutive missing observations (in days), and no more than *miss* missing points in the sample (in %). These figures differ for each data type (swap rates, bond prices etc.) and are reported below, table C.1. For asset-level data I consider the lifespan of each bond to calculate the statistics on the data quality (and not the entire sample as for market data). Once the data series are selected, I replace the missing observations with a 5-day moving average, for a maximum of *maxdays* consecutive missing days.

[Table C.1 about here.]

Table C.2 reports a summary of the available data for each country after the cleaning procedure. I differentiate between market- and bond-level data. For the former, I report a blue tick under the column IRS if there is sufficient data on the time-series of swap rates (for the different maturities), and on the underlying (float leg) deposit rate. A black tick indicates unavailable data on the time-series of deposit rates, but sufficient data on the swap rates. Similarly for the other data types I indicate with a tick if the data is available after the screens, and with a cross otherwise. In the bond panel, I count the number of bonds that have complete time-series of clean and dirty prices (P), the corresponding yields to maturity (y), dollar duration (DV01), and z-spread (Z). I adopt a bottom-down approach and out of a total of 20 countries I select nine that have a sufficiently large cross-section of bonds with available data, and complete market data. The countries include: Australia, Canada, Japan, the UK, the US, and France, Germany, Italy, and Spain for the Euro Area.

[Table C.2 about here.]

I now I report further details on bond-level data. For my universe I only consider bonds that were issued by the national governments from January 1st 2010 up to the end of October 2021 and that match certain characteristics. In particular, I exclude inflation-linked bonds, green bonds, international bonds (those issued in a foreign currency), retail, exchange-traded, when-issued, sinkable, and funged bonds, as well as certificates of deposits. I only include bonds with standard characteristics (i.e. option-free, non bullet bonds) with a maturity of 2 to 30 years.

For each bond I gather from Bloomberg the following data series (field in brackets):

- Clean price (PX_LAST): the last price received from the pricing source (BGN). For all bonds this coincides with the clean price, expressed per 100 face value (not in currency terms).
- Dirty price (PX_DIRTY_MID): the mid price of the bond that includes the accrued interest that the seller is entitled to receive, dirty price $\geq$ clean price.
- Yield conventional (clean) (YLD_CNV_LAST): conventional yield computed from the clean price (in %).
- Yield annual (dirty) (YLD_ANNUAL_MID): using the dirty price, the yield of the bond (in %), Yield annual $\geq$ yield conventional.
- Dollar duration (RISK_MID): the dollar value of a basis point change in mid yield times 100.
- Z-spread (Z_SPRD_MID): the spread (in bp) that must be added to the spot curve so that the bond's discounted cashflows equals its mid price (dirty), with each dated cashflow discounted at its own rate.

The total number of securities with complete series of end-of-day clean prices is 972, 788 if I consider the series of dirty prices, 717 and 706 for yields on the clean and dirty prices respectively, 540 bonds with historical dollar duration series, and 288 bonds with complete Z-spread daily values.

C.2 Carry Factor

In regards to carry, given the characteristics of my data, i.e. CUSIP-level at daily frequency, adapting the measures of Koijen et al. (2018) requires careful consideration of duration risk. Koijen et al. (2018) examine carry for US Treasuries in the cross section from one to ten years of maturity, and adjust their position sizing to account for the different volatility (and duration) of the short and long legs. Consider for instance a (term-spread) portfolio that invests long 1$ of 10-year bonds and shorts 1$ of 1-year bonds. The 10-year bonds are far more volatile than the 1-year, and they adjust their positions by multiplying the synthetic futures price F_t^τ by the duration D_t^τ, $X_t^\tau = F_t^\tau D_t^\tau$. This implies that a riskier bond with a larger duration is supported by a larger amount of capital and its return and carry are scaled down accordingly. The key equations from Koijen et al. (2018) are

$$R_t^{\tau} = \frac{F_t^{\tau} - F_{t-1}^{\tau}}{X_{t-1}^{\tau}} \tag{C.1}$$

$$C_t^{\tau}(X = F_t^{\tau} D_t^{\tau}) = \frac{C_t^{\tau}(X = F_t^{\tau})}{D_t^{\tau}} \tag{C.2}$$

$$C_t^{\tau}(X = F_t^{\tau}) = \frac{(1 + y_t^{\tau})^{\tau}}{(1 + r_t^f)(1 + y_t^{\tau-1})^{\tau-1}} - 1 \tag{C.3}$$

where $C_t^{\tau}(X = F_t^{\tau} D_t^{\tau})$ is the time-t carry on a bond with τ periods to maturity, which is a function of the bond's duration D_t^{τ} and the carry of a fully collateralised position $C_t^{\tau}(X = F_t^{\tau})$, y_t^{τ} is the zero-coupon bond yield. The return and carry of the slope-of-the-yield portfolio in my example are

$$C_t^{slope} = C_t^{10Y}(X = F_t^{10Y} D_t^{10Y}) - C_t^{2Y}(X = F_t^{2Y} D_t^{2Y}) \tag{C.4}$$

$$R_t^{slope} = R_t^{10Y} - R_t^{2Y}. \tag{C.5}$$

The approach of Koijen et al. (2018) of calculating carry for constant-maturity securities differs from my setup under many aspects: they use synthetic futures as base assets, instead of traded bond data, they bootstrap the yields y_t^{τ} for all (continuous) maturities of a single issuer's curve, and adjust their allocations by duration, rather than time-to-maturity. An extension to my study is to compare the performance of portfolios formed on bond-level data against the constant-maturity benchmark yields that I plot in figure 3.2. Using the latter, I can price the synthetic bond portfolios for all available maturities (together with the bootstrapped yields), calculate the carry of a fully collateralised position using equation (C.3), and finally use equations (C.4) and (C.5) to calculate carry and returns. I can then compare its performance against a (local or global) portfolio formed on bond-level data that match a desired duration target D^* (ten and two years in the example above)

$$w_t^*(D*) = min\{D_t^{*\ (c,b)} - D^*\} \tag{C.6}$$

$$D_t^{*\ (c)} = \sum_i^{N_{t,c}} w_{i,t}^* \ I(w_{i,t}^* > 0) \ D_{i,t} + \sum_i^{N_{t,c}} w_{i,t}^* \ I(w_{i,t}^* < 0) \ D_{i,t}. \tag{C.7}$$

The resulting portfolio returns (for a fixed target D^*) can be calculated using equation (3.3), and the equivalent slope-of-the-yield portfolio is thus

$$R_t^{*slope} = f_t^{(c)}(D^* = 10Y) - f_t^{(c)}(D^* = 2Y). \tag{C.8}$$

Effectively, one can think of the constant-maturity portfolios $f_t^{(c)}(D^*)$ as primitive assets in this approach, which allows reconciliation with standard practices from the literature, eg. Durham (2015), Brightman and Shepherd (2016), Brooks and Moskowitz (2017) and Brooks, Palhares, and Richardson (2018).

TABLE C.1: **Data Cleaning Parameters**

The table reports the parameters that I use to select the data series after the screens are applied. Column *maxdays* refers to the number of maximum consecutive missing observations (in days), and *miss* to the percentage of missing data in the sample period.

Data	*maxdays* (days)	*miss* (%)
Bond Data (prices, yields etc.)	10	10
IRS	10	50
Deposit Rates	800	70
Generic Yields (nominal, real)	800	70

TABLE C.2: **Data Availability**

The table reports a summary of the available data for each country after the cleaning procedure. In the *Market Data* panel, I report a blue tick under the column *IRS* if there is sufficient data on the time-series of swap rates, and on the underlying deposit rates. A black tick indicates unavailable data on the time-series of deposit rates, but sufficient data on the swap rates. For the other columns a tick indicates that the data is available after the screens, and a cross otherwise. In the *Bond Data* panel, I count the number of bonds that have complete time-series of clean and dirty prices (P), the corresponding yields to maturity (y), dollar duration, *DV01*, and z-spread, *Z*.

		Market Data			Bond Data					
Issuer	Included	IRS	Nominal	Real	Clean (P)	Dirty (P)	Clean (y)	Dirty (y)	DV01	Z
Australia	✓	✓	✓	✓	32	28	25	25	18	0
Canada	✓	✓	✓	✓	62	54	52	52	36	36
Denmark	x	x	✓	✓	8	6	6	6	5	5
Japan	✓	x	✓	x	184	149	94	87	61	61
New Zealand	x	✓	✓	✓	13	12	11	11	11	0
Norway	x	x	✓	✓	8	8	7	7	6	6
South Korea	x	x	✓	✓	21	21	16	14	0	0
Sweden	x	✓	✓	✓	8	8	8	8	5	5
Switzerland	x	x	✓	✓	13	12	12	12	7	7
United Kingdom	✓	✓	✓	✓	35	35	35	35	27	27
United States	✓	✓	✓	✓	385	275	272	268	231	0
Euro Area		✓								
Austria	x		✓	✓	22	21	19	19	14	14
Belgium	x		✓	✓	24	24	23	23	15	15
France	✓		✓	✓	45	39	38	38	27	27
Germany	✓		✓	✓	72	68	63	63	45	42
Greece	x		✓	✓	16	16	15	15	14	14
Italy	✓		✓	✓	98	86	84	84	61	61
Netherlands	x		✓	✓	20	19	18	18	11	11
Portugal	x		✓	✓	15	15	15	15	11	11
Spain	✓		✓	✓	59	54	54	54	34	34
All	20	7	20	19	1140	950	867	854	639	376
Included	9	5	9	8	972	788	717	706	540	288