UNIVERSITÉ PARIS OUEST - NANTERRE LA DEFENSE

ECOLE DOCTORALE ECONOMIE, ORGANISATIONS, SOCIETE

N° attribué par la bibliothèque

Année 2009 |_|_|_|_|_|_|_|_|_|_|

T H È S E

pour obtenir le grade de

Docteur en sciences économiques

de l'Universite de Paris Ouest - Nanterre La Defense

Discipline : Sciences Économiques

présentée et soutenue publiquement par

Juliana CAICEDO-LLANO

le 11 décembre 2009

Marchés Actions de pays Emergents: analyse de la dynamique des rendements et stratégies d'investissement

---

Directeur de thèse :

Mme. Catherine BRUNEAU, Professeur à l'Université de Paris Ouest - Nanterre La Défense

---

JURY

M.	Patrice FONTAINE,	Professeur à l'Université de Grenoble
Mme.	Carole GRESSE,	Professeur à l'Université Paris - Dauphine
Mme.	Sandrine LARDIC,	Professeur à l'Université de Paris Ouest - Nanterre La Défense
M.	Duc K. NGUYEN,	Professeur de Finance à l'ISC Paris School of Management
Mme.	Catherine BRUNEAU,	Professeur à l'Université de Paris Ouest - Nanterre La Défense

L'UNIVERSITE DE PARIS OUEST NANTERRE LA DEFENSE n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses; ces opinions doivent être considérées comme propres à leurs auteurs.

Résumé

Marchés Actions de pays Emergents: analyse de la dynamique des rendements et stratégies d'investissement

par

Juliana CAICEDO-LLANO

Directrice de recherche: Catherine BRUNEAU

Cette thèse évalue des modèles de gestion de portefeuille et d'évaluation d'actifs financiers avec des données de marchés d'actions de pays émergents. Son objectif est d'analyser la dynamique des comovements, la prévision des rendements et leur intégration dans la construction de stratégies de portefeuille. Elle met en évidence des différences de structure des rendements de marchés développés et émergents ainsi que l'importance de la gestion du risque dans le temps.

Le premier chapitre étudie un large échantillon de séries d'indices boursiers avec une analyse factorielle pour estimer des facteurs communs principaux par rapport auxquelles des comovements sont calculés. La structure factorielle des rendements et la dynamique des comovements sont analysées et associées à l'intégration financière et la diversification.

Le deuxième chapitre propose un modèle de prévision des rendements inspiré de modèles de la valeur présente. Ce modèle permet d'estimer des rendements espérés qui sont utilisés dans le dernier chapitre pour une application à la gestion de portefeuille. Un modèle de cointégration est utilisé pour étudier les relations de long terme entre les variables fondamentales et les prix des indices.

Le troisième chapitre développe des techniques de construction de stratégies de portefeuille. Des informations concernant la structure dynamique des rendements et les différents niveaux de sont prises en compte pour modifier les poids issus d'une optimisation de portefeuille. Cette illustration permet de comprendre l'importance de la gestion des comovements dans la gestion de portefeuilles au cours du temps.

Mots-clefs : Marchés émergents, Comovements, Intégration financière, Prévision de rendements, Cointégration, Théorie du portefeuille, Modèles d'évaluation d'actifs

Classification Jel: C61; G11; G15

Abstract

Emerging equity markets: analysis return dynamics and portfolio strategies

by

Juliana CAICEDO LLANO

Supervisor: Catherine BRUNEAU BRUNEAU

This dissertation aims at investigating some particular features of emerging equity markets that are central to the construction of equity portfolios. These features are confronted to the financial theory of portfolio selection and asset pricing. We provide theoretical and empirical analyses of the behavior of emerging equity indexes.

We focus on the dynamic of comovements among emerging capital markets, the predictability of equity returns and the integration of these parameters in the construction of portfolio strategies. Our study identifies the specific features of emerging equity markets It highlights the importance of risk management over time and the variability of the structure of equity returns.

The thesis is organized around three chapters developing the topics of comovements, estimation of expected returns and construction of portfolio strategies.

The first chapter analyses a large number of equity indexes corresponding to specific sectors and countries and uses a factor models technique to estimate the principal factor that would be used as a reference to calculate the comovements. This study shows that the structure of equity returns is different for emerging or developed markets and that it has a time varying nature.

The second chapter proposes a prediction model for equity returns influenced from present value models which takes into account the results concerning the time varying structure of returns and the number of factors to include in a factor model. Our model estimates expected returns that would be used in the last chapter in a portfolio strategy application. A cointegration model is used to study the long term relation of fundamental variables with the level of prices for equity indexes.

Finally, the last chapter of the thesis investigates the idea of the dynamic structure of return and suggests to modify the weights issued from the portfolio optimization taking into account the comovements calculated with the methods described in the first chapter. This study highlights the importance of the comovements management over time in

portfolio management.

Key words: Emerging Markets, Comovements, Financial integration, Return Predictability, Cointegration, Portfolio theory; Asset pricing
Jel Classification Codes: C61; G11; G15

Contents

General Introduction **1**

1 Stock markets returns comovements and their dynamics **19**

1.1 Introduction 19

1.2 Review of the related literature 24

1.2.1 Prior research on comovements 24

1.2.2 Factor models and asset pricing literature review 26

1.2.3 Integration and diversification studies 29

1.3 Methodology 34

1.3.1 Factor models 34

1.3.2 Determination of the number of factors 40

1.3.3 Measuring the strength of comovements 42

1.3.4 Interpretation of the factors and the measure of comovements . . . 44

1.4 Description of the Data 47

1.4.1 Description of the data base 47

1.4.2 Descriptive statistics 49

1.4.3 Stationarity 53

1.4.4 Factor decomposition 54

1.5 Main Results 59

1.5.1 Factor Model and Comovements 60

1.5.2 Integration 68

1.5.3 Diversification 75
1.6 Summary and Conclusions 78
1.7 Appendix 81

2 Forecasting Equity Returns **89**

2.1 Introduction 89
2.2 Literature Review on Comovements 93
2.3 Studies on the prediction of equity risk premia 93
2.4 Methodology 96
2.5 Methodology 96
2.5.1 Explaning Expected returns 96
2.5.2 Cointegration 98
2.5.3 Financial and econometric challenges 101
2.6 Data set description 104
2.6.1 Samples 104
2.6.2 Equity Returns and risk premia 104
2.6.3 Descriptive statistics 108
2.6.4 Stationarity tests 109
2.6.5 Non Stationarity of stock valuation indicators 112
2.7 Main Results 113
2.7.1 Cointegration analysis 113
2.7.2 Analysis for different emerging markets 119
2.7.3 Comparison to alternative models 121
2.7.4 Robustness 122
2.8 Summary and Conclusions 124
2.9 Appendix 125

3 Portfolio construction and investment strategies **131**

3.1 Introduction 131

3.2 Prior Research on Portfolio Construction 135

3.2.1 Estimation of expected returns 135

3.2.2 Market neutral portfolios 136

3.2.3 Shrinkage 137

3.2.4 Risk Budgeting 138

3.2.5 Transaction costs estimation 139

3.2.6 Performance measurement 140

3.3 Methodology 142

3.3.1 Estimation of expected returns 143

3.3.2 Portfolio Optimization 143

3.3.3 Shrinkage estimators 145

3.3.4 Risk Budgeting 145

3.3.5 Transaction costs estimation 146

3.3.6 Performance measures 147

3.4 Description of the Data 149

3.5 Main Results 151

3.6 Summary and Conclusions 161

General Conclusion **168**

List of Figures

1.1 Cross Correlation Emerging and Developed Markets 52

1.2 Cross Correlation Emerging and Developed Markets - Subsamples: 1994:200 - 2000-2006 53

1.3 Part of variance explained by main factors 55

1.4 Loadings from Principal Components analysis on countries indexes returns 57

1.5 Measures of comovements and number of factors for developed and emerging markets 63

1.6 Confidence Intervals and cumulative variance explained for DM and EM . 66

1.7 EWI vs F1 69

1.8 Commovements and Confidence intervals DM-EM 71

1.9 C1 for developed markets and C2 for emerging with CI 72

1.10 Percentage of Variance explained 74

1.11 Percentage of Variance explained 76

1 -A Cloud of points and estimation of PCA factors 85

2.1 Levels and differences: RP, DY, G and VIX for the US (2001-2006) 110

2.2 Levels and differences: RP, DY,G and VIX for Chile (2001-2006) 111

2 -A Input variables for the US market (2001-2006) 127

2 -B Input variables for Chile (2001-2006) 128

3.1 Long-short beta neutral payoff and integration measure- Emerging Markets 154

3.2 Long-short beta neutral payoff and integration measure- Developed Markets 154

3.3 Long-short beta neutral payoff and integration measure 158
3.4 Long-short beta neutral payoff and integration measure 158
3 -A Long-short beta neutral payoff and integration measure 166
3 -B Long-short beta neutral payoff and integration measure 166

List of Tables

1 Size statistics for Emerging and Developed markets and economies 2

2 Emerging equity markets 7

3 Emerging Market capitalization by region 8

4 Emerging equity markets ratio of market capitalization to GDP (2005) . . 10

1.1 Countries and Sectors 48

1.2 Sector codes and names 49

1.3 DS 49

1.4 ADF Tests for developed and emerging markets 54

1.5 Number of factors for different regions and different sample size 61

1 -A Descriptive Statistics calculated for the period 81

1 -B Correlation coefficients for the set of developed countries 82

1 -C Correlation coefficients for the set of emerging countries 83

1 -D Loadings from Factor Decomposition 84

1 -E OLS regression of C_t measure on a linear trend 85

2.1 Stationarity 110

2.2 Stationarity 111

2.3 Rank Tests 114

2.4 Cointegrating vectors 115

2.5 VECM for Chile and US (2000-2006) 117

2.6 Rank Tests 118

2.7 Cointegrating vectors 119
2.8 VECM for different countries (2000-2006) 120
2.9 AR US and Chile (2000-2006) 121
2.10 VECM for different countries (2003-2006) 123
2 -A EM and US market excess returns statistics (2001-2006) 126
2 -B Descriptive statistics for the US-w3m 127
2 -C Diagnostics for different lags in VAR 127
2 -D Descriptive statistics for Chile-w3m 128
2 -E Diagnostics for different lags in VAR for Chile 128
2 -F RankTest 129
2 -G RankTest 129

3.1 Results In local Currencies - Emerging Markets 152
3.2 Results In local Currencies-Developed Markets 156
3.3 Performance measures for different strategies 160
3 -A Descriptive Statistics - Emerging markets 163
3 -B Correlation coefficients for the period 01-1993 to 08-2006-Emerging Markets 164
3 -C Descriptive Statistics - Developed markets 164
3 -D Correlation coefficients for the period 01-1993 to 08-2006-Developed Markets 165
3 -E Results In USD - Emerging Markets 167
3 -F Results In USD - Developed Markets 167

Acknowledgment

There are many people to whom I feel grateful for their help in many different forms while writing this dissertation. Thanks first of all to Thomas whose help, support, patience and passion for philosophical discussions were absolutely essential to enter and get out from this long project. Thanks also to my family who encouraged me even in during the hardest times and special thanks to my grand mother for being so cheerful.

Thanks also to my dissertation advisor, Mme. Catherine Bruneau who supported this project since the beginning encouraging me to be a candidate for a dissertation fellowship and then being so patient but at the same time imposing precision and accuracy in my work during all these years.

I also thank the committee members for accepting to read and examine this dissertation.

Some parts of this dissertation were made in collaboration with Catherine Bruneau and Thomas Dionysopoulos, thanks to both for the great discussions that made possible to write these articles. I have also to thank all the referees and discussant in different conferences who gave invaluable advises to improve each work.

I also thank the group of doctoral students of EconomiX and former Thema who were devoted to the coffee machine and who were always open for any kind of discussion which made the years dedicated to this dissertation more pleasant. My appreciation also goes to my colleagues at the department of Economics at Nanterre, especially the members of the for providing a stimulating atmosphere for this dissertation. The discussion, encouragement, and friendship of Helene, Pascal, Marie-Laure, and Skander were very important to keep believing that this project would be accomplished wone day. I am also deeply grateful to Sophie and Veronique for their devoted friendship and unconditional support. Thanks to Tunisia for giving me so great office mates. I also thank my new colleges at ESG: Olivier, Eric and Pierre for their encouragement, their experience and useful advises

have been a lot of help in this final step. To Armand Derhy, thanks for his patience.

To my friends and family who understood how important was this project to me and were patient and supportive. Natalia, Mami, Santi, Papa, thank you.

General Introduction

In 1970, the amount of portfolio investments in equity markets of emerging countries was close to zero. In 2006, these markets received more than $100 billion US Dollars in portfolio investments[1]. The world market capitalization of emerging markets grew up from less than 100 billion US Dollars in 1983 to almost $4000 billions in 2003 and exceeded $5000 billions in 2005. In ten years, the size of emerging equity markets have been multiplied nearly by 40. This market capitalization figure represented less than 2% of world market capitalization in 1983 and topped 24% in July 2009[2] (See Table 1).

In 1986, the International Finance Corporation (IFC) was the first to create an emerging market mutual fund invested in securities from these markets. The launch of this fund, seeded with $50 million US Dollars and invested in more than 30 different countries has interested pension funds and insurance companies and has open up the path for the development of investments in emerging countries.

Not only equity markets attracted international investors during the last decades, there is also an increasing interest bond markets from emerging countries. The level of foreign investment in emerging market local currency bonds has risen dramatically in recent years[3]. Indeed, assets under management of emerging market funds benchmarked to emerging market bond indices grew from $27 billion US Dollars in late 2000 to about $230 billion as of mid-2006.

Various reasons have made emerging markets to become so attractive for investors.

[1]Source: IMF, International Financial Statistics database
[2]Sources: Standard & Poor's Global Stock Markets Factbook 2005 and Bloomberg
[3]Source: Global Financial Stability Report, September 2005, International Monetary Fund.

First of all, emerging countries achieved higher long term growth rates than developed markets. Some emerging economies have been growing twice as fast as advanced economies, in a continuous way during the most recent years. In 2007, the growth rate of high income economies was close to 3% while emerging economies were growing at more than 7% on average[4]. The rapid growth of these economies is the product of several improvements in key economic indicators such as inflation, current accounts, or net direct investment.

Table 1: Size statistics for Emerging and Developed markets and economies

	1990	**2000**	**2005**
Part of world market capitalization			
Emerging Countries	4.3%	5.6%	10.6%
Developed Countries	94.6%	92.9%	89.4%
GDP growth			
Emerging Countries	0%	0%	0%
Developed Countries	0%	0%	0%
Part of world GDP			
Emerging Countries	0%	0%	28%
Developed Countries	0%	0%	72%

The four famous BRIC countries (Brazil, Russia, India and China) are currently part of the ten largest worldwide economies and according to the projections of Goldman Sachs by 2050[5], these countries would be part of the six largest economies in the world.

Other economic statistics that are interesting to highlight are the part of the world population accounted by emerging economies and the part of economic activity they generate. The population represents more than 80% of the world population with 3.5 billion inhabitants. In terms of economic world activity, emerging markets generate more than 20% of it measured by GDP. The growth perspective of companies in such an environment, large population and rapid economic growth, are extensive even if the capital markets are not sufficiently developed. However, capital markets in emerging markets have followed a process of liberalization during the 1980's and are now more competitive and interesting

[4]World Bank 2008

[5]"Dreaming with BRICs: The Path to 2050" Goldman Sachs (2003)

for investors. In fact, their capital markets have been growing and transforming their structures to gain in efficiency and competitiveness. Even if they are still far from the levels of efficiency of developed countries, the increasing efficiency has reduced transaction cost and make than them more valuable for international investor. The second reason for emerging markets to be attractive for investors is that these markets have high perspectives of performance which is a key element for investors. George Hoguet, and investment strategist for the firm SSGA[6], argues that for the 18 years ending on August 2008, the MSCI Emerging Markets Index had outperformed the S&P 500 by more than 76 basis points per year.

The third reason is that these markets have been relatively cheap compared to developed markets. Their average valuation ratios, such as the price to earnings ratio, during the last years is close to 14 to 16, excluding the recent bubble of asian markets, compared to 17 to 18 for markets in developed countries. Value investors would be interested in stocks with low price to earnings ratios and would therefore be interested in emerging equities.

Finally, emerging markets offer important levels of diversification to a global portfolio due to their relatively low correlations to other emerging markets and mainly to markets of developed countries. This last point would be largely studied in this dissertation as we consider that it is essential for the construction of portfolio strategies.

Besides of the attractiveness of emerging markets, there are also some negative characteristic that push investors out from these markets.

First of all, these markets present a higher risks. The volatility of their returns is higher as well as the risk of bankruptcy of their companies. Political risk is also higher in these countries and have a significant influence in investing decisions. In emerging markets, the number of participants could be lower than in advances markets making the risk of

[6]State Street Global Advisors

liquidity an important parameter to be considered. In addition to this set of risks, the investor that is interested in emerging securities has to deal with other potential risks. The combination of securities from different countries can bring an additional currency risk. Another feature related to the cases where multiple securities is considered, is the similar behavior of these markets in times of crisis. In fact, contagion effect have been observed in emerging markets in recent years. These factors, market risk, credit risk, liquidity risk and political risk would be taken into account by investors when taking the decision to allow some capital to this type of assets. And when different stocks are included, currency risk and extreme correlation issues have to be integrated in the analysis. Second, these markets are less efficient and the integration of news in the prices take a longer time. The bad transmission of information is due to the lower levels of transparency, the scarcity of corporate information, and the general lower level of efficiency. And finally, these markets are subject to difficult legal framework and other restrictions such as higher transaction costs or lower execution times.

As a result, investors have to find a trade off between the risks and rewards generated by these markets. Indeed, even after accounting for all the risks associated to these markets, emerging equities continue to attract an important number of professional investors.

A specific group of investors interested in emerging markets are hedge fund companies. These funds are quite heterogeneous and adopt different kinds of strategies such as equity long/short, event driven, global macro and fixed income arbitrage. In the first quarter of 2008, emerging market hedge funds managed approximately $110 billion US Dollars, x% of assets under management for this category of professionals. Another type of investors follow capitalization-weighted indexes, there is a important amount of money held in indexed portfolios tracking different emerging indexes.

The growing interest in emerging equity markets not only comes from practitioners, since the years 1990, multinational institutions such as the World Bank or IMF have

been publishing reports of the international financial outlook including statistics from the equity market and since 2000 a number of academic journals have been created, witnessing the interest of academia in this kind of markets.

The main questions addressed by academics have been summarized by Harvey (2003). The study of integration and liberalization issues has been at the center of academic studies on emerging markets. Other financial issues such as the level of returns, the different kind of risks, the changing structure of capital flows and diversification have been studied during the 1990's and 2000's. Other topics of interest have been the effect of contagion, the specific features of crises and different models having the object of predict them. The decomposition of risk based on asset pricing models, the non-normal distribution of returns and its implication to portfolio construction as well as different asset allocation issues have also intrigued researchers for several years.

Since the seminal work of Solnik in 1974, international investors are aware of the potential benefits from diversification when they include foreign equities in their portfolios. In practice, only during the 80's, the development of emerging markets funds has allowed investors to profit from this issues. However, a large strand of the financial literature presented, since the mid-1970's, emerging markets as providing benefits in terms of diversification.

According to Campbell (1996), in a well diversified international portfolio a US investor should include 15% of emerging markets assets. Non-US equity markets represent more than 50% of world market capitalization, however US investors only hold an average less than 14% in foreign stocks within their equity portfolios.(See Solnik (2008) [7].) In Europe, the statistics similar. In fact, the part of a typical European portfolio invested abroad is much higher than in the case of a US portfolio but after excluding the investments within countries of the European area, the part of the portfolio allocated to other international markets represents less than 10% of the investment. In brief, after 30 years of internation-

[7]Data Coordinated Investment Portfolio Survey (CIPS) 2003

alization of financial markets, a small part of an international portfolio is invested abroad. Many reasons can explain this suboptimal behavior of international investors.

All the risk presented above can push the investors to reduce their exposure to emerging markets. Lower liquidity, higher transaction costs, a complicated law framework or the ”home bias” behavior are only some of the possible reasons that could explain this situation. Despite of this, a UK pension funds analysis by Babilis and Fitzgerald (2005) shows that the share of overseas equities in these portfolios rose from 26% in 1992 to 35% in 2001 and the share of emerging market equity within the portfolio of overseas equity increased almost consistently over the sample period from 13% in 1992 to **28%** at the end of 2001.

Even if the BRIC term has made of emerging markets a popular subject for any kind of investor, and even if now, every financial journal has an emerging markets section, we have observed that there is no broadly accepted definition or clear categorization of an emerging market. The IMF, classifies countries into three categories: advanced economies, developing countries, and countries in transition. The World Bank classifies them according to their gross national income (GNI) per capita, as low-income, middle income and high income economies. According to World Bank classification, there are in 2005 more than 200 economies in the world and among them only 60 are considered high income economies, emerging markets are composed of low-income and middle-income countries. Emerging countries are very diverse in economic as well as in political terms, countries such as Zimbabwe are among the poorest on earth but Korea and the Check Republic are members of OECD.

From a financial point of view an emerging market is an investment class of instruments uncorrelated to developed markets. This is a more complicated and subjective definition because it depends on the levels of correlation chosen to accept a market as a non-correlated one to the Developed Markets . In the financial markets, investor are

guided by the list the countries included in indexes that are labeled as emerging markets. Over the last years, some countries have changed from emerging to developed indexes according the criteria established by the company constructing the index, usually the world largest data providers such as S&P, MSCI or FTSE[8].

The focus of this document is on 50 countries corresponding to the highest market capitalization markets. We have 24 developed countries and 26 Emerging countries that we list below:

Table 2: Emerging equity markets

EMERGING MARKETS	
ARGENTINA	MEXICO
BRAZIL	PAKISTAN
CHILE	PERU
CHINA	PHILIPPINE
COLOMBIA	POLAND
CYPRUS	ROMANIA
CZECH REP.	RUSSIA
HUNGARY	SOUTH AFRICA
INDIA	SRI LANKA
INDONESIA	TAIWAN
ISRAEL	THAILAND
KOREA	TURKEY
MALAYSIA	VENEZUELA

In the first chapter of this dissertation we use a framework where the inclusion of a large number of data is important and therefore we include the maximum number of emerging time series. In the second chapter we focus on a reduced sample of 6 countries, 5 emerging and the US, as we do not consider the cross-sectional dimension. In the third chapter, we study two different sample: one composed of 17 emerging countries and one with 14 developed ones.

[8]S&P: Standard and Poors, MSCI: Morgan Stanley Capital International Inc., FTSE: Company owned by Financial Times and the London Stock Exchage

From the 26 countries, a small set represents a large part of the whole set of emerging markets.

The BRIC countries represent nearly 50%of the total Emerging Markets capitalization and only 8 countries (BRIC countries plus Korea, Mexico, South Africa and Taiwan) represent about 75%. India and China together account for 15% and Brazil and Mexico almost 20%. The distribution of market capitalization by regions is the following:

Table 3: Emerging Market capitalization by region

Region	**Market Capitalization (End 2005)**
Asia	47%
Latin America	25%
Emea	28%

EMEA: Europe, middle east and Africa region

Emerging countries have changed dramatically during the last 20 years. This period has been a time of growth and transformation for these economies and their markets. During the 80's the asian tigers (Hong-Kong, Singapore, Taiwan and South Korea) have experience a very rapid economic growth. In Europe, countries such as Portugal and Greece have integrated the European Union accelerating their growth rates and modernizing the markets. During the 90's, Latin America privatized state companies and liberalized of financial markets. In eastern Europe, the end of communism has boosted the growth countries such as Hungary, Poland, and the Czech Republic. During the last years China and India have been the center of attraction of Emerging Markets but also African countries such as Zimbabwe and South Africa have increased their openness to external markets enhancing the profitability of their equity markets. However these years where also times of crisis: Brazil in the beginning of 1994, then Mexico at the end of 1994 (Tequila Crisis), Asian crisis in mid 1997 followed by the Russian crisis in September 1998. And to finish Argentina in 2001.

As a results, many countries considered as emerging markets not that long ago, are now

part of efficient and modern financial markets. For instance, the Greek market was pull out officially from the MSCI Emerging Index integrating the indexes of developed markets in 2000. Then, defining what an emerging market is could be a difficult task. And as the set of emerging markets is continually changing the features that are considered today as specific of emerging markets could be perceived differently tomorrow.

We have stated that there is no specific definition for emerging markets, however the analysis of several features can help to categorize these markets. Rather than defining an emerging market, we focus on describing the principal features that can be useful to distinguish these markets from the more advanced ones. In the following paragraphs, we present three elements that we consider important to categorize emerging markets: the level of concentration, the degree of financial development and the industrial structure.

Let's start with the level of concentration which measures the dispersion of market capitalization in a market. A very concentrated market would have a large part of its market capitalization represented by only a few stocks. And a market with low concentration would have its biggest stocks representing only a small part of the total market capitalization. Usually developed market have low concentration, the US market biggest securities represent only 3% of the whole US capitalization. But some developed countries can be very concentrated, it is the case for Switzerland having 10 of the biggest firms representing almost 70% of the market.

Emerging and developed countries also differ in their degree of financial development. The ratio of market capitalization to GDP is often used to gauge this feature.

Table 4: Emerging equity markets ratio of market capitalization to GDP (2005)

Name	CODE	Mkt Cap	GDP	Ratio
SINGAPORE	SG	175 677	116 775	1.50
UK	UK	2 907 154	2 229 472	1.30
US	US	14 030 934	12 455 825	1.13
SOUTH AFRICA	SA	261 008	239 419	1.09
CANADA	CN	1 184 669	1 132 436	1.05
JAPAN	JP	4 072 844	4 567 441	0.89
FRANCE	FR	1 615 964	2 126 719	0.76
BELGIUM	BG	264 501	371 695	0.71
BRAZIL	BR	460 472	795 666	0.58
RUSSIA	RS	422 103	763 287	0.55
INDIA	IN	380 748	771 951	0.49
PHILIPPINE	PH	35 775	98 371	0.36
MEXICO	MX	227 434	768 437	0.30
CZECH REP	CZ	35 989	124 310	0.29
CHINA	CH	156 756	2 234 133	0.07

A special case is Hong-Kong with a market capitalization... more than for times larger than its

Most of Developed countries have a ratio of market capitalization to GDP of about 1 and the majority of emerging countries have lower ratios closer to 0.5.

The industrial structure is also very different if we compare emerging and developed countries. Developed countries have usually multiple vigorous sectors that contribute in more evenly distributed proportions to the economic activity than in emerging countries where very often not all the economic sectors are present or only one or two sector represent the main economic activity of the country. If we consider only the BRIC countries, we can see that each country has a strong sectoral advantage: Agriculture for Brazil, Energy for Russia, Services for India and manufacturing for China. Some mayor world companies are located in Emerging Markets reinforcing sector concentration. For example, the leading semiconductor manufacturer, Samsung electronics, is a Korean firm, the world's leader in Internet security, Checkpoint software, is an Israeli company and some of the world biggest energy firms, Surgutneftegaz and Gazprom, are located in Russia.

We will see in the first chapter of this dissertation, that the differences between emerging and developed markets can be captured with the analysis of the structure of equity

returns.

The three features presented above, combined with other elements such as the higher volatility, higher returns make that some investors consider these markets as a different asset class. In fact, as the characteristics and behavior are so different as well as the operational and legal frameworks, the correlation of equity returns between these two types of markets is low making them attractive in terms of diversification.

It's important to recall the basic theories used in this study. First of all, the modern portfolio theory developed by Harry Markowitz in 1952 which formalized the idea of diversification, and second, the asset pricing literature initiated with the work of Sharpe (1964) and Lintner (1965) which developed the idea that investors demand extra returns for taking on more risk. Both nobel laureates, changed the course of finance and transformed the way investors and managers used to take decisions.

The central idea of modern portfolio theory is that it is possible to combine individual securities or financial assets in certain proportions that would produce a portfolio with a lower overall level of risk and average level of return. This is the principle of diversification which constitutes the center of financial theory. In the model of Markowitz, it is assumed that all investors are rational and risk averse and that the distribution of asset's returns are completely described by their two first moments and they are assumed to be distributed normally.

The Capital Asset Pricing model (CAPM) of Sharpe and Lintner stated that there is a linear relation between the excess returns of a security and the excess return of the market above a given level of risk free rate. Other model were developed later suggesting a more general specification of asset pricing including a higher number of factors, non-linearities or dynamic specifications. The CAPM model styates that:

$$E[R_i] = r^f + \beta(E[R_m] - r_t^f)$$

where $E[R_i]$ is the expected excess return of asset i, r^f is the risk-free rate, β is the

sensitivity of the asset's expected excess returns to the expected excess market returns or market risk premium that correspond to the term $(E[R_m] - r^f)$. To test empirically the CAPM with a data set, we calculate expected return from asset and market returns time series allowing for an error term.

"The Gordon formula states that the real return from buying and holding a stock forever and consuming the dividends is the dividend yield plus the real rate of dividend growth."

The main objective of this dissertation is to analyze different features of emerging equity markets that are important in the context of portfolio construction. We focus on the changing structure of portfolios over time. In particular we study the comovements of equity markets and the structure of equity returns as main inputs for the portfolio construction. In fact, an estimation of a covariance matrix is needed to find the optimal weights and an estimation of expected returns. For this, we analyze different models that are able to capture some predictability in equity returns and we use the estimates in our portfolio construction.

In this dissertation we use the framework of asset pricing as a basis for each part of the analysis. In the first chapter we decomposed equity returns into orthogonal principal factors by the use of factor analysis models. In the second chapter, we try to explain equity returns with other fundamental variables that are used in the framework of present value models. In the last chapter, we show that in practice equity returns differ from the asset pricing model and a component of equity returns not related to common factors would be of great interest in the structuring of a portfolio strategy.

In the first chapter of this dissertation, we focus on the time varying nature of comovements for emerging and developed markets. We start by quantifying the level of comovements for the period of study and then we analyze its impact on integration and diversification.

We analyze a comovements measure for a set of international country-sector indexes over time. Recent techniques of factor analysis were applied to study of the time varying feature of comovements. The main factors are used as a reference to calculate the comovements. We analyze a large number of equity market return variables and its decomposition into main factors in order to understand the structure of equity returns. The measure is calculated at periodic dates in order to observe the dynamics of the comovements. We show that the comovements are increasing and try to analyze the impacts of this phenomenon. We show that a higher level of comovements is related to a higher level of financial integration and a lower level of potential diversification. The decomposition of equity return into factors leads to the rightful question of the number of relevant factors to include in the factor structure. The use of a large data setting allows to use inference results recently presented by Stock and Watson (1998), Bai (2003) and Bai and Ng (2003). More specifically, several information criteria can be used to determine the number of factors and confidence intervals are use to analyze if observed factors correspond to the true factors.

We compare the diversification ratio introduced by Solnik (1974) with the measure of comovements analyzed in this chapter, we observe that both measures are closely related and we conclude that the opportunities of diversification have been reduced over time with the integration of equity markets.

Other important results that would be presented in this first chapter are the finding that the first factor is closely related to an equally weighted index calculated form the different variables in the data set. From the analysis of the measure of comovements over time we conclude that for both kinds of markets - emerging and developed - there is an

upward trend in the level of comovements. However, for the set of developed markets this trend is stronger confirming that the behavior of emerging and developed markets are different.

From the comparison of the different analysis of the number of factor to include in the model, we conclude that for the set of developed markets only one factor is needed whereas for the set of developed markets at least two factors would be part of the structure of equity returns. This result would be confronted to the results from the other chapters.

This chapter shows that the measure of comovements could be related to important economic and financial issues such as the process of integration of equity markets and the time varying opportunities of diversification offered by international markets. Both having an impact in portfolio selection problems.

The second chapter focuses on the predictability of equity returns for a reduced set of emerging countries and the US market that is included as a example of developed markets and is useful to compare the results from the analysis of emerging markets. We analyze a set of five Latin American countries. In the first chapter we have found that only one factor extracted from the whole set of variables is needed to explain the structure of equity returns for developed markets. Here, we aim at finding variables having some predictive power on equity returns. If only one variable is needed to forecast developed markets and more that one variable is needed for the set of emerging markets, the results would be in line with those of the first chapter. This chapter presents, first, a classic model founded on a present value relation to analyze the predictability of equity markets. Based on this fundamental approach, we start by analyzing the equilibrium relation of the set of variables in levels and we determine if there is a cointegration relation among the variables. Then, we analyze the impact of this error component on the dynamics of the

equity risk premia. Our work presents several innovations to the literature. First of all we focus on the changes of equity risk premia rather than the risk premia variable itself in order to include the maximum possible information and to respect the econometric restrictions of the model. Also, we study if there is an improvement when combining a set of fundamental variables with variables that, to our best knowledge, have rarely been studied as forecasts of excess returns such as the implied volatility[9]. The main result of our analysis is the finding of different cointegration relations within the selected set of variables could be established. The results are confirmed for different countries. We set this conclusion after running the rank tests of Johansen which are confirmed with the estimation of a negative and significant coefficient for the error correction term for several of the countries that we have analyzed. From the estimated model we can conclude that there is some evidence of predictability for this set of countries. We also tested a specification including a growth variable but even if the rank tests suggest the presence of cointegrating relations, the estimated models have no significant coefficient and poor adjusted R^2.

The introduction of the implied volatility variable in our model helps to improve the prediction. This variable has rarely been used in the literature as a forecasting variable for equity returns. The limits of this chapter are that we haven't analyzed cointegration relation for different horizons of returns or different sample. This work would have helped to determine the robustness of the estimations. Also we should have analyzed the predictive power of the models out-of-sample in order to really asset the value of the estimations.

In the first two chapters we analyzed the comovements of equity markets and a model of predictability. These two elements are the essence of the classic portfolio model. In fact, when we determine with an optimization technique the best quantities to be invested in each different asset, the elements that we need to include in the optimization problem

[9]See Bollerslev and Zhou (2007a)

are the expected return of every asset and an estimate of their variance/covariance matrix. Then, we can use the results of the previous chapters to analyze the construction of a portfolio strategy. Here, we focus on a special case of portfolio optimization which correspond to the case where we do not include a constraint for the quantities to be invested as positive values. It means, that we focus on a portfolio that allows to short some assets. The two main contributions of this study are the decomposition of the portfolio risk into a market component and an alpha component and, the analysis of integration as an elements related to the market risk component. And second, we suggest an innovative way of managing risk through time, introducing integration as a key variable to manage long-short portfolios. We consider a simple long-short portfolio selection process based on forward earnings yields[10] and we build a zero-beta optimized portfolio using a covariance matrix corrected with a shrinkage procedure. We measure integration for every time period and we derive a scaling factor for our long-short weights at each date, that modifies the optimal weights in a way such that when markets are strongly integrated we take a more conservative position. With this procedure, we protect the strategy from the downside risk that a pure alpha investor would face during periods of high integration, when markets tend to exhibit a strong directional trend. In this chapter, we have analyzed separately emerging and developed market for the same period of study. First, we have built a beta neutral portfolio strategy and analyzed its performance indicators. We have used indicators such as the information ratio, the Sortino ratio and maximum drawdown which are pertinent to evaluate this portfolio strategy and are commonly used by practitioners. Then, we have constructed a strategy that takes into account the actual levels of integration and reduces the risk taken when the measure of integration is higher than a given threshold. This strategy presents better performance indicators so we conclude that the management of risk over time based on the criteria we have defined could be very useful for portfolio managers. We have also observed that for the set of emerging

[10]Earnings per share divided by the actual Stock/Index Price.

markets using our risk budgeting techniques enhances more the performance than for the set of developed market . In order to analyze the robustness of each portfolio strategy , we have compared the compared the results of a data set in US Dollars with a data set in local currencies. We have observed that for the set of emerging markets there is no significant change between the two data sets, however, for the set of developed market the strategy calculated in US Dollars is penalized compared to the one calculated in local currencies. In addition to the robustness of the currency used, we have also analyzed the impact of transaction costs. For this, we have calculated an annual average cost per transaction and a measure of turnover based on the changes of weights for the instruments in each portfolio. We compute the performance of each strategy taking the measure of transaction costs into account and we observe that all the strategies are penalized but still deliver good estimates of performance letting us to conclude that the idea of considering the level of integration could be important for the management of equity portfolios and particularly for emerging markets .

Chapter 1

Stock markets returns comovements and their dynamics

1.1 Introduction

The analysis of comovements is essential in the context of portfolio theory as it has a direct impact on the risk reduction potential by means of diversification. As stated in the European Central Bank document by **?**, stock returns comovements are "at the heart of finance". The interest in studying comovements across assets (or asset classes) in finance comes from the need of constructing better portfolios in a mean-variance framework. Since the formalization of the portfolio theory with the work of Markowitz (1952), financial investors have changed from selecting securities on an individual basis leaded by the risk and return features of each sock to a decision process where the core is the risk reward of the overall portfolio. In this context, understanding the comovements across securities becomes the key of portfolio construction because the existence of imperfect comovements permits to diversify the overall risk of the portfolio.

Despite their important role in portfolio construction, equity returns comovements are not easy to master. One of their principal features is that there are not constant through time. It is therefore important to understand their dynamics. In this chapter, we study the

time varying structure of comovements and its impact on integration and diversification issues. For this, we analyze a measure of comovements for a set of international country-sector indexes over time, using a factor model approach. As we will discuss in section 2.3, many studies have analyzed these issues, this study contributes to the literature using recent findings of the factor models literature and focusing on equity markets of emerging countries which have a secondary place in the literature on comovements. We use actual techniques of factor analysis mainly used in macroeconomic studies that we apply to a financial subject such as the study of varying comovements. Many documents have presented results about comovements for developed markets but only a few have been interested in analyzing the comovements with emerging markets or among them. Our study gives particular attention to emerging equity markets pointing out the differences in their factor structure compared to developed ones.

Our work lies within different strands of the literature. First of all, it is a study of the linkage between equity returns. This field has been largely explored, specially for advanced markets. This study is similar to Longin and Solnik (1995) and Goetzmann et al. (2005) who analyzed correlation coefficients of different sets of developed markets and study their stability across time concluding that correlations are time-varying and that they rise during periods of high integration and higher volatility.

Second, this study is based on the financial and economic literature of factor models initiated by Ross (1976), Chamberlain and Rothschild (1983) and Stock and Watson (1988). To analyze the structure of comovements, we use a factor decomposition of returns in a large data setting environment. We suggest a dynamic measure capturing the intensity of the comovements based on the part of variance explained by orthogonal factors extracted with principal components analysis and estimated with a rolling procedure. The idea, derived from the APT literature[1], is that markets are ruled by common factors and in the way that common comovements are stronger, specific risk and diversification potential

[1]See Ross (1976) or Burmeister and McElroy (1988)

are lower. The factor estimation procedure is similar to the one employed by Connor and Korajczyk (1993). We use a data set with a large number of variables and observations so that we can use the inference results recently obtained by Stock and Watson (1998), Bai (2003) and Bai and Ng (2003). Thus, the principal components can estimate consistently the factor space spanned by the true factors. Moreover, we can use information criteria to estimate the number of factors consistently as in Bai and Ng (2002). Additionally, we build confidence intervals around the true factors in order to find if an observed factor lies in, and thus give an interpretation of the estimated factors.

Third, we analyze the comovements with the perspective of relating them to two important financial topics: the integration of financial markets and the concept of diversification. The use of comovement related measures to capture integration has been largely debated without a definite response, we can mention the work by Bekaert and Harvey (1995) or Errunza et al. (1999) as some examples. We argue that using a large data setting, comovements could be used to gauge financial integration. In fact, if we extend the analysis of comovements to the international context, comovements would be impacted not only by the features of the assets and the markets to which they belong, but also by the barriers existing between different markets. In this study, we suggest different approaches to measure integration based on comovements measures and we conclude that the level of financial integration has increased during the period of analysis.

The other subject of interest in this chapter is the relation between comovements and portfolio diversification. The literature on portfolio construction and diversification is extensive. However, many authors only mention that an increase in the comovements has an impact on diversification without really analyzing it. Here, we focus on the diversification ratio defined in Solnik (1974) in order to study its relation with a measure of comovements. We show that both measures are closely related and therefore we can conclude that the diversification potential has been reduced over the analyzed period with the increase of comovements.

The appropriateness of using the comovements as an estimate of integration has been longly debated in the literature and, as an important point of this chapter, it deserves some discussion. Indeed, Bekaert and Harvey (1995) argue that correlation is not a good measure of market integration when the industrial structures of the markets are very different, however they admit that correlation in all cases give an indication on integration and they present evidence of a significant increase in correlations after an event of liberalization. Nevertheless, Bekaert and Harvey (1997) state that the increasing influence of world factors on expected returns may appear in increased correlation with global markets. While, Errunza et al. (1999) advocate that higher correlation is a signal of greater integration but not necessarily a cause of it, Carrieri et al. (2005), on the other side, defend the unworthiness of using correlations of market wide index returns as a measure of market integration arguing that correlations underestimate the degree of integration. As we can see, there is no ultimate response to the appropriateness of using comovement related measures to gauge integration. Our point of view is that the integration of markets includes a lot of observable barriers as well as some invisible and unmeasurable ones and for this reason it is impossible to quantify exactly integration. Observable barriers can include legal constraints, transaction cost, trade policies or economic changes and unobservable barriers could include the attitude of investors (changes in market tastes) or cultural customs. Consequently, market integration would be a time-varying process with a trend towards full integration as communications become faster, transaction cost smaller and market policies more flexible while different countries engage in regional trade agreements. We think that for the purpose of this dissertation and the type of data that we analyze (country-sector indices and not individual stocks), we can use the comovements as a proxy of integration and we will use the two terms interchangeably over the sections. The reasons is that as we use aggregates, some effects on an individual stock would be offset by others stocks included in the same index. The choice of a comovement measure

as an estimate of integration imply that we adopt several assumptions. We assume that there is a linear relation between the variables and that the variables are mostly symmetric. In the context of linear factor models with a large number of variables, the use of correlation related measure is then a reasonable choice.

This first chapter is organized as follows: Section 2.3 places this work within the existing three main axis of related literature. Section 1.3 describes first the methods used in our analysis namely the factor model approach and its derived measures of comovements as well as the links to integration and diversification. Section 1.4 describes the data set used in this study, the decision taken about the estimation of the factor model and a broad analysis of correlations and factor decomposition. The main results of our analysis are presented in section 1.5 which focus on the dynamics of comovements, starting with the estimation of the main components and the optimal number of factors to include, following with the results about integration measures and diversification. Some conclusions and comments are found in the last section.

1.2 Review of the related literature

1.2.1 Prior research on comovements

Equity markets comovements are usually captured by statistical measures such as the covariances or correlation coefficients. The number of studies analyzing this statistical measures is extensive. Pioneer studies on this subject include the work by Grubel (1968) or Levy and Sarnat (1970) who showed the importance of including international securities to reduce portfolio risk. The first studies focusing on the statistical measures of comovements have calculated correlations levels and analyzed their stability over time, later its asymmetry has been explored as well as the presence of extreme correlations.

Longin and Solnik (1995) studied the correlation of monthly excess returns for seven major countries over the period 1960-1990 and Goetzmann et al. (2005) have examined the correlation structure of the major world equity markets over 150 years. Both analyze correlation coefficients of different sets of developed markets and study their stability across time. Both conclude that correlations are time-varying and that they rise during periods of high integration and larger volatility. In this chapter, we aim at measuring to what extent this feature is also present in emerging markets.

A different procedure to measure comovements is to focus on cross-sectional dispersion. This method is often used for large data sets studies. Dispersion averaged over a period of time is closely related to the opposite correlation coefficients[2].

Ang and Chen (2002), Karolyi and Stulz (1996) or Forbes and Rigobon (2002) studied the asymmetry of correlations between returns of international equity portfolios. In fact, correlation coefficients present some bias caused by heteroscedasticity and non-linearity. Ang and Chen (2002) among others[3] have studied the asymmetry in correlations of equity portfolios conditional on the sign of market returns; in bear markets, conditional correla-

[2]Campbell et al. (2001) examine cross-sectional dispersion of individual stocks, showing an increase in recent periods. Adjaouté and Danthine (2004) study cross-sectional dispersions separating country and industry effects. A period of low (resp. high) dispersion would correspond to high (resp. low) correlations.

[3]Longin and Solnik (2001), Cizeau, Potters and Bouchaud (2000)

tions tend to be higher than in bullish markets[4]. The measure of comovements we present in this chapter confirms this statement on the comovements' asymmetry.

Other authors have focused on extreme correlations. Longin and Solnik (2001) focus on the correlation of extreme returns and find that equity market correlation increases conditional upon market drops, but it is not the case for bull markets. Karolyi and Stulz (2003) capture extreme simultaneous market movements using a measure of "co-exceedances". This part of the literature is closely related to studies about contagion. Bekaert and Harvey (1995), Longin and Solnik (1995) or de Santis and Gerard (1997) analyzed conditional correlations using multivariate GARCH models which are indeed more flexible and allow to capture heteroscedasticity; however these models require the estimation of a large number of parameters and are therefore inappropriate for large data sets.

Our study is in some way similar to Goetzmann et al. (2005). In their paper, they have focused on a very long period of time and they have studied developed markets. Here, we use only a 12 years sample and we focus on emerging markets. We present an alternative method to capture the intensity of comovements based on the literature of large-scale factor models and we give special interest to emerging markets.

Statistical measures such as the covariance or correlation capture the whole variation among a couple of variables. In this chapter, we suggest to decompose the structure of returns into main components and analyze the comovements of each asset with these common components. Our measure of comovements appears to be an average of these comovements across all the markets certainly due to the fact that we consider stock market indexes better diversified than individual stocks. We use this measure of comovements to gauge financial integration and diversification.

Our study is similar to the recent paper by **?**, they also analyze comovements with

[4]Cho and Engle (2000), Bekaert and Wu (2000), Kroner and Ng (1998), and Conrad, Gultekin, and Kaul (1991) use multivariate asymmetric GARCH models to analyze the covariance asymmetry of stock portfolios.

a measure based on the part of variance explained by common components. They focus on a longer sample and impose a number of factors arbitrarily. Here, we go further in the analysis of the number of factors to include and also in the interpretation in terms of integration and diversification. Moreover, as previously announced, we focus on emerging equity markets and we compare the results for those markets with developed ones.

1.2.2 Factor models and asset pricing literature review

The second strand of the literature we rely on is the factor models and asset pricing literature. Factor models are frequently used to understand the structure of the variables and its covariability, to parsimoniously summarize the information contained in a set of variables, to identify risk sensitivities and to generate predictions. With the growing quantity of information that is now available for researchers and analysts, this kind of models can be very useful.

We can distinguish two main types of factor models depending on the observability of the factors. Economic factor models linearly relate a random variable to several observed variables. The coefficients of the linear function can easily be estimated with regression methods.

In macroeconomics and finance, many observable variables have been used to model different economic phenomena. The asset pricing literature applies factor models to the analysis of risk premia, models such as the CAPM of Sharpe (1964) and Lintner (1965) the APT of Ross (1976) or the Fama and French (1992) three factor model use this approach. We can consider the Arbitrage Pricing Theory (APT) as the most general model. In fact, Ross introduced in 1976 the APT showing that equity excess returns could be decomposed statistically with a linear factor representation into common and idiosyncratic components. The idiosyncratic part of asset returns having no price of risk, since investors can diversify them away by holding portfolios, makes that the security risk premia are therefore associated to the equity return's covariance with the common

components. Many asset pricing models can be considered as particular cases of the APT, a model that has been explored for more than 30 years in the asset pricing literature.

When the factors are not observable (latent variables), we have a statistical factor model requiring the estimation of both the factor variables and the loadings. Regression analysis impose that the predictor variables have to be linearly independent (their matrix must have full rank) so that coefficients can be estimated by least squares methods. In statistical factor models, this constraint is respected by construction. However, the interpretation of the estimated factors is in this case often difficult. Depending on the structure of the model and the different assumptions on which it relies, different specifications and estimation techniques would be considered Maximum Likelihood Estimation (MLE) or Principal Components Analysis (PCA)). The most restrictive specification is the classic factor model presented in Anderson (1963). Since the pioneer work by this author who formalizes the structure of factor models for multiple time series, a number of advances in the estimation techniques and inference results have been developed. Parallel to the development of statistical techniques, many applications in macroeconomics and finance have contributed to enrich the literature on this field. The classic factor model will be presented in Section 1.3.1 as well as other specifications derived from it. The strict version is based on the premise stating that the idiosyncratic or specific components have to be uncorrelated. This assumption has been considered to be too restrictive and difficult to be validated with economic data.

Chamberlain and Rothschild (1983) introduced an approximate factor model allowing for limited cross-sectional correlation in the factor structure. They showed that principal components analysis could be used instead of factor analysis to estimate the factors when the number of variables tends to infinity. Connor and Korajczyk (1986) show that, up to rotational indeterminacy, a consistent estimate of the factors could be given by the eigenvectors of a cross-product matrix when the approximate specification is considered.

Recent works on factor models focus on inference, estimation techniques and the pre-

sentation of economic and financial applications. Stock and Watson (2002) extend the asymptotic principal components method of Connor and Korajczyk (1986) proving the consistency of the estimates of the factors and loadings in the cases where the time and the cross sectional dimensions tend jointly to infinity. They also contribute to the literature by suggesting a Quasi-MLE as estimation method.

Bai (2003) derived inferential results relative to rates of convergence and limiting distributions. He determines in which circumstances the estimation error is negligible and how to calculate a confidence interval for the true factors. We make use of some of these inference results to build our analysis.

In a series of papers, Stock and Watson (1999, 2002a, 2002b) have used factor models to forecast macroeconomic variables. The basic idea of Stock and Watson was to extract some principal factors using a PCA from the correlation matrix of the input variables and then use this factors to forecast any variable of interest.

In this chapter, we follow the estimation method described in Connor and Korajczyk (1986) or Bai and Ng (2002) which is a static estimation. In fact, Forni et al. (2000) introduced dynamic factor models allowing for lags to enter in the factor specification. More details of the differences in these factor models specifications are presented in Section 1.3.

We use a static analysis that is repeated for different rolling windows allowing us to study the dynamics in the intensity of comovements.

Another important question relative to factor models is how to estimate the number of relevant factors in the model. Bai and Ng (2002) suggest to estimate consistently the number of common factors minimizing several information criteria. We used this method to determine the number of factors for each window.

Factor models are useful in large data sets environments as they allow to reduce the dimension of the problem by summarizing the information contained in all the variables by a smaller set of factors. They also eliminate measurement error in the idiosyncratic

movements when a large number of variables is included. We believe that factor models are useful methods to understand the structure of equity returns and to analyze their comovements.

1.2.3 Integration and diversification studies

The state of perfect integration implies that all the existent barriers between markets have been eliminated and as a result the risk premia associated with global factors are identical in any of those integrated markets. In other words, the law of one price should hold and no arbitrage opportunities should appear. Another consideration of the state of financial integration is that the pricing kernel for all the asset should be unique. The determination of the state of integration is therefore crucial to the correct pricing of assets. Integration has been explored for different kind of markets: real markets as well as financial markets. Within the financial markets, every asset class could be considered as a different market and thus integration could be measured separately for each instrument type. We revise here the literature related to integration focusing on the studies related to equity markets.

In a document from the European Central Bank (ECB), Baele et al. (2004) have classified integration measures into three broad categories[5]:

- First, price-based measures study on the declining importance of the country component in equity returns compared to the sector component is interpreted as a feature of increasing integration. The focus of this category of studies is on cross-sectional dispersion in sector and country index returns which is a measure inversely related to the degree of correlation.[6]

- Second, news-based measures analyze the increase in the common news component of equity returns, considered as a signal of integration. Equity returns are therefore explained by common factors or local news. A large number of studies about

[5]This classification has been used in many ECB publications, namely the "Indicators of financial integration in the euro area" as well as in studies by other authors.

[6]Authors such as Adjaouté and Danthine (2004) or Solnik and Roulet (2000) investigate this issues.

integration are part of this category namely all the asset pricing based studies of integration[7].

- Finally, quantity based indicators where the increase of international exposure is sign of financial integration. These indicators can be constructed by analyzing the amounts of investments abroad in equity funds: investment fund industry, pension funds and insurance companies[8].

All these approaches provide only approximate analyses of integration, other studies have tried to stick to the law of one price to test for integration despite the difficulty of finding identical assets in different markets. Hunter (2006) and Levy Yeyati et al. (2005) have compared the prices of Depositary Receipts[9] to those of their underlying shares in the market of origin. The problem of this approach is that only certain securities have ADRs traded in other markets and therefore this kind of tests are based on reduced data sets specially for emerging markets.

The most common kind of integration studies are those relying on asset pricing models such as the international version of CAPM or the APT which could be considered as belonging to the first category of studies according to the ECB classification. The first tests of market integration consisted in testing if markets were perfectly integrated as in the work by Stulz (1984) or perfectly segmented as in Lucas (1978) and Breden (1979). If markets were perfectly integrated then the international version of the CAPM should be validated and in the case where the local version of the CAPM was verified then markets should be considered perfectly segmented. Later, the concept of partial states of integration and segmentation was introduced in the work by Stulz (1981), Black (1972)

[7]For instance, Fratzscher starting from a uncovered interest rate parity (UIRP) relation establishes an equation to explain how shocks in world markets are transmitted to a specific equity market. Rouwenhorst (1995) or Brooks and Del Negro (2004) have tested if excess returns could be explained by global, local and specific factors.

[8]The "home bias" is used to assess integration as in Adam et al. (2002) or the recent work of De Santis and Gerard (2006) who analyze changes in "home bias" based on the comparison of actual and optimal share of foreign holdings.

[9]or ADRs: American Depositary Receipts

or Errunza and Losq (1985) allowing for different levels of integration.

In the perfect case of integration, the return of a security should be explained only by global factors.

The previous studies considered integration in a static way, without measuring integration through time. More recent studies have allowed not only for partial integration but also for a dynamic measures of integration as in **?**, Bekaert and Harvey (1995) and Korajcyk (1996) among others[10]. Bekaert and Harvey were the first to consider a time-varying measure of financial market integration. Using data on equity returns, they found that a number of markets exhibit time-varying integration, thus allowing them to identify the reasons for rejecting the international CAPM.

Our study is closer to the work by Korajcyk (1996) who also used factor models and then quantified integration based on an asset pricing model that includes the factors. He used a rolling measure of mispricing. Instead, we focused on the part of variance explained by the common factors estimated by rolling windows[11]. Our measure of integration derived from the measure of comovements is not exact. In fact, to measure integration correctly - respecting the law of one price- we should have analyzed different marketable assets traded in different markets as individual stocks rather than aggregate indexes as we did in our empirical study. However, as it is very difficult to find similar securities in different markets, a part from ADRs[12] or multinational companies that issue stocks in different markets, the analysis of equity indexes can be helpful in explaining equity markets comovements.

We also rely on studies concerning international diversification issues. The first studies on international portfolio diversification are the ones by Grubel (1968) and Solnik (1974), both exploring the benefits of diversifying internationally. After the publication of this

[10]Also the work by Ferson and Harvey (1991), Hardouvelis et Al. (2006), Stultz and Karolyi (2001)
[11]This approach is also considered in the forthcoming article of **?**
[12]ADR: American Depositary Receipt

studies, the interest for international diversification issues has not ceased to grow.

The benefits of investing internationally appear in two ways. First, the investment opportunity set is enlarged and it is possible to find assets in the expanded set having better features in terms of risk and return. Second, if the international assets are not fully correlated to the domestic ones, there should be opportunities to reduce the risk of the new portfolio while maintaining the levels of expected returns - the diversification of individual risks is then possible.

Different approaches allow to capture the benefits of diversification, always derived from the risk return framework. Huberman and Kandel (1987) and Bekaert and Urias (1996) analyzed the benefits of diversification applying spanning tests to determine if there is some gain by investing abroad. In fact, they test whether the addition of assets to an initial set composing a benchmark enlarges the space spanned by the initial set and consequently reduces the overall risk of the portfolio. In the same line and focusing on the changes of diversification benefits, Driessen and Laeven (2007) present a recent study for the period 1985-2002 where they find decreasing diversification benefits. Similarly, Lewis (2006) explores the risk reduction of different international portfolios analyzing the period 1970-2004 and concluding that there is a slight decrease in the gains due to the diversification.

Some studies analyzing the behavior of local investors towards foreign investments have signaled the presence of a "home bias". This behavior describes the tendency of investing larger amounts in domestic assets than in the optimal international allocation. **?** have analyzed data from the US and Japan concluding that these countries exhibit some "home bias". Explanations for this suboptimal behavior include the analysis of cost and risk that could offset or even exceed the gains from diversification[13], the presence of barriers and market imperfections [14] and several behavioral explanations suggesting that

[13]such as the work by Black (1974) or Stulz (1981)
[14]as documented by Lewis (1999) and Karolyi and Stulz (2002)

investors have a preference for closer markets[15].

A simple indicator of diversification benefits has been introduced in Solnik (1974). Indeed, he has defined a diversification ratio accounting for the variance of a portfolio over the average variance of a typical security in order to show that that the benefits increase as more stocks (or assets) are included in the portfolio up to a certain level. We use this the ratio of diversification in our study in order to relate it to the measures of comovements. We indeed show that there are very close links between the diversification ratio of Solnik and the measures of comovements that we present.

[15]such as in Kang and Stulz (1997) or Huberman (2001)

1.3 Methodology

In the following subsections of this chapter, we present the general framework of factor models with the asymptotic principal components estimation strategy exposed in the work by Connor and Korajczyk (1986) or Bai and Ng (2007) as well as the method developed by Bai and Ng (2002) to determine the number of relevant factors. Based on this factor decomposition we suggest a measure that could trace the level of comovements. The last subsection describes some useful tools employed to relate the results about the comovements to integration and diversification properties.

1.3.1 Factor models

Statistical models with multiple variables involve a large number of parameters. However, factor models frequently used in this kind of setup, are implemented as dimension reduction techniques that simplify the task of estimation. Nowadays, factor models have become very useful with the growing quantity of financial and macroeconomic information available for researchers and the continually improving speed of computation.

In general, factor models express an observable random variable X_{it} as a linear combination of some K observed or unobserved factors F_{kt} and an idiosyncratic component ε_{it}. To formalize it, let us note X_{it} the variable corresponding to the ith individual where $1 \leq i \leq N$, at the time where $t,\ 1 \leq t \leq T$.

This set of N variables obeys a factor model with K factors F_k at every time t such that:

$$X_{it} = \sum_{k=1}^{K} \Lambda_{ik} F_{kt} + \varepsilon_{it} \tag{1.1}$$

where Λ_{ik} the sensitivity (or loadings) of the variable i to factor k and ε_{it} represents the idiosyncratic component not correlated with the factors. This relation can be rewritten in matrix notation and in a more general specification allowing for a constant term for each

individual as:

$$X = 1_T \alpha'_N + F\Lambda' + \varepsilon \tag{1.2}$$

where $X = (X'_1,...,X'_T)'$ with $X_t = (X_{t1}, ..., X_{tN})'$ denoting the N -dimensional vector of the N random variable X_{it}. $\Lambda = (\Lambda_1, ..., \Lambda_N)'$ is the NxK matrix of factor loadings (with $\Lambda_i = (\Lambda_{1i}, ..., \Lambda_{Ki})'$) and $F = (F_1, ..., F_T)'$ (with $F_t = (F_{1t}, ..., F_{Kt})$) are the K factors. The term α_N generalizes the model for the case where a constant is required, and the vector 1_T is a vector of ones.

We can write Equation 1.1 in cross sectional form as $X'_t = \alpha'_N + F_t\Lambda + \varepsilon_t$ or in time series form as $X_i = \alpha_i 1_T + \Lambda_i F + \varepsilon_i$.

The variables X_i are determined by a common component ruled by the set of common factors and an idiosyncratic component represented by the term ε_i. There are different specifications for factor models depending on the different assumptions adopted for each component of the model. The *classic factor model* is usually determined for a fixed number of variables N much smaller than the number of observations T, the factors in the matrix F are non-correlated and identically distributed and also independent of ε_t; the idiosyncratic term ε_t is also independent and identically distributed with mean $E[\varepsilon_t] = 0$ and variance $V[\varepsilon_t] = \Sigma_\varepsilon$; and the idiosyncratic term does not present correlation at the cross-sectional level.

The basic assumptions of the classic model are the following :

$$N << T \text{ and } T \to \infty \tag{1.3}$$

$$E[\varepsilon_i] = 0 \text{ for all } i \neq j \tag{1.4}$$

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j \tag{1.5}$$

$$Cov(\varepsilon_t, \varepsilon_s) = 0 \text{ for all } t \neq s \tag{1.6}$$

$$Cov(F_t, \varepsilon_t) = 0 \text{ and } F_t, \varepsilon_t \text{ are i.i.d.} \tag{1.7}$$

In the case where ε_t is normally distributed, the factors and loadings can be estimated using the maximum likelihood technique.

The variance covariance matrix of X can be noted Σ_N. In the case where the factors are uncorrelated with the error term (as stated in Equation 1.5), the variance covariance matrix can be decomposed into:

$$\Sigma_N = \Lambda'\Omega\Lambda + D_N \tag{1.8}$$

Where $E[F_t F_t'] = \Omega$ and D_N is a diagonal matrix implying that the idiosyncratic shocks have no cross-sectional correlation.

In the case where the factors are uncorrelated and have unit variance $E[F_t F_t'] = I_k$ (factors are normalized, their variance covariance matrix is thus an identity matrix), the covariance matrix Σ_N is decomposed into:

$$\Sigma_N = \Lambda'\Lambda + D_N \tag{1.9}$$

This factor model version relies on several assumptions that are difficult to verify with economic and financial variables. The diagonality of the idiosyncratic covariance matrix would rarely be observed and it is very often possible to have data sets where the number of individuals is higher than the number of dates of observation.

Recent developments in the literature on factor models have released some of the assumptions of the *classic factor model* renewing the interest in these models as a number of applications can be studied in macroeconomics and finance with this more flexible specification. Indeed, Chamberlain and Rothschild introduced in 1983 an *approximate factor model* which allows to model more realistically the structure of different economic variables. They have shown that when the number of observations N tends to infinity, it is possible to relax the assumptions of the classic model corresponding to the serial

and cross-sectional correlation of the disturbances[16] and still have the convergence of the factors vector space and the loadings.

If the disturbances are weakly correlated across series and across time, Chamberlain and Rothschild showed that the variance covariance matrix can be decomposed into:

$$\Sigma_N = \Lambda'\Lambda + R_N$$

where R_N is a matrix with uniformly bounded eigenvalues and Σ_N has only k eigenvalues that are unbounded as N tends to infinity ($N \to \infty$).

The use of models with large N and T allows to use a different procedure of estimation than the traditional one achieved with maximum likelihood estimation which requires normality of the errors. In fact, Chamberlain and Rothschild (1983) showed that a principal components estimation, which is a much simpler estimation method when the number of parameters is large, provides a consistent estimators when N increases to infinity and the number of factors is much smaller than N.
In the same line of investigation, Connor and Korajczyk (1986, 1993) showed that when the number of series is higher than the number of dates, it is also possible to estimate the factors but in this case by referring to the eigenvectors of the TxT centered cross-product matrix (XX').

By centering and normalizing the series, the principal components are obtained as the eigenvectors of the correlation matrix of the initial matrix. We use the latter procedure in our application.

More recently, the asymptotic estimation within large panel settings has been completed by Bai and Ng (2007). We follow these authors but in our case, we use a standardized matrix Z calculated from X because we have a set of variables with very different standard deviation . The factors $\hat{F^k}$ and the loading $\hat{\Lambda^k}$ are estimated with a principal components method such that:

[16]Provided that they remain limited.

- $N > T$:
  - $\hat{F^k}$ is $\sqrt{T}$ times the eigenvector corresponding to the k largest eigenvector of the TxT matrix ZZ' and
  - $\hat{\Lambda^k} = T^{-1}\hat{F^k}'Z$
- $N < T$:
  - $\hat{\Lambda^k}$ is $\sqrt{N}$ times the k largest eigenvector of the NxN matrix $Z'Z$ and
  - $\hat{F^k} = N^{-1}Z\hat{\Lambda^k}$

Here, we analyze financial time series corresponding to returns of equity markets. We can use this kind of factor models as long as the stationarity assumption of the returns is respected. For equity returns, stationarity is almost always verified. Tests of stationarity are presented in Section 1.4.3.

The first estimated factor $\widehat{F_1}$ can be interpreted as an average of the variables of the data set. The loadings can be either positive or negative and reflect how correlated each variable is with each common factor. In our case, the data set is composed of country indexes or country-sector equity indexes that are positively related, therefore we expect that all the loadings associated with the first factor will be mainly positive. In a PCA, the factors are extracted in a sequential way, with the first factor explaining the largest part of variance in the data set, the second factor explaining the largest variation not explained by the first factor, and so on.

A more general specification of factor models allows some dynamics of the different elements of the model: the factors, the loadings or the idiosyncratic component. This means that the factors, loadings or disturbances are allowed to vary over time and that their dynamics can therefore be modeled. A common approach to specify the dynamics is to consider that the time varying elements of the model follow a stationary process with weak serial autocorrelation. Some authors have considered for example to model the factors as autoregressive processes of order p having the following form:

$$F_t = \gamma_1 F_{t-1} + ... + \gamma_p F_{t-p} + u_t \tag{1.10}$$

Thus, a general formulation of the *dynamic factor model* for this case is given by:

$$X_{it} = \alpha_i + \Gamma^p(L) F_t + \varepsilon_{it}$$

where $\Gamma^p(L)$ is a polynomial of order p in the lag operator L. In the same way the loadings, or the idiosyncratic term, can be modeled enriching the structure of the model but also increase the number of parameters to estimate.

Stock and Watson (1998) use a dynamic specification of this model to forecast macroeconomic time series. They have proved that the estimated common factors can be used to improve forecasting accuracy.

In this chapter, we make use of a static approximate factor model to analyze the comovements of equity returns. This choice can be justified by the low levels of serial correlation found in the original variables. For an illustration we have presented this feature in the subsection 1.4.1 where we present the descriptive statistic of the data set.

Factor models thus summarize the information contained at a given date in the set of returns within a few statistical variables that can be considered representative of the key features of the whole data set. Two main reasons justify the choice of factor models in the context of large-data sets environments. First, consistent estimates are possible to construct and the estimators of the factors converge to the true factor space as the number of variables and the number of observations increase. And second, inference is available including methods to determine the number of relevant factors and methods to build confidence intervals for the true factors useful to make comparisons of the factor with observable series.

1.3.2 Determination of the number of factors

Many theories have been developed to estimate the relevant number of factors in the context of factor models. Traditional methods to test the number of relevant factors in a multivariate model consist in analyzing different overall variability measures as the trace or the determinant of the covariance matrix as well as the eigenvalues of the covariance or correlation matrix. Trzcinka (1986) has tested a set of American firms, extracting the eigenvalues from the NxN covariance matrix and testing with a χ^2 test [17] the null hypothesis that the $k+1$ to N eigenvalues: $\lambda_{k+1}, ..., \lambda_n$ are equal. He found evidence for the presence of one factor and no evidence for k factors ($k > 1$). Brown (1989) derived explicit formulas for the rates of change of eigenvalues with N, he also performed some simulations to analyze the eigenvalues as N increased. He showed that all eigenvalues grow with N. Connor and Korajczyk (1993) developed a test statistic based on the idea that if k is the number of factors then there's no significant decrease in the average cross-sectional adjusted squared residual when the model is tested with one additional factor ($k+1$). They showed that this residual quantity diminishes as k goes to infinity. Chamberlain and Rothschild (1983) have shown that the $(N-k)$ eigenvalues diverge as N goes to infinity. A recent paper by Bai and Ng (2002) presents some panel based information criteria to determine the number of factors in the case of *approximate factor models*. Their method has some advantages over other studies because the results are still valid even in the presence of heteroscedasticity and weak autocorrelation and because, in the context of large data settings, the estimated number of factors converge to the true number of factors. Bai and Ng suggested that their method could be used in macroeconomics and finance. Many empirical studies have used it in macroeconomic applications but, to the date, there are fewer applications in finance.

In large N settings and using the *approximate factor model* framework, Bai and Ng have proposed to use some information criteria to consistently determine the number of factors

[17] $\chi^2 = -(T-1)\sum_{j=k+1}^{n} Ln(\lambda_j) + (T-1)Ln(r\sum_{j=k+1}^{n}\lambda_j)$ and the degrees of freedom $df = \frac{1}{2}(r+1)r$ where $r = n-k$

as N and T tend to infinity.

Based on a k-factor model, they have studied the overall sum of the estimated squared residuals divided by NxT:

$$V_k = \frac{1}{NT}\sum_{t=1}^{T} \hat{e}_t \hat{e}_t'$$

where $\hat{e}_t = Z_t - \widehat{\Lambda}\widehat{F}_t$ and have presented three different panel criteria $PC(k)$ as well as three different information criteria $IC(k)$, of the form:

$$PC(k) = V(k) + kg(N,T) \tag{1.11}$$

$$IC(k) = Ln(V(k)) + kh(N,T) \tag{1.12}$$

where $g(N,T)$ and $h(N,T)$[18] are penalty functions taking into account the effective number of observations NxT. The estimated number of common factors $\hat{k}$ as the number that minimizes the information criterium. Bai and Ng have analyzed different criteria and they showed that they can be asymptotically equivalent; but have different properties in finite samples. Therefore, they run a number of simulations considering different finite values of N and T and different specification including the cases of homoscedastic and heteroscedastic errors as well as a case where serial and cross-sectional correlation of the error terms are allowed. They compare the selected criteria that take into account both dimensions N and T to other conventional criteria used in time series applications such as AIC or BIC[19]. In the cases where the samples are small, they show that the set of $IC(k)$ criteria tends to underestimate the true number of factors and the set of $PC(k)$ criteria tends to overestimate it. They conclude that in the cases where the error terms are not independent, the first two information criteria $IC(k)$ - from the three different criteria presented in their paper - are the most robust. They note these criteria $IC_1(k)$ and $IC_2(k)$. In our application, we have a matrix of equity returns which present serial

[18]Function g depends on the variable σ^2 estimated as $(NT)^{-1}\sum\sum\varepsilon_{it}^2$ that is based on the choice of a maximum number of factors $kmax$, while the function h does not depend on $kmax$ which is an advantage of $IC(k)$ criteria in practical terms

[19]Akaike information criterion (AIC) and Bayesian information criterion (BIC) or Schwarz Criterion

and cross sectional correlation; therefore, we use these two information criteria $IC_1(k)$ and $IC_2(k)$:

$$IC_1(k) = \ln V(k) + k \log(\frac{NT}{N+T}) \frac{N+T}{NT} \quad (1.13)$$

$$IC_2(k) = \ln V(k) + k \log(\frac{\min(N,T)}{N+T}) \frac{N+T}{NT} \quad (1.14)$$

More precisely, we evaluate the number of factors for different data sets, and we combine these results with the measures of comovements.

1.3.3 Measuring the strength of comovements

We measure the comovements by referring to a factor structure with k factors. The intensity of the comovements at time t, are calculated from the estimated correlation matrix with W observations[20], as:

$$C_t^{1:k} = \frac{\sum_{i=1}^{k} l_{it}}{\sum_{i=1}^{N} l_{it}} \quad (1.15)$$

where l_i corresponds to the ith eigenvalue of the sample correlation matrix. This measure indicates the intensity of markets comovements and represents the part of variance explained by the k first common factors over the set of standardized variables. We can also define a measure accounting only for the comovements explained by one of those factors as:

$$C_t^k = \frac{l_{kt}}{\sum_{i=1}^{N} l_{it}} \quad (1.16)$$

After analyzing the contribution of the different number of factors, we observe that for the set of developed markets, the comovement measure C_t^1 which is based on the first factor is the only one having an important variation over time. We conclude that the main factor is responsible for fundamental comovements for the set of developed countries and

[20] from the date $(t - W)$ to the date t

therefore we focus on it for this set of markets for the rest of the analysis. However, for the set of emerging markets, there are several periods where the second and third factor drive a relevant portion of the comovements. For this reason, we consider that it should be meaningful to include more than one factor for these markets. In what follows, we note C_t^1 as C_t for simplicity.

It is worth noting that C_t is equivalent to the average squared correlation of each of the N variables with the estimated factor $\widehat{F_{1t}}$.

$$C_t = \frac{1}{N}\sum_{i=1}^{N} \widehat{r}^2(Z_{it}, \widehat{F}_{1t})$$

with

$$\widehat{r}(Z_{it}, \widehat{F}_{1t}) = \frac{\sum_{s=t-W}^{t} Z_{is}\widehat{F}_{1s}}{\sqrt{\sum_{s=t-W}^{t} Z_{is}^2 \sum_{s=t-W}^{t} \widehat{F}_{1s}^2}}$$

where W denotes the number of observations in the rolling window retained foe estimation and $\widehat{F_{1t}}$ is the first factor estimated from the correlation.

We construct confidence intervals for the eigenvalues of the correlation matrix to be able to compare different levels of C_t based on the asymptotic results of **?** for PCA with standardized variables. He has defined a confidence interval for the eigenvalues of the correlation matrix using the limiting distribution of $\sqrt{N}(log(\widehat{\lambda_i}) - log(\lambda_i))$ defined in Anderson (1963). The interval applied to the measure C_t^k is given by:

$$\hat{C_t^k} exp(-1.96\sqrt{\frac{2}{N-1}}) < C_t^k < \hat{C_t^k} exp(1.96\sqrt{\frac{2}{N-1}}) \tag{1.17}$$

Many authors have reported similar measures of comovements without analyzing them over time and without including confidence intervals allowing to compare the levels from different periods.

1.3.4 Interpretation of the factors and the measure of comovements

Economic Interpretation of the factors and financial integration

A disadvantage of factor analysis is the difficulty to interpret the statistical factors economically. Heston et al. (1995) showed that the main factor is highly correlated with an equally weighted index (EWI) composed of all N assets. It is a first indication to identify the factor as an observable variable. We pushed further this analysis using the method of Bai and Ng (2006) to build a confidence interval around the true factor and to evaluate if an observed factor lies within its bounds[21]. The resulting interval for an observed variable G has the following form:

$$\widehat{G} + \widehat{\varepsilon}_{jt} - 1.96\widehat{s}_{jt} < G < \widehat{G} + \widehat{\varepsilon}_{jt} + 1.96\widehat{s}_{jt} \tag{1.18}$$

Where $\widehat{\varepsilon_{jt}}$ is the error term from the regression of the observed variable G on the estimated factors $\widehat{F}$ and $\widehat{s}_{jt}$ is the asymptotic variance of this error term[22].

If the EWI lies within the confidence interval for the first true factor, we can conclude that the first factor represents a market factor which is close to the EWI.

If one can conclude that the first factor represents a market factor, C_t can be used as a measure of integration by referring to studies of integration based on an asset pricing model. We used two different approaches to measure integration. First, we analyze different regions and compare the values of the associated C_t^{F1R} measures. Second, we measure integration to global markets[23]. We calculate the measure of comovements explained by the main factor for each region: a region with a higher value of C_t is considered more integrated. We compare it with the measure of comovements based on the first factor extracted from the whole set of variables, noted C_t^{F1G}. We make the comparison for two

[21] The details of the method are found in the appendix.

[22] The details of this calculation are found in the appendix of this chapter

[23] This approach used in other integration studies is similar to international asset pricing models (i.e. CAPM) based on an observed global factor.

groups, emerging and developed markets. We compare the importance of the global factor compared to a regional factor. Here, we have to note that for several calculations, we don't have a very large number of variables and therefore, the asymptotic results presented in in Section 1.3.1 are not valid. However, the finite statistical sense of the measures remains the same.

Interpretation of the measure of comovements in terms of diversification

The idea behind this kind of factor models is that the returns are supposed to share a systematic (market) risk represented by a limited number of factors. The risk corresponding to the idiosyncratic components is diversifiable which means that for a relevant composition of the portfolio, this risk can be eliminated.

In the same way that the variance covariance matrix of the returns was decomposed, the variance of the return of a portfolio $\alpha = Vec\{\alpha_k\}_{1\leq k\leq N}$ can be decomposed into:

$$Var(\sum_{k=1}^{N}\alpha_k X_k) = \alpha'\Lambda'\Lambda\alpha + \sum_{k=1}^{N}\alpha_k^2 Var(u_k)$$

If the quantity $\sum_{k=1}^{N}\alpha_k^2$ tends to zero when N tends to infinity, one can prove that the idiosyncratic risk also tends to zero, adopting relevant hypothesis to exclude any arbitrage opportunity. Thus, the so-called diversifiable part of the risk $u^{(N)}$ is eliminated by diversification and the term $\sum_{i=1}^{N}\alpha_i^2$ can be interpreted as a measure of the diversification opportunities offered by the portfolio[24].

There exist sufficient conditions for the Data Generating Process (DGP) obeying to a weak k-factor structure[25]. The problem with such conditions is that they are difficult to test, because one has to look for maxima over N, whose value is of course fixed in practice.

[24]This measure is proved to be a function of $\overline{\lambda}$ =$\sup_N \lambda_u^{(N)} < \infty$ (Chamberlain (1983)). However, as presented in Solnik (1974), when the number of assets in a portfolio attains a given level, the reduction of the idiosyncratic risk is negligible

[25]Note that a sufficient condition for the DGP obeying to a weak k-factor structure is given by:

$$\sup_N \lambda_k^{(N)} \to \infty,\ \sup_N \lambda_{k+1}^{(N)} < \infty\ \inf_N \lambda_N^{(N)} > 0$$

As the level of comovements arises, the part of variance explained by XXXX and the idiosyncratic variance decreases accordingly. Less opportunities of diversification are therefore left for the investor. Comovements are therefore inversely related to diversification. Solnik (1974) has introduced a measure capturing the part of risk that cannot be reduced by diversification, defined as the ratio of the variance of a portfolio of n assets over the average variance of a typical security[26]. Thus, the diversification ratio DR_t, for a set of variables x_{it} representing stock or index returns, is defined as:

$$DR_t(n) = \frac{Var(\sum_{i=1}^{n} \frac{x_{it}}{n})}{\frac{1}{n}\sum_{i=1}^{n} Var(x_{it})} \tag{1.19}$$

We give empirical evidence that DR_t lies within the confidence intervals of C_t confirming that C_t is closely related to this measure DR_t.

if the eigenvalues of the matrix $Var(X^{(N)})$ for a given N, are ordered according to:

$$\lambda_1^{(N)} > \lambda_2^{(N)} > ... > \lambda_N^{(N)}$$

[26]The average variance of a single stock computed with all the available stocks or assets.

1.4 Description of the Data

Sections 1.4 and 1.5 concern the empirical part of this chapter. This section presents a static analysis of the data and the next section focuses on a dynamic analysis based on rolling estimations. In this section, we present the main features of the different variables used in our study. The first subsection presents the different series focusing on the geographical and sectoral distribution. The second and third parts of this section present descriptive statistics and a factor decomposition of the country based indexes.

1.4.1 Description of the data base

We have used Datastream[27] total return country-sector indexes in US Dollars for 10 sectors and 46 countries[28]. The reasons for using these indexes is that Datastream uses the same methodology for their calculation; moreover, the country indexes include at least 80% of the capitalization of each market and there is a large coverage on emerging markets. The period of study is 1994:07-2006:06. The indices are weighted by capitalization. We have decomposed the set of countries into developed and emerging markets and distinguished three different regions within emerging markets[29]: Asia, Latina America and EMEA region (Europe Middle East and Africa). We present the details of the different countries, regions and sectors included in the data set in Table 1.1 and a basic description of the sectors in Table 1.2.

We have studied the returns of indexes rather than stocks, but the present study can be easily extended to individual stocks, which would be more useful from a practical point of view and would be more consistent with the assumptions required by the methodology, namely the high number of required variables. The difficulty in this case would be to

[27]Datastream International. Datastream International Ltd., Data retrieved in September, 2007.

[28]The 46 countries were selected from a set of 50 country indexes available in the Datastream database, we have excluded those countries having missing data or long series of consecutive equal prices. An alternative could have been to replace the missing information by their last known data or a related average but we should have introduced undesired spurious correlations.

[29]according to the classification given by Datastream

handle a very large data base[30].

Table 1.1: Countries and Sectors

COUNTRY	CODE	TYPE	REGION	BASIC	CYCGD	CYSER	GENIN	ITECH	NCYCG	NCYSR	RESOR	TOTLF	UTILS	NB	NB-R	NB-T
CANADA	CN	DM	AM	1	1	1	1	1	1	1	1	1	1	10	20	204
U.S.A	US	DM	AM	1	1	1	1	1	1	1	1	1	1	10		
AUSTRALIA	AU	DM	AS	1	1	1	1		1	1	1	1	1	9	47	
HONG.KONG	HK	DM	AS	1	1	1	1	1	1	1	1	1	1	10		
JAPAN	JP	DM	AS	1	1	1	1	1	1	1	1	1	1	10		
NEW.ZEALAND	NZ	DM	AS	1	1	1	1		1	1	1	1	1	9		
SINGAPORE	SG	DM	AS	1	1	1	1	1	1	1	1	1		9		
DENMARK	DK	DM	EUP	1	1	1	1		1	1		1	1	8	44	
NORWAY	NW	DM	EUP	1	1	1	1	1	1		1	1	1	9		
SWEDEN	SD	DM	EUP	1	1	1	1	1	1	1		1		8		
SWITZERLAND	SW	DM	EUP	1	1	1	1	1	1	1		1	1	9		
UNITED.KINGDOM	UK	DM	EUP	1	1	1	1	1	1	1	1	1	1	10		
AUSTRIA	OE	DM	EUR	1		1	1		1	1	1	1	1	8	93	
BELGIUM	BG	DM	EUR	1		1	1	1	1	1		1	1	8		
FINLAND	FN	DM	EUR	1	1	1	1	1	1	1		1	1	9		
FRANCE	FR	DM	EUR	1	1	1	1	1	1	1	1	1		9		
GERMANY	BD	DM	EUR	1	1	1	1	1	1	1		1	1	9		
GREECE	GR	DM	EUR	1	1	1	1	1	1		1	1		8		
IRELAND	IR	DM	EUR	1	1	1	1		1		1	1		7		
ITALY	IT	DM	EUR	1	1	1	1	1	1	1	1	1	1	10		
NETHERLANDS	NL	DM	EUR	1	1	1	1	1	1	1	1	1		9		
PORTUGAL	PT	DM	EUR	1	1	1	1		1	1		1		7		
SPAIN	ES	DM	EUR	1	1	1	1		1	1	1	1	1	9		
CHINA	CH	EM	EAS	1	1	1	1		1		1	1		7	72	172
INDIA	IN	EM	EAS	1	1	1	1	1	1	1	1	1	1	10		
INDONESIA	ID	EM	EAS				1		1	1	1			4		
KOREA	KO	EM	EAS	1	1	1	1		1	1	1	1	1	9		
MALAYSIA	MY	EM	EAS	1	1	1	1		1	1	1	1	1	9		
PAKISTAN	PK	EM	EAS	1	1	1	1		1	1	1	1	1	9		
PHILIPPINES	PH	EM	EAS	1		1	1		1	1	1	1	1	8		
TAIWAN	TA	EM	EAS	1	1	1	1	1	1			1		7		
THAILAND	TH	EM	EAS	1		1	1	1	1	1	1	1	1	9		
CZECH.REPUBLIC	CZ	EM	EEMEA	1	1	1	1	1	1	1	1	1	1	10	51	
HUNGARY	HN	EM	EEMEA	1	1	1			1	1		1	1	7		
ISRAEL	IS	EM	EEMEA	1			1	1	1	1	1	1		7		
POLAND	PO	EM	EEMEA	1					1			1		3		
RUSSIA	RS	EM	EEMEA	1					1	1	1	1	1	6		
SOUTH.AFRICA	SA	EM	EEMEA	1	1	1	1		1	1	1	1		8		
TURKEY	TK	EM	EEMEA	1	1	1	1	1	1	1	1	1	1	10		
ARGENTINA	AR	EM	LA	1	1	1	1		1	1	1	1	1	9	49	
BRAZIL	BR	EM	LA	1			1		1	1	1	1	1	7		
CHILE	CL	EM	LA	1		1	1		1	1	1	1	1	8		
COLOMBIA	CB	EM	LA	1		1			1	1		1	1	6		
MEXICO	MX	EM	LA	1	1	1	1		1	1	1	1		8		
PERU	PE	EM	LA	1	1	1	1		1	1	1	1		8		
VENEZUELA	VE	EM	LA	1								1	1	3		
TOTAL														376		

[30]The goal of our study being to understand the structure of equity returns from a global perspective, the number of country-sector series is enough to meet our goals

Table 1.2: Sector codes and names

Sector Code	**Sector Name**	
BASIC	Basic Industries	(Chemicals, construction)
CYCGD	Cyclical Consumer Goods	(Automobiles, Household, Textiles etc)
CYSER	Cyclical Service	(Leisure, media, transportation)
GENIN	General Industrials	(Aerospace, Industrials, machinery)
ITECH	Information Technology	(Hardware, Software)
NCYCG	Non Cyclical Consumer Goods	(Food, Beverage etc)
NCYSR	Non Cyclical Services	(Food retailers, telecom services)
RESOR	Resources	(Mining, Oil and Gas)
TOTLF	Financials	(Banks, Insurance, Investments)
UTILS	Utilities	(Electricity, Gas, Water)

1.4.2 Descriptive statistics

Log returns were calculated from the price series[31] in US Dollars at weekly horizons. We have analyzed some descriptive statistics for the returns of the set of country-sector indexes. The results are summarized by regions in Table 1.3.

Table 1.3: DS

	Mean	**Std**	**Skew**	**Kurt**	**AC1**	**AC2**	**AC3**	**AC4**	**ADF**
AM	10.52%	16.11%	- 0.340	1.532	- 0.026	0.003	0.051	0.028	- 15.087
AS	3.46%	19.42%	- 0.290	1.301	0.040	0.044	0.038	- 0.012	- 15.364
EUP	9.36%	19.70%	- 0.520	2.353	- 0.059	0.033	0.042	0.004	- 17.135
DEVELOPED	8.18%	19.32%	- 0.454	2.052	- 0.035	0.033	0.042	0.003	- 16.572
EAS	0.56%	32.06%	- 0.278	3.189	0.027	0.076	0.068	- 0.006	- 12.393
EEU	13.09%	35.67%	- 0.286	1.673	0.025	0.065	0.094	- 0.018	- 15.424
LA	2.67%	28.89%	- 0.355	3.942	0.030	0.064	0.056	0.018	- 14.135
Others	7.85%	23.61%	- 0.632	2.068	0.005	0.065	0.050	0.003	- 18.236
EMERGING	6.12%	27.49%	- 0.390	2.817	0.000	0.058	0.061	0.000	- 14.773

For the set of developed countries, the average return is 8.2% and the average annualized volatility is 19.3%. The figures for the set of emerging markets present lower levels of expected returns for the period considered and higher volatility[32]. The average return for these markets is 6.1% and the average annualized volatility is 27.5%.

[31] $R_{t,t+h} = ln(P_{t+h}/P_t)$

[32] These features make that emerging markets are not very attractive during the sample analyzed as their are dominated on average by developed ones in risk and return terms

To get a better understanding of our data set, we have performed a factor decomposition of the set of country indexes.

Table 1 -A in the appendix of this chapter presents detailed descriptive statistics for the different country based indexes returns. We observe that annualized average returns in US Dollars for the different country indexes vary from -1% for Japan to 13% for Finland for the group of developed markets and from -6.5% for Thailand to 27% for Russia for the emerging ones. The average return for emerging markets is 6.1% while for developed markets it is close to 8.2%. Hungary, Russia and Turkey have average returns above 10%, and countries such as Argentina, Thailand, Philippines and Malaysia have average returns that are lower than -2% in US Dollars for this period. The only developed country presenting negative average return for the period is Japan with -1.3%. The risk figures present also a clear difference between emerging and developed markets. Volatilities range from 15% to 50%, the average volatility for emerging markets is 31% and for developed markets is 19%. Countries such as Turkey or Russia have annual volatilities of more than 40% and countries as Austria, Australia or United Kingdom have annual volatilities of only 15%. We have also reported skewness and kurtosis coefficients. The skewness coefficients present dissimilar patterns for developed and emerging markets. Most developed markets are skewed to left, the average skewness coefficient for these countries is -0.45. And most emerging markets have also negative coefficients of skewness but slightly closed to zero, the extreme cases are South Africa having a very negative coefficient and Taiwan a slightly positive coefficient. The figures for kurtosis coefficients do not present a marked difference for developed and emerging markets. However the coefficients are higher for emerging than developed markets. These descriptive statistics give us important information about the distribution of returns for the two sets of countries. The data present some positive autocorrelation for emerging markets at the second and third lags. All the series of returns are stationary, according to the very negative values of the ADF statistic, this point is very important because it is necessary to have stationary series in order to be able to use

factor models.

The goal of this chapter is to analyze the comovements between different international markets. A good start is to analyze the correlation coefficients. We present correlations across zones and by samples. For the complete sample we observe that developed countries are more correlated than emerging ones indicating a higher tendency towards integration. We have also decomposed the data set into two subperiods, before and after 2000. Many authors have presented analysis of correlations for developed countries[33]. We analyze those same figures finding similar results. The average correlation of the US with other developed markets is 0.4 before 2000 and 0.53 after this date. We have also studied correlations for emerging countries for which the levels are significantly lower, close to 0.25 on average with maximum correlation close to 0.5 and minimum of approximately 0. Within the emerging group we highlight the strong correlations for three different groups of countries: first the Latin American countries (in particular Brazil, Mexico and Chile over 0.5 of average correlation within the group), second the south Asian countries (for Thailand, Malaysia, Philippines 0.45) and third the emerging European countries (Hungary, Poland and Czech Republic with 0.45). Within the developed markets, the group of European countries, specially those making part to the Euro zone, present correlations of above 0.75. We also observe a high correlation for USA and Canada, and looking at the lowest values we can see that Japan is the country less correlated with every other country in the sample (0.31 on average). When we look at the correlation between emerging and developed countries, we observe that Singapore is highly correlated with Malaysia and Thailand as they belong to the same geographical area. And we have also noticed that some countries such as Pakistan, Colombia or Venezuela are poorly correlated with all the other developed

[33]Goetzmann et al. (2005) find that the average correlation among 4 main markets is 0.475 for the period 1972-2000. Longin and Solnik (1995) present a graphical analysis of average correlation of the US market with other seven developed markets showing that their level varies from 0.3 to 0.75 for the period 1959-1991, they show correlation close to 0.45 for the first years and close to 0.65 from the last years. Brooks and Negro (2003) have declared that correlation of the U.S. with developed markets has risen from a level close to 0.4 from the mid-1980s through the mid-1990s to 0.9 approximately for the more recent period.We do not find levels of correlation so high as Brooks and Negro but we have also observed an increase across time.

countries. Tables 3 -D and 3 -B in the appendix summarize these results.

To see if there is a change in correlations through time, we have plotted correlations in a graphical way where darker colors represent stronger correlation, Figure 1.1 presents correlation for the whole sample and Figure 1.2 presents the results for two different subsamples: one with the first half of the observations (first period) and the second one with the rest. In the following figures, the set of emerging markets are denoted by X and the set of developed markets by Y. The cross correlation graphic combines the two sets of markets. The different colors indicate the levels of correlation. In each figure we can see the difference in the levels of correlation within the group of emerging markets and developed markets. The set of developed markets, in red color indicating correlations of 0.5 or more is highly correlated compared to the set of Emerging markets, in green indicating levels of correlation close to 0. (See Figure 1.1).

Figure 1.1: Cross Correlation Emerging and Developed Markets

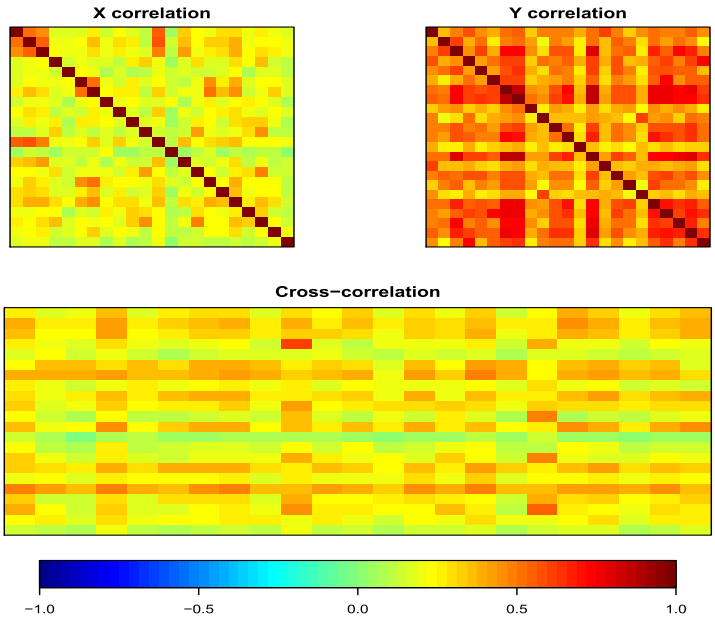
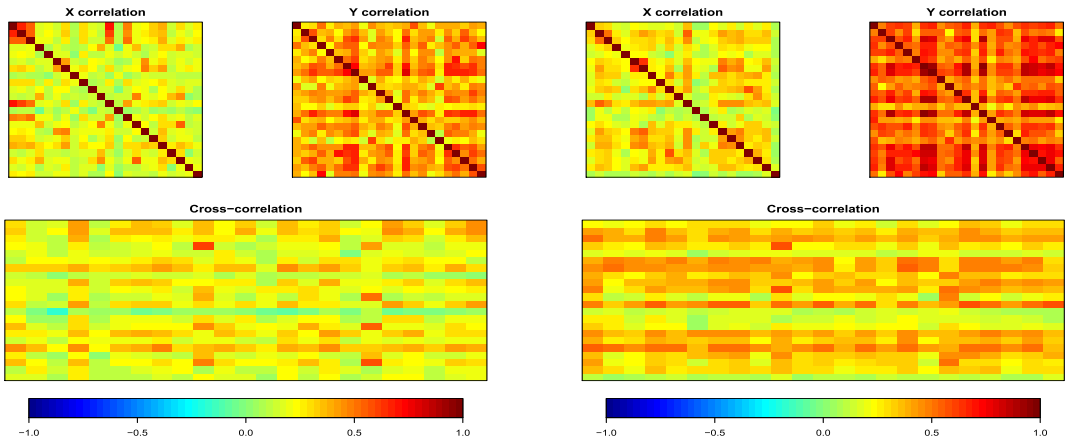


Figure 1.2: Cross Correlation Emerging and Developed Markets - Subsamples: 1994:200 - 2000-2006



Analyzing Figure 1.2, we observe that the correlations for developed and emerging countries have increased significantly from the first period to the second, the levels of correlation for developed markets in the second period are close to 0.6 with more than 10% of the pairwise correlation being higher than 0.75, based on averages calculated from the tables in the appendix.

1.4.3 Stationarity

We have also studied the stationarity of the countries time series returns. As described in Section 1.3.1, for the class of factor models that we have chosen, it is necessary to use stationary series. The results are found in Table 1.4. We observe that every country index return series is stationary.

Table 1.4: ADF Tests for developed and emerging markets

	ADF			ADF	
DM	**Levels**	**Returns**	**EM**	**Levels**	**Returns**
AU	1,819	- 23,601	AR	- 0,849	- 23,072
OE	3,274	- 23,547	BR	0,388	- 12,383
BG	1,482	- 17,051	CL	0,515	- 10,655
CN	2,302	- 22,799	CH	0,350	- 10,595
DK	2,010	- 24,601	CB	- 0,511	- 25,626
FN	- 0,454	- 7,993	CZ	2,108	- 10,578
FR	1,181	- 17,584	HN	0,775	- 11,632
BD	0,485	- 13,344	IN	0,446	- 6,157
GR	0,069	- 23,751	IS	1,117	- 24,755
HK	- 0,087	- 10,339	KO	0,306	- 13,055
IR	1,333	- 24,070	MY	- 1,091	- 5,184
IT	0,637	- 24,417	MX	0,398	- 11,178
JP	- 0,873	- 25,055	PK	- 0,448	- 12,594
NL	0,495	- 10,150	PE	0,762	- 6,078
NZ	1,444	- 9,942	PH	- 1,534	- 10,010
NW	1,522	- 22,998	PO	0,149	- 23,640
PT	0,508	- 9,206	RS	1,477	- 6,960
SG	- 0,366	- 5,721	SA	0,793	- 12,166
ES	1,626	- 13,530	TA	- 0,582	- 23,309
SD	0,626	- 17,459	TH	- 1,686	- 8,091
SW	0,685	- 15,956	TK	- 0,325	- 24,321
UK	0,996	- 17,546	VE	- 0,884	- 15,910
US	0,823	- 6,192			

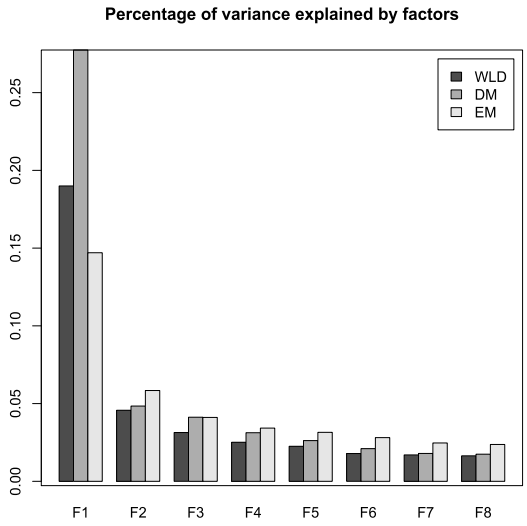
1.4.4 Factor decomposition

A broad decomposition of returns into common factors is performed with a PCA on the whole set of variables (376 series) as well as on each type of market (emerging or developed) for the complete set of observations. This decomposition gives a first look into the structure of returns. We decomposed the data into 8 factors, with the number of factors arbitrarily chosen. We observe in Figure 1.3 that, on average for the period of study, the first factor F1 accounts for 18% of the total variance of the complete data set[34], 28% of the variance of a set of country sector indexes from developed countries and only 15% for the same sample for emerging countries. We also observe from this figure that

[34]All world countries are included and this set is denoted as WLD in the figure.

the second factors only accounts for approximately 5% of the variance of every sample of countries and the subsequent factors account for even less than 5%.

Figure 1.3: Part of variance explained by main factors



Now, we focus on the set of returns of country indexes (46 series). To get a more detailed information, we have decomposed the standardized returns into principal components and have analyzed the values of the loadings associated to each factor, for each country. The loadings are represented in four different bidimensional plots in figure 1.4. The first plot represents the loadings corresponding to factor F1 and F2. Developed countries are represented in red and emerging countries in green. The same information is also contained in Table 1.7 in the Appendix. In that table, the first row of the table contains the total percentage of variance explained by each factor. The first factor F1 accounts for 36% of the total variance if the data set[35]. The second factor F2 only explains 7.1%, F3 explains 6.9 % and the following factors less than 6%. The rest of the table contains the loadings for each country, the table is sorted according to the exposures to the first factor F1.

[35]This data set contains 46 country indexes

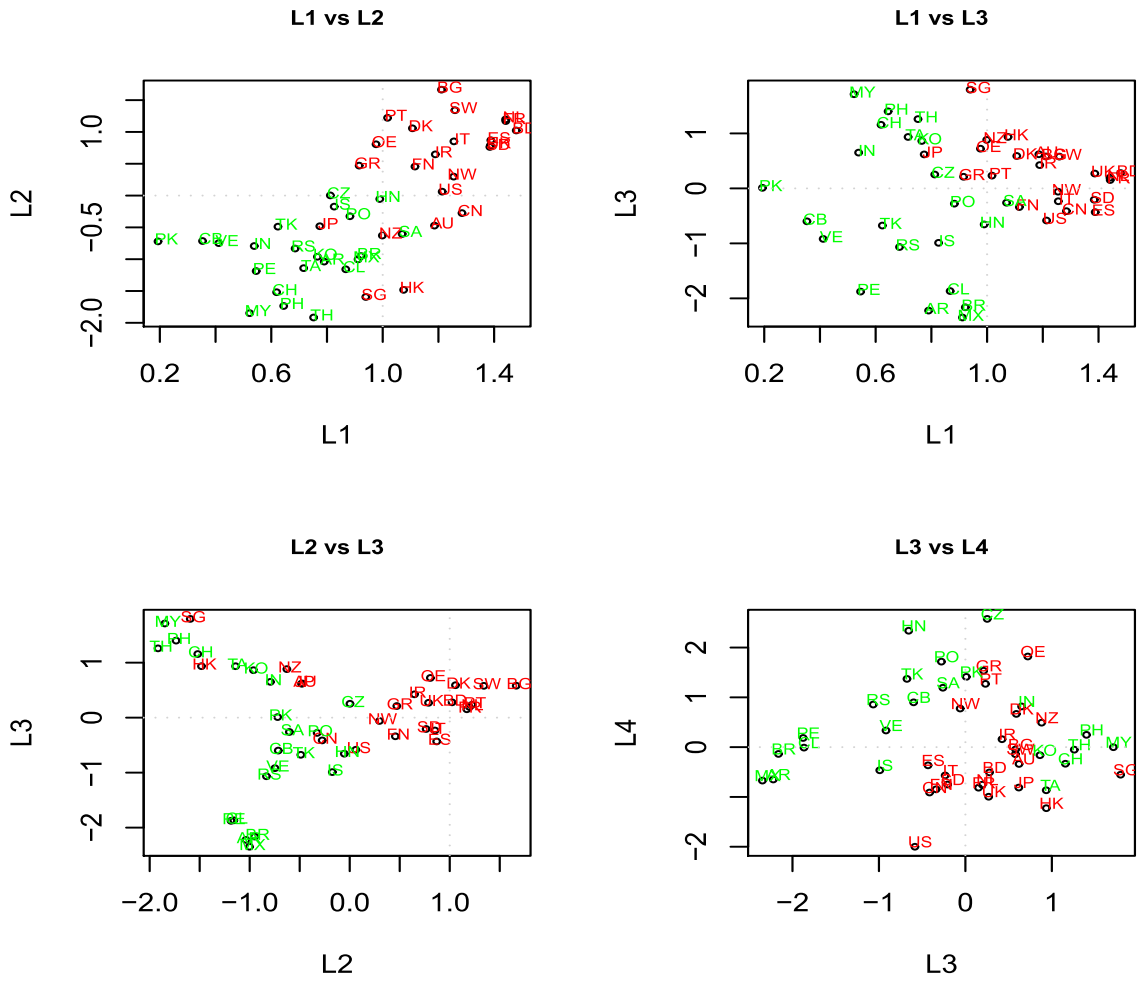
We can observe in Figure 1.4 that the loadings corresponding to this first factor are positive for all countries which means that there is no country index that behaves contrary to the global market. The majority of developed countries (in red) have loadings associated to F1 higher than one (the exceptions are Japan, Singapore and Greece) and countries having loadings lower than one are mainly from the emerging region (in green)[36]. The second factor also affects in a different way emerging and developed markets. For the set of developed countries the loadings are mainly positive with a few exceptions (For example Canada, Australia and some Asian countries). For the set of emerging countries almost all the loadings are negative. And for the set of European countries the loadings are close to one. Looking closer at the loadings we observe that the most negative values correspond to Asian countries and the higher loading correspond to European countries. The second factor therefore allows to separate the different countries in geographical regions, mainly Asia and Europe. The geographical difference can be explained by different economic episodes having affected each region. We can think that the second factor is a factor related to an Asian particularity, probably the Asian crisis, that could have affected other emerging countries. Indeed, when we run a decomposition of returns with a PCA, the first factor is chosen in a way so that it explains the higher part of the variability of the data set. Therefore, some events affecting mainly a group of countries but with very large variability can be captured by the common factors. The loadings corresponding to the third factor are close to 1 for the majority of Asian countries and close to -2 for the set of Latin American countries. The loadings corresponding to the fourth factor give less clear information, however, the loadings for the US are very negative and the loading for a set of eastern European countries (Hungary, Czech Republic, Poland and Austria) are positive and close to 2, this information could have an economic meaning but we are unable to identify it.

The use of factor models in finance is very common and the first factor commonly rep-

[36]The reader can reefer to Table 1.1 to find the codes for each country as we have used country codes instead of names in Figure 1.4 for purposes of clarity.

resents a global market which means that every asset follows in some way the movements of this global market. In has been largely debated in the literature if other factors have to be included or not and the studies including several factors debate on which could be the best variables to explain the returns. Our procedure is different in the sense that we do not choose a set of economic variables to explain the returns but we compute statistically uncorrelated factors that best describe their variability.

Figure 1.4: Loadings from Principal Components analysis on countries indexes returns



In section 1.5.2, we will show that the first factor is, indeed, a global market factor. The question remains open for the subsequent factors. The famous work by Fama and

French (1992) suggests to use three factors namely a market factor, a size factor and a value factor.

This section has presented the data set that we use for our analysis. We analyzed standard descriptive statistics as well as stationarity tests for the different country series. We observed the basic descriptive statistics are different for emerging and developed markets. From the analysis of correlation, we concluded that for both type of markets, correlation increases over time and that there is a significant difference in the levels of correlation for developed or emerging markets. Finally, we decomposed the set of returns into orthogonal factors and we analyzed the loadings corresponding to the main factors. From this analysis, we concluded that the structure of returns is also different for the two types of markets. In the next section we will study the importance of each factor for each type of market.

1.5 Main Results

In this section we present the results corresponding to the application of the model presented in section 1.3 using the complete data set composed of weekly returns of 376 country-sector indexes. We implemented a rolling estimation procedure in order to capture the changes in the comovements. The size of the window in the rolling estimations has to be large enough to include a sufficient number of observations so as to refer to the asymptotical results by Bai and Ng. Using rolling windows also ensures that only the most recent information is used in each estimation. At each date, we run a new calculation and, this way, we can observe the dynamic behavior of comovements. We used different window lengths varying from one to five years but we report principally the results for the windows of one year length as we can have more details on the dynamics of comovements. The series were normalized before estimating the factors, as we have series with very different levels of variance.

The structure of this section is the following. In the first subsection, we analyze the levels of comovements and the number of factors, we show the dynamics of comovements during the period of study and we show that for the different types of markets there is an upward trend in the level of comovements. Second, based on the results of the first subsection, we analyze the levels of comovements as a measure of integration, we show that with any of the different approaches we focus on, we arrive to the conclusion that integration has been rising during the last years. Finally, in the last subsection, we analyze the measure of comovements as linked to the diversification ratio presented in section 1.3.4 and we show that this two measures are closely related.

1.5.1 Factor Model and Comovements

Rolling estimations and window size determination

We start by analyzing the number of common factors within each data set with Bai and Ng's methodology presented in section 1.3.2. The first question we try to solve is what should be the correct size for the rolling windows. It is worth noting that Bai and Ng suggest to use at least 40 observations or 40 variables to have consistent estimators. Even if the theory suggests to use a large number of observations, the inclusion of more data would make the task of explaining the returns with common factors more difficult as it may imply a higher heterogeneity. We decide to include the maximum available information at the cross-sectional level and to reduce the windows to one or two years in order to have the maximum information about the changing behavior of comovements and factor structure over time. In order to give an answer to the question of the window size, we estimated the number of factors with the two criteria presented in equations 1.13 and 1.14 for different window's sizes as well as for different groups of markets. Both criteria give similar results, however, we present and analyze only the results corresponding to IC2 criterium which provides are more stable estimates. We calculate the number of factors considering the whole data set and also the average number of factors for rolling estimations with window sizes of one and two years [37]. We observe that the number of factors is different for each zone, it ranges from 4 to 6 if we use the whole sample of data and from 1 to 4 for the rolling estimations with a window length of one year[38]. The results are presented in table 1.5:

[37] 50 and 100 observations for our data set of weekly observations.
[38] This number of factors corresponds to the average number over the different windows.

Table 1.5: Number of factors for different regions and different sample size

	ALL	**DM**	**EM**	**EU**	**EUP**	**EEU**	**LA**
50	2	2	1	2	2	4	3
100	3	3	2	3	3	5	4
T	5	6	6	6	4	5	4

ALL: All countries; DM: Developed Markets; EM: Emerging Markets; EU: Europe ; EUP: Euro countries; EEU:Emerging Europe; LA: Latin America

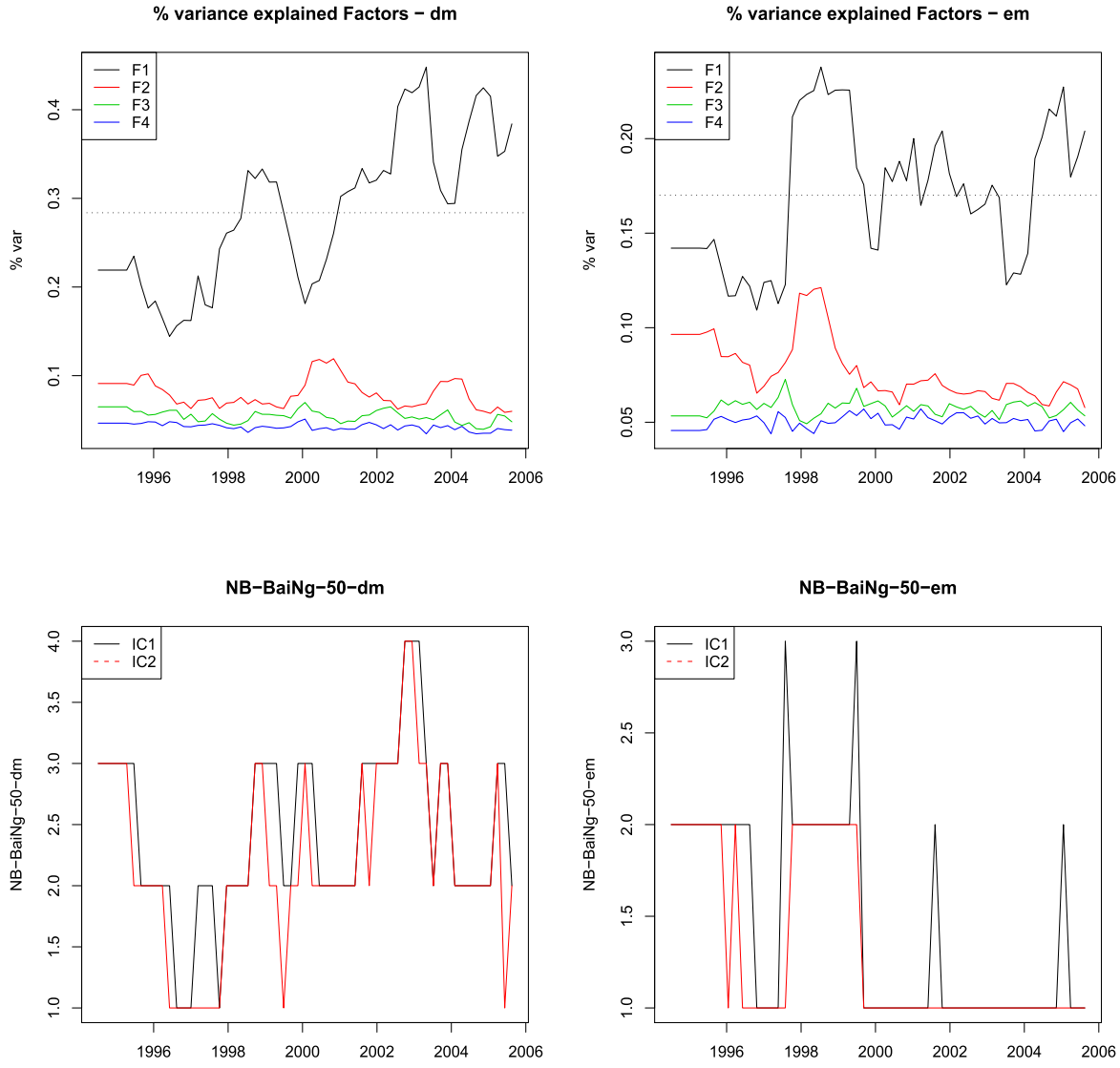
Several information from this table are worth to be highlighted. First, the estimated number of factors calculated with the maximum number of observations[39] is equal to 6 for each type of market. Second, the number of factors increases with the length of the window. These two points conducts us to the following analysis. We can think that, as the number of observations increases, the number of different events affecting a large number of markets can also increase leading to a larger number of factors. However, even after including a very large number of observations (here, the whole sample), the criteria suggest to include a limited and small number of factors. Third, for every sample, we observe a higher number of common factors for the sets of developed markets than for the emerging ones. This last observation leads us to question of whether the number of factors should be lower for a set of highly correlated markets than for a set with lower correlations. This question could be extended to whether we expect that during periods of high correlations, the number of factors would be reduced compared to periods of higher correlations. No answer to these questions have been found in the literature, therefore, we expect to find answers from our empirical analysis. Considering that larger windows would smooth too much the curve representing the comovements and would not allow to analyze the changes in the comovements, and considering that very short windows can produce unstable estimations, we decide to use windows of one and two years of length.

[39] 590 weekly observations.

Measure of comovements and number of factors

The different criteria presented by Bai and Ng are based on the maximization of the percentage of variance explained by a given number of factors taking into account a penalty function depending on N and T. Therefore it would be sound to jointly analyze the number of factors and the measures C_t. We analyze the percentage of variance explained by different number of factors, for emerging and developed data sets separately. Figure 1.5 presents these results.

Figure 1.5: Measures of comovements and number of factors for developed and emerging markets



First, we analyze the behavior of the measure C_t calculated with different number of factors. For the set of developed markets[40] presented in the left side of the figure, we have plotted the different measures of comovements with different number of factors in the plot of the top and the number of factors calculated with the method of Bai and Ng in the lower plot. From the first plot it worth noting that the measure C_t^1 has a level of

[40] noted as *dm* in Figure 1.5.

0.3 on average which means that nearly 30% of the variance of the whole standardized series is explained by the first estimated factor $\widehat{F1}$[41]. We observe an increase from a level close to 20% at the earlier period of study to a level close to 40% at the end. There is a positive trend in the comovements measure for both types of markets. We analyze this trends later in this chapter.

The comovements calculated with the other factors have a level close to 7% on average. Those corresponding to the second factor, C_t^2, are between 7% and 12%, C_t^3 have a level of 5% and C_t^4 less than 4%. The figure corresponding to the set of emerging markets presents similar features but lower levels of comovements. C_t^1 is on average 16% varying from 12% to 24%. The average level for C_t^2 is 7% and for C_t^3 or C_t^4 it is close to 5%. Looking closer to Figure 1.5, we observe different shapes for the comovements measure based on the first and second factors, in emerging or developed markets. The cumulative variance explained by the factors should explain the whole variability of the data set when all the factors are included. Consequently, if the measure C_t^1 increases due to a more similar behavior of the different series, automatically the part explained by the rest of factors would be lower. This explains why for developed markets, during some short periods, the measure C_t^1 drops and C_t^2 (and even C_t^3) rise accordingly. Another explanation for the presence of the second or third factors should be the temporary different behavior of a subset of markets or the presence of specific events affecting a large number of variables for a short period of time. Again, an increase in C_t^2 would make C_t^1 looks less important. We relate the peaks of C_t^1 with downturns of the market and, in the same way, comovements' drops should correspond to bullish phases of the market[42]. For the set of developed markets, we observe three peaks (marked in the dates axis by the years 1999, 2002-2003 and 2005) and three minimums (in 1996, 2000 and 2004). Consequently, the years peak in 1999 corresponds to years of high comovements or a crisis period in 1997-1998[43], we observe for

[41] The average level of variance explained by the common factors is represented in figure 1.5 by a dotted horizontal line.

[42] When investors are more confident and there is less variability in the equity market

[43] We recall that our measures are rolling over a window of one year, therefore a peak of comovements

those years the presence of a relevant second factor. During those years the Asian region has suffered one of the most dramatic financial crisis of its history. Therefore, the peaks in the comovements, observed for both types of markets for these dates, are certainly related to the Asian crisis with a larger and faster impact on emerging countries. In fact, for the set of Emerging Markets we observe a peak in the comovements during this period which is also the maximum value of comovements for the whole period of analysis and a few months later, a similar feature for Developed Markets is observed but it does not represent the maximum level in this case. Therefore, we can infer that there was an impact from this crisis on developed markets. In the same way, at the year 2000-2001 the comovements drop to a minimum level for developed countries and the second factor reaches one of its top levels, these observations are undoubtedly associated to specific factors related to the technological bubble. In the end, the level of comovements decomposed by factors, reveal different phases of the market.

The second row of graphics in figure 1.5 presents the optimal number of factors calculated for the same windows with the method of Bai and Ng. The combined analysis of the number of factors with the results from the measure C_t^k gives important results. We recall that their criteria chooses k so that the maximum percentage of variance is explained. During periods where the second or third factor account for an important part of the variance, the criteria suggest to use two or three factors, otherwise only one factor should be relevant. If the second and third factors reveal specific phases of the market, it is therefore important to include them, but only during some short periods. We observe a lower level in the comovements of emerging markets but also a lower number of factors according to the applied criteria. For developed markets, the number of factors is very unstable and difficult to interpret economically, probably because the episode of the IT bubble has an impact on the larger number of factors. An important result from this analysis is that the number of factors indicated by these criteria is a number close to 2 or

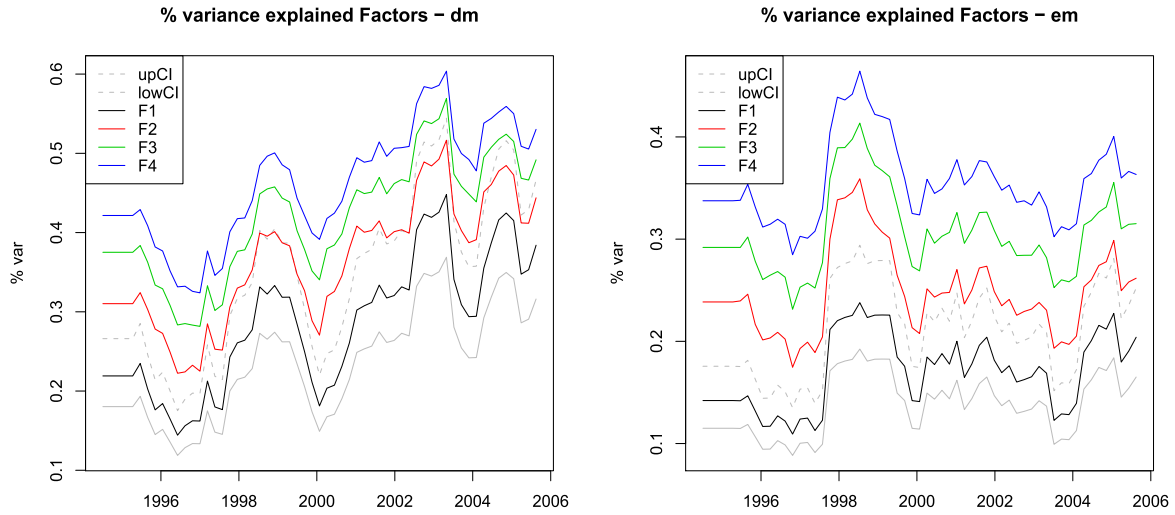
observed at the beginning of a year correspond to comovements calculated with the data of the whole previous year ending that date.

3 on average and never higher than 4. The standardization of the series before the calculation of the criteria plays a role in reducing the number of factors. We have concluded from the analysis of this rolling measure that there is some instability in determining the number of factors. However, a regularity appears if we use the criteria of Bai and Ng to analyze what should be the maximum number of factors. These criteria could be used to determine that information.

Cumulative measure and alternative way to determine the number of factors:

A complementary view to the parts of variance explained by each factor is to present the cumulative parts of variance explained by k factors. The following figure presents it:

Figure 1.6: Confidence Intervals and cumulative variance explained for DM and EM



We have also plotted, in Figure 1.6, the upper and lower confidence intervals around the measure C_t^1 calculated according to equation 1.17. The information corresponding to the confidence intervals allows to determine if the comovements measured with one or two factors are similar or not, that is if C_t^1 and $C_t^{1:2}$ could be considered statistically different. Two main results arise from this analysis. The first one is that the main

variation in the comovements is dominated by F1 in the sense that the dynamics of the comovements are ruled by this factor. The second result concerns the comparison of $C_t^{1:2}$ with the confidence interval of C_t^1 in order to determine whether these two measures are statistically different. For the set of developed markets, the measure $C_t^{1:2}$ is very often inside the bounds, particulary after 2001. However, it is not the case for the set of emerging markets where $C_t^{1:2}$ is always outside the bounds. Based on this observation, one could argue that for the end of the period and for the set of developed markets, only one factor is relevant as the measure of comovements based on one factor is not significantly different from the one based on the first two factors. However, this result does not confirm the findings presented in figure 1.5 concerning the number of factors calculated with the method of Bai and Ng where the number of factors is higher than 1 for developed market in many periods.

In brief, after analyzing the comovements associated with different numbers of factors, we conclude that the results are different for developed and emerging markets. For the developed markets, we decide to focus only on the first factor. The reasons are the following: first, we have observed that only the first factor drives the major dynamics in comovements; second, the number of factors is very unstable for the set of developed markets when it is determined by the criteria of Bai and Ng; third, there is a lack of interpretation for the resulting number of factor and finally, based on the analysis of C_t^1 and its confidence intervals, the measures C_t^1 and $C_t^{1:2}$ are statistically indistinguishable during the most recent period.

However, for the set of emerging markets, it is quite different. If we observe individually each measure C_t, we can see that not only C_t^1 characterizes the comovements but also that, during some short periods, the inclusion of a second or third factor should be relevant. The criteria of Bai and Ng suggest to use only one or two factors for the most recent period but these result is not confirmed by the alternative analysis of figure 1.6. We conclude that it is reasonable to include at least two factors for the set of emerging countries.

1.5.2 Integration

The use of a factor structure allows to relate the analysis of equity returns to the asset pricing literature, particularly to the concept of market integration. In fact, as we have highlighted in section 1.2.2, the largest part of the asset pricing literature is based on factor models as the CAPM or APT. Therefore, we can use the analysis of comovements as a way to gauge integration if the first factors explain a large part of the dynamics of the data set. Most of the literature studying integration with asset pricing models starts from the hypothesis stating that there should be only one global factor and in the way that this factor commands the returns there is a shift towards integration. Indeed, we have found in the literature many studies that test international versions of the CAPM by regressing the asset's returns on the global market factor and a constant that should not be significant in order to validate this model. However, we have seen from our previous results, that for emerging markets it should be sound to integrate more than one global factor. We consider this result in the following analysis.

In this subsection, we start by analyzing the possible interpretation of the first statistical factor as an global market factor. Then, we analyze the comovements measure in terms of integration in different ways. First by studying the comovements as the intensity of correlation of the returns with the first factor for both kind of markets as well as in a model with two factors for the set of emerging markets. We have also analyzed the contribution of regional or global factors for developed and emerging markets.

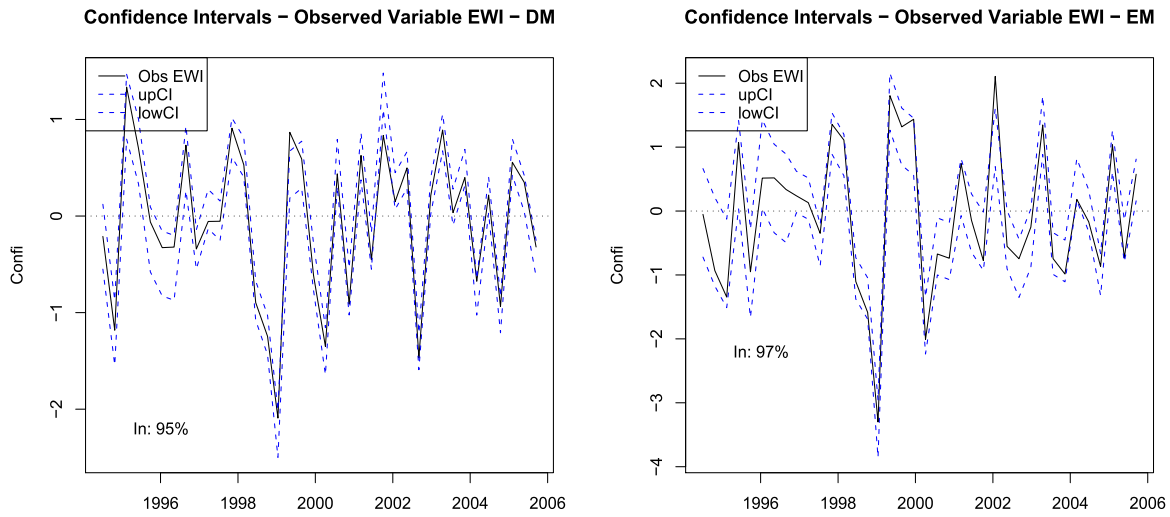
Identification of the main factor

The first point is then to test empirically if the first factor could be interpreted as a market factor. The factor $F1$ is estimated as the factor that explains the largest part of the variance of the data set. An image frequently presented to explain this construction is to represent a two dimensional data set by the cloud of points representing it[44]. The

[44]We present a similar picture in Figure 1 -A in the Appendix

first factor is positioned in the middle of the cloud of points, therefore, the estimated first factor could be similar to an Equally Weighted Index (EWI) which is an average of the data set. An EWI composed of the indexes in the data set is very easy to build by computing the average return at each date with all the components of the index. We want to evaluate how close this EWI is to the main factor. For this, we calculate a 95% confidence interval for the true factor according to equation 1.18 and we expect that the observed variable lies within these bounds. The results of this analysis are presented in Figure 1.7 for the sets of developed and emerging markets. We observe that the EWI lies within the confidence interval for almost the whole period of study. The same result is observed for the other data sets. Indeed, we have also calculated the number of times that the observed variable lies within the bounds of the true factor, we find 95% of the times for the set of developed markets and 97% for the set of emerging ones. We conclude that the main factor can be interpreted as an EWI composed with all the variables in the data set. This is why we can use the first factor or an EWI as global market factors in an international asset pricing approach to test for integration.

Figure 1.7: EWI vs F1



Integration measures

In the case where only one factor explains the comovements of a given data set, the tests of integration based on the asset pricing literature would consist in verifying that the global factor or the market portfolio prices each asset (in this case country indexes) correctly. The market portfolio should be the portfolio containing all the available risky assets in the market. According to the international version of the CAPM, the market risk premia is just given by the risk premia of the global market.

In fact, when all the barriers between markets are eliminated, the prices of traded assets in both markets should be the same if th, wey share the same part of common risk. Thus, we can say that integration has been achieved. This theoretical definition of integration is impossible to test for various reasons already discussed in section 1.2.3. First, the set of barriers are difficult to determine and even more to quantify. Second, it is difficult to observe assets sharing the same common risk in different markets. Third, we observe a changing process of integration and not only the two perfect states of integration and segmentation of the markets. Finally, this approach would require to test integration for each pair of markets within the set. The reality is obviously very different. However, one can claim that the gradual process of eliminating barriers should be reflected by the changing level of comovements between the markets. Also, we want to focus on the integration of a region rather than pairs of markets, therefore we focus on aggregated measures of comovements. If the first factor represents the market[45] as showed before, then we can build two approaches to measure integration based on the measure C_t described in equation 1.3.3.

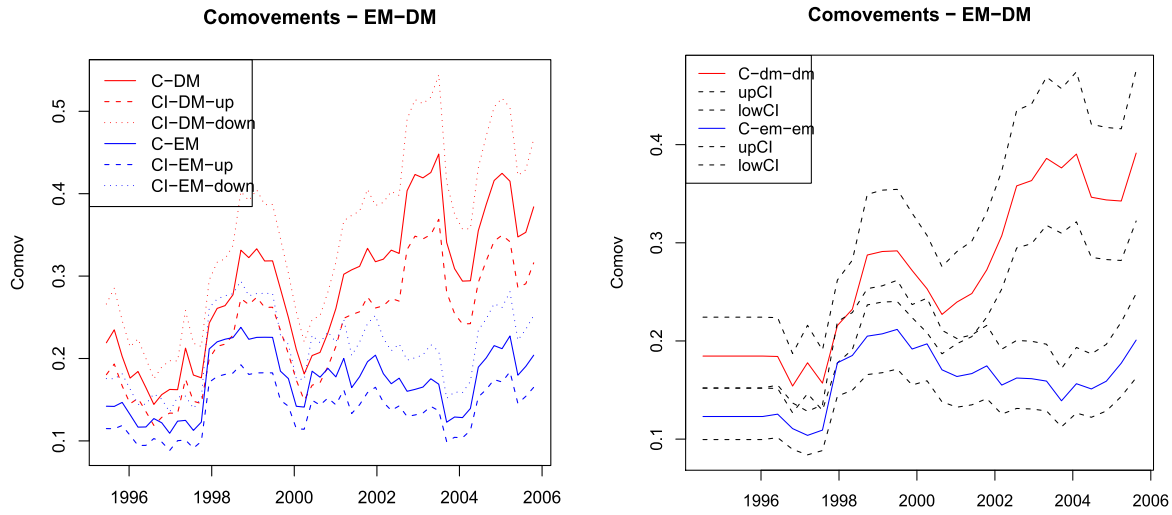
To begin with, we evaluate if there is an increase in the comovements within each type of market[46]. This would reflect a process of increasing integration for these sets of markets. In Figure 1.8, we present the measure C_t for the two main sets of markets: developed and emerging. In order to compare these two measures and their changes over time, we have

[45] Represented by an EWI.

[46] measured with C_t which is based on part of variance explained by the principal factor (market factor)

built confidence intervals around the measure C_t according to equation 1.17.

Figure 1.8: Commovements and Confidence intervals DM-EM



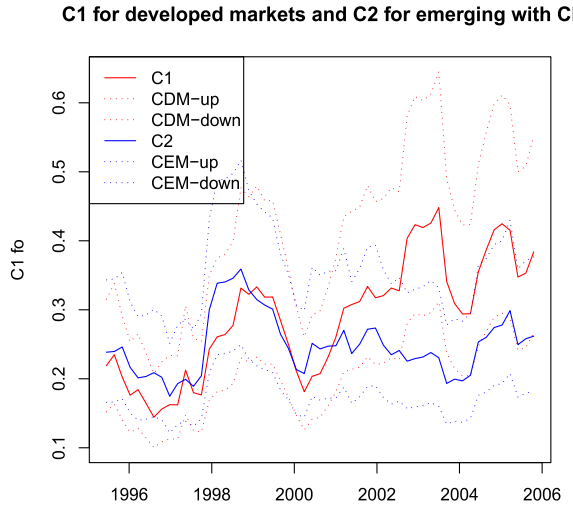
We observe in Figure 1.8 the comovements measure calculated over rolling windows of one year on the left and two years on the right, for developed and emerging markets. The comparison of these two pictures shows the effect of calculating the comovements with larger windows as we have analyzed in the first paragraph of section 1.5.1. The picture on the right has smoother curves and show better the trends, the picture on the left shows more details in the changes of the comovements. During the initial period, these two measures are not significantly different. However, for the last period, the difference is significant. We observe that after 2001, the comovements among developed countries is continually increasing. In terms of integration, it reveals a process of integration intensified after the year 2000. It is possible that the introduction of the Euro **were** responsible for this increase in comovements for the set of developed markets[47]. Also, we observe that only the group of developed countries presents a significant increase in integration over the period. We can confirm this observation with a regression of the measure C_t for each

[47]The European (euro zone) countries represent and important part of the developed world, and as we do not take into account size effects, their importance is even higher.

type of market on a linear trend. The coefficients are positive for both types of markets and higher for the set of developed markets. The results are presented in Table 1 -E in the Appendix.

We repeat the same analysis but considering that for the set of emerging markets there should be two global factors in the asset pricing model. The results are presented in Figure 1.9.

Figure 1.9: C1 for developed markets and C2 for emerging with CI



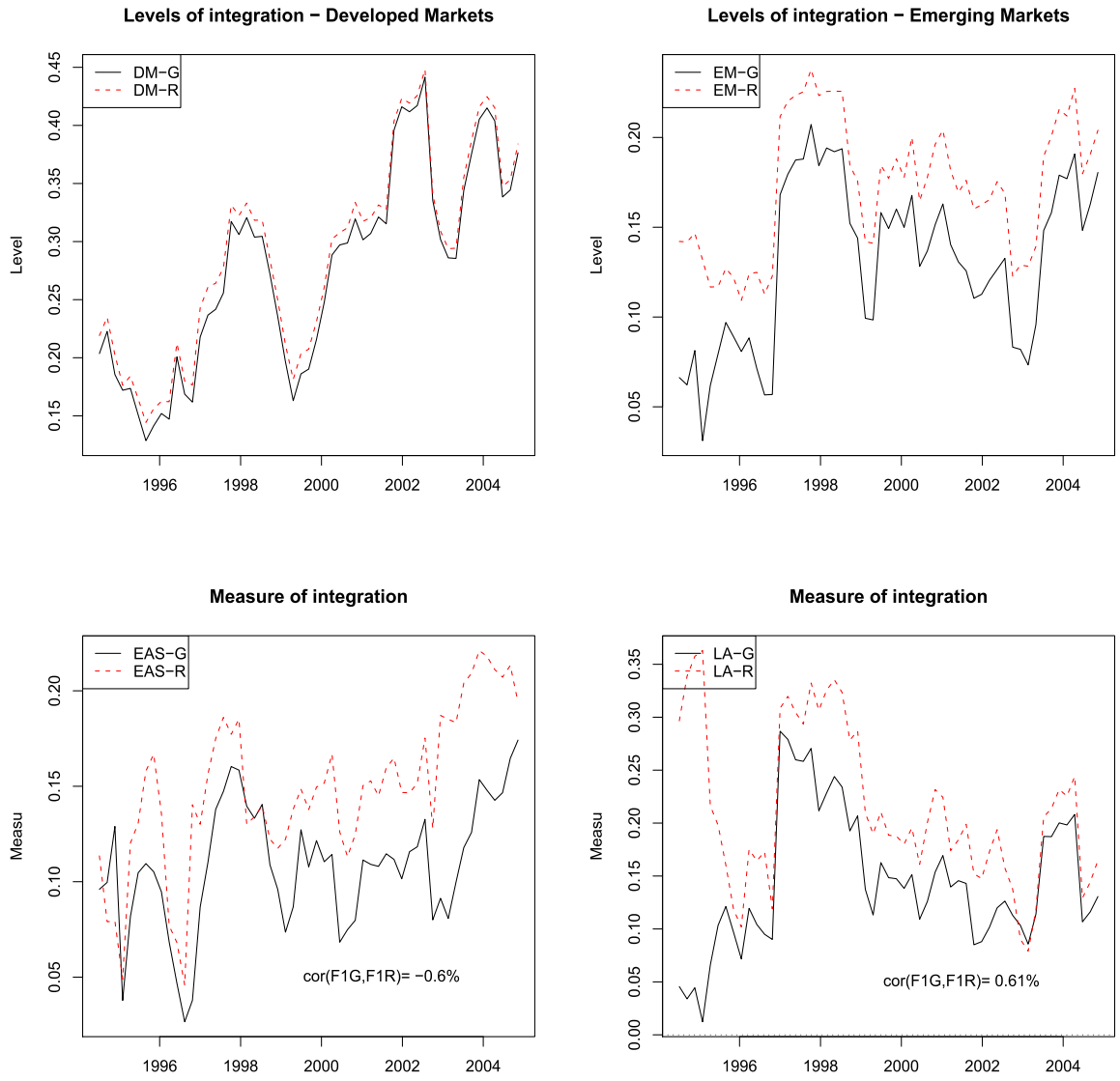
We can see that the results are similar to the previous ones but the two measures are closer as we have included two factors for the set of emerging countries.

The next analysis consist in investigating if a set of variables is better explained by a global factor or a regional factor. If the contribution of the global factor - as measured by C_t^{F1G} - is small compared to the contribution of the regional factor C_t^{F1R}, one can think that the analyzed markets are partially segmented from the global market[48]. Moreover, we compare C_t^{F1R} and C_t^{F1G} for developed and emerging markets, as well as for the group

[48]We compute the comovement measure C_t^{F1R} from the first factor estimated from the returns observed in the region.

of Latin American (LA) markets and the group of Asian (EAS) countries. Within each region, an increase in the comovements with global or regional factors would be related to an increase in the level of integration. The results are presented in Figure 1.10. The graphic on the top left of Figure 1.10 corresponds to the set of developed markets, the one in the top right to emerging countries and the graphics in the bottom of the figure present the results for a set of Asian emerging countries (EAS) on the left and Latin American countries (LA) on the right.

Figure 1.10: Percentage of Variance explained



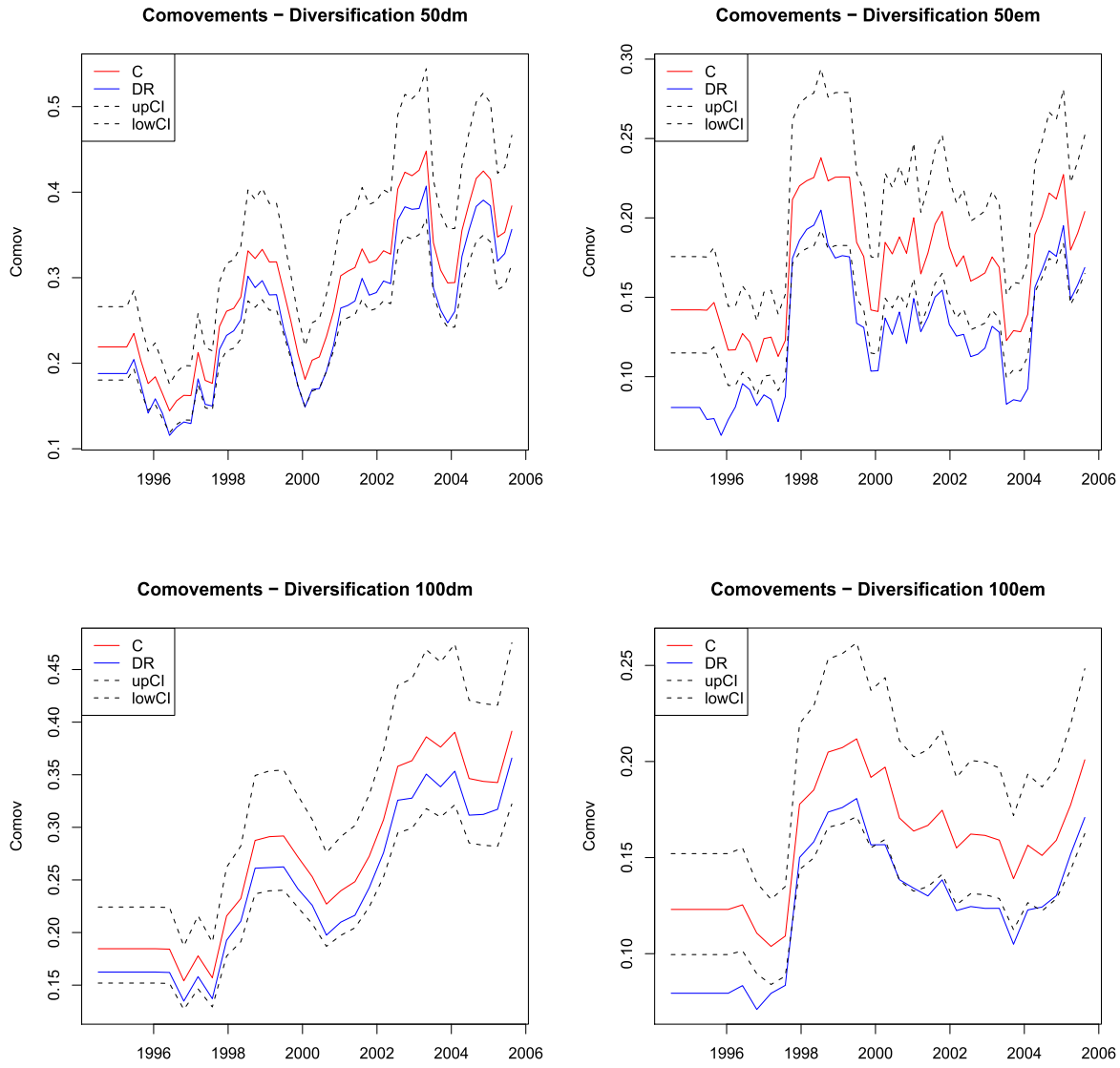
For each data set, the regional factor $F1R$ accounts for a larger part of the comovements than the global factor. However, for the set of developed markets, we observe that the two measures are very similar. It means that the information contained in emerging markets does not help to explain the comovements in Developed Markets. This result confirms that only one factor is needed to explain returns in developed countries. Moreover, it suggests that developed markets are integrated to the corresponding global factor which is not the case for the different groups of emerging markets. Focusing on emerging

regions, only the group of Asian countries presents a slight upward trend in the comovements, suggesting that they have become more integrated to the associated global factor. At the beginning of the period, the part of comovements explained by C_t^{F1G} for the set of Latin American markets is insignificant compared to C_t^{F1R}, indicating a clear segmentation before 1997. We have also reported correlation coefficients between the factors C_t^{F1G} and C_t^{F1R} for each figure. For the set of developed markets, the two factors are highly correlated. This coefficient combined with the analysis of C_t^{F1R} and C_t^{F1G} over time give important information for asset pricing applications. Indeed, we observe that this coefficient of correlation is positive for the set of Latin American markets but negative for the set of Asian Markets. Thus, in the case of Latin America, the addition of the regional factor to the pricing model would not be that important, contrary to the case of the Asian countries.

1.5.3 Diversification

The last important feature that we want to highlight about these measures of comovements is that they are closely related to diversification issues. For this, we have computed the diversification ratio defined in equation 1.19 and we have compared it to the measure of comovements. Figure 1.11 shows that both measures are closely related. Both measures are defined between zero and one, where one indicates the highest level of comovements and no diversification opportunities; and zero indicates the opposite: weak comovements and the maximum level of diversification. The diversification ratio captures the part of variance that can not be diversified away, which should correspond to the part of risk explained by the common factors or the market factor. The measure of comovements can therefore be used to capture the dynamics of the potential diversification.

Figure 1.11: Percentage of Variance explained



The diversification ratio proposed by Solnik (1974) was calculated for a limited size of portfolios, here we calculate this ratio for a portfolio of the largest possible size. Hence, we can interpret this measure of diversification as the risk reduction limit available from our data set. It takes a value close to 0.15 for the global set of variables, this means that 15% of the average risk of an index could not be diversified away with diversification. We observe that for emerging countries this value is 0.06 at the beginning of the period and 0.14 at the end. For developed markets, it takes a value close to 0.18 at the beginning

compared to 0.30 at the end of the period. This means that there are fewer diversification opportunities within developed markets than within emerging markets and that for both kinds of markets, the diversification potential has been reduced over time. We have also plotted in Figure 1.11, confidence intervals for the measure of comovements C_t. We observe that the diversification ratio lies within these confidence intervals for the set of developed markets (plotted on the left) and it is close to the lower confidence band for the emerging markets data set (plotted on the right). We conclude that for developed countries, the measures C_t and DR_t are not significantly different during the whole period of analysis and for the set of emerging countries these two measures are not significantly different during the phases of high comovements. An high level in the measure of comovements can therefore be used as an indicator of lower diversification opportunities for both kind of markets.

1.6 Summary and Conclusions

In this chapter, we have analyzed the comovements among a set of global equity markets for a 12 years period based on a factor decomposition of country-sector based index returns in a large data environment. After describing the data set and decomposing the structure of returns into principal components, we have presented and analyzed a measure that accounts for the level of comovements based on the principal factors extracted from the data set. From this decomposition of equity returns into factors, we have shown: that the principal factor is responsible for the largest part of the comovements so that it drives its changes and that the structure of equity returns is different for emerging or developed markets. Additionally, we have determined that the first factor is highly correlated with an equally weighted index composed of the variables in the data set while the second factor is useful to account for episodes of crises such as the Asian and the Argentinean crisis. A main result is that there is an upward trend in the global comovements and also that this trend is stronger for developed than emerging countries.

In addition, taking advantage of the large cross-section and time series dimensions, we have estimated the number of relevant factors over a rolling window in the lines of Bai and Ng. Thus, we have concluded that for emerging countries, this number is constantly decreasing and that it is a number close to 4 for emerging markets and 3 for the developed markets on average. We have also analyzed the correspondence between the measure of comovements an the number of factors indicated with the Bai and Ng criteria by comparing for the same dates the behavior of this two indicators. We expected to have a diminishing number of factors while observing an increasing level of comovements, but the data does not confirm this statement. We conclude that a decreasing number of factors should not be interpreted as a sign of integration.

We have shown that the measure of comovements could be related to important economic and financial issues such as the process of integration of equity markets and the time varying opportunities of diversification offered by international markets.

To begin with, we have analyzed the measures of comovements considering them as proxies for the levels of integration. The analysis of integration has been conducted in two ways. First, we have shown that the level of integration is significantly higher for the set of developed markets and that integration has increased for these markets during the period. We have observed that these integration measures behave differently depending on the area, on the period and on the data frequency. Second, we have analyzed the effects of global and regional factors as indicators of integration concluding that only the set of developed markets are integrated to a global market. To complement the analysis of integration we have evaluated some asset pricing models based on the findings of the structure of equity returns for the different types of markets.

Finally, we have shown that our measure of comovements can be used as a measure of diversification like the diversification ratio proposed by Solnik (1974). Important for portfolio strategies is the finding that for the global set of indexes a maximum of 15 % of the risk of a typical security could be reduced with diversification. Moreover, for developed markets, this percentage is of 18% at the beginning of the period and 30% at the end, compared to 6% and 14% respectively for emerging markets.

Many extensions of this analysis are possible. First, we have focused on contemporaneous comovements leaving aside prediction considerations. Hence, the analysis of comovements we have proposed should be extended to the past and for any portfolio application it would be relevant to estimate the future level of comovements. Second, we have analyzed the comovements for a global set of indexes and we have identified periods of high comovements and fewer diversification opportunities and periods of low comovements and higher opportunities of diversification. It would be interesting for future analysis to further investigate how to take this information into account for the management of equity portfolios, to build portfolio strategies based on the different levels of global comovements.

Such extensions are treated in the following chapters. In the second chapter, we

analyze a prediction model for different emerging equity markets. The results about the structure of returns where several factors are needed is a guide to build this model. In the third chapter, we use the information about the time varying comovements to construct a portfolio strategy for emerging and developed markets.

1.7 Appendix

Table 1 -A: Descriptive Statistics calculated for the period

	Mean	**Std**	**Skew**	**Kurt**	**AC1**	**AC2**	**AC3**	**AC4**
AUSTRALIA	8.16%	15.78%	- 0.337	0.462	0.021	0.043	0.002	- 0.001
AUSTRIA	8.75%	15.46%	- 0.453	1.217	0.005	- 0.024	0.082	0.064
BELGIUM	8.33%	17.71%	- 0.630	3.793	- 0.116	0.050	0.046	- 0.019
CANADA	12.28%	16.05%	- 0.464	1.171	0.039	0.007	0.071	0.070
DENMARK	10.78%	17.03%	- 0.484	1.935	- 0.050	0.025	0.009	- 0.012
FINLAND	13.48%	30.33%	- 0.695	2.062	- 0.056	0.055	0.065	- 0.032
FRANCE	8.77%	18.85%	- 0.424	2.414	- 0.144	0.039	0.079	- 0.016
GERMANY	6.00%	19.32%	- 0.559	2.181	- 0.085	0.051	0.042	- 0.019
GREECE	10.55%	25.87%	- 0.299	1.439	0.011	0.043	0.001	0.084
HONG.KONG	4.64%	23.57%	- 0.566	1.806	0.046	0.054	0.125	- 0.065
IRELAND	11.72%	17.99%	- 0.517	1.943	- 0.015	0.011	- 0.014	- 0.008
ITALY	6.82%	20.66%	- 0.308	1.237	- 0.020	0.018	0.106	- 0.045
JAPAN	-1.32%	20.91%	0.181	0.816	- 0.029	0.025	0.018	- 0.001
NETHERLANDS	7.46%	18.36%	- 0.814	4.746	- 0.162	0.105	0.026	- 0.060
NEW.ZEALAND	4.70%	16.63%	- 0.543	1.452	0.062	0.064	- 0.015	0.051
NORWAY	11.61%	20.55%	- 0.741	2.734	0.018	- 0.051	0.034	0.102
PORTUGAL	7.83%	17.36%	- 0.151	1.191	0.031	0.060	0.045	0.098
SINGAPORE	1.12%	20.19%	- 0.186	1.966	0.102	0.031	0.059	- 0.047
SPAIN	10.99%	19.04%	- 0.606	1.525	- 0.082	0.064	- 0.003	0.072
SWEDEN	11.14%	24.42%	- 0.510	1.958	- 0.097	0.022	0.056	- 0.042
SWITZERLAND	8.53%	17.20%	- 0.901	5.150	- 0.050	0.047	0.034	- 0.020
UNITED.KINGDOM	7.05%	15.01%	- 0.220	2.115	- 0.133	0.018	0.068	- 0.077
U.S.A	8.76%	16.16%	- 0.215	1.893	- 0.091	- 0.002	0.032	- 0.014
DEVELOPED	8.18%	19.32%	- 0.454	2.052	- 0.035	0.033	0.042	0.003
ARGENTINA	-3.82%	33.30%	- 0.356	4.073	0.048	0.029	- 0.001	0.003
BRAZIL	8.42%	34.74%	- 0.693	3.749	- 0.044	0.137	0.027	0.041
CHILE	3.63%	18.75%	- 0.157	2.805	0.136	0.096	0.106	0.002
CHINA	7.17%	32.90%	- 0.324	2.374	0.012	0.059	0.120	0.010
COLOMBIA	-1.35%	29.47%	- 0.414	3.367	- 0.055	0.017	0.088	0.081
CZECH.REPUBLIC	9.90%	23.32%	- 0.475	1.182	0.005	0.061	0.136	- 0.037
HUNGARY	11.77%	28.30%	- 0.282	2.569	0.020	0.106	0.069	0.055
INDIA	4.76%	27.63%	- 0.501	2.000	0.043	0.131	- 0.006	0.010
ISRAEL	8.70%	23.05%	- 0.410	1.480	- 0.026	0.041	0.010	- 0.005
KOREA	5.87%	41.39%	- 0.209	2.362	- 0.078	0.114	0.002	0.030
MALAYSIA	-2.22%	29.43%	- 0.504	8.401	0.025	- 0.069	0.164	0.023
MEXICO	5.18%	29.96%	- 0.568	4.145	0.091	0.065	0.082	- 0.016
PAKISTAN	0.47%	32.42%	- 0.408	2.440	0.094	0.110	0.002	- 0.041
PERU	7.33%	20.15%	- 0.198	5.764	- 0.014	0.051	0.040	0.080
PHILIPPINES	-5.56%	28.53%	- 0.273	4.873	0.024	0.127	0.142	- 0.032
POLAND	6.66%	32.02%	- 0.026	1.662	0.024	0.039	0.074	- 0.062
RUSSIA	26.75%	44.28%	- 0.389	2.069	0.075	0.085	0.117	- 0.039
SOUTH.AFRICA	6.99%	24.16%	- 0.854	2.656	0.036	0.089	0.089	0.011
TAIWAN	0.52%	28.09%	- 0.082	1.307	0.041	0.036	0.019	- 0.009
THAILAND	-6.56%	36.10%	0.077	1.752	0.054	0.102	0.100	- 0.040
TURKEY	10.39%	50.43%	- 0.258	0.884	0.003	0.036	0.075	- 0.006
VENEZUELA	-0.72%	35.85%	- 0.096	3.691	0.049	0.055	0.048	- 0.066
EMERGING	4.74%	31.10%	- 0.336	2.982	0.026	0.069	0.068	- 0.000
WORLD	6.46%	25.21%	- 0.395	2.517	- 0.004	0.051	0.055	0.001

Table 1 -B: Correlation coefficients for the set of developed countries

	AU	OE	BG	CN	DK	FN	FR	BD	GR	HK	IR	IT	JP	NL	NZ	NW	PT	SG	ES	SD	SW	UK	US
AU	1.00	0.37	0.47	0.57	0.46	0.38	0.52	0.54	0.32	0.49	0.48	0.48	0.38	0.52	0.62	0.48	0.32	0.45	0.52	0.50	0.50	0.55	0.46
OE	0.37	1.00	0.60	0.40	0.52	0.26	0.49	0.53	0.43	0.19	0.49	0.44	0.26	0.51	0.36	0.54	0.50	0.29	0.53	0.41	0.52	0.44	0.24
BG	0.47	0.60	1.00	0.46	0.60	0.38	0.74	0.74	0.46	0.28	0.59	0.59	0.26	0.80	0.37	0.52	0.53	0.26	0.66	0.56	0.73	0.68	0.48
CN	0.57	0.40	0.46	1.00	0.41	0.51	0.61	0.62	0.33	0.51	0.49	0.53	0.39	0.61	0.38	0.57	0.36	0.40	0.54	0.61	0.49	0.61	0.73
DK	0.46	0.52	0.60	0.41	1.00	0.43	0.58	0.63	0.39	0.30	0.53	0.50	0.26	0.61	0.33	0.59	0.48	0.27	0.58	0.57	0.53	0.56	0.36
FN	0.38	0.26	0.38	0.51	0.43	1.00	0.60	0.61	0.33	0.39	0.41	0.49	0.33	0.55	0.28	0.48	0.41	0.30	0.53	0.72	0.44	0.55	0.56
FR	0.52	0.49	0.74	0.61	0.58	0.60	1.00	0.85	0.44	0.44	0.59	0.75	0.37	0.85	0.39	0.58	0.57	0.38	0.78	0.76	0.72	0.78	0.62
BD	0.54	0.53	0.74	0.62	0.63	0.61	0.85	1.00	0.51	0.47	0.62	0.70	0.38	0.85	0.42	0.58	0.58	0.38	0.77	0.76	0.73	0.74	0.64
GR	0.32	0.43	0.46	0.33	0.39	0.33	0.44	0.51	1.00	0.25	0.43	0.38	0.28	0.43	0.29	0.40	0.42	0.24	0.44	0.38	0.44	0.38	0.28
HK	0.49	0.19	0.28	0.51	0.30	0.39	0.44	0.47	0.25	1.00	0.36	0.37	0.37	0.43	0.39	0.36	0.23	0.62	0.41	0.45	0.33	0.51	0.48
IR	0.48	0.49	0.59	0.49	0.53	0.41	0.59	0.62	0.43	0.36	1.00	0.53	0.28	0.61	0.37	0.53	0.44	0.34	0.57	0.56	0.57	0.64	0.51
IT	0.48	0.44	0.59	0.53	0.50	0.49	0.75	0.70	0.38	0.37	0.53	1.00	0.29	0.68	0.36	0.51	0.47	0.30	0.70	0.65	0.58	0.65	0.52
JP	0.38	0.26	0.26	0.39	0.26	0.33	0.37	0.38	0.28	0.37	0.28	0.29	1.00	0.35	0.26	0.30	0.22	0.35	0.31	0.36	0.29	0.33	0.32
NL	0.52	0.51	0.80	0.61	0.61	0.55	0.85	0.85	0.43	0.43	0.61	0.68	0.35	1.00	0.42	0.61	0.54	0.35	0.74	0.72	0.75	0.80	0.63
NZ	0.62	0.36	0.37	0.38	0.33	0.28	0.39	0.42	0.29	0.39	0.37	0.36	0.26	0.42	1.00	0.37	0.34	0.40	0.39	0.37	0.39	0.43	0.30
NW	0.48	0.54	0.52	0.57	0.59	0.48	0.58	0.58	0.40	0.36	0.53	0.51	0.30	0.61	0.37	1.00	0.44	0.37	0.57	0.59	0.53	0.58	0.44
PT	0.32	0.50	0.53	0.36	0.48	0.41	0.57	0.58	0.42	0.23	0.44	0.47	0.22	0.54	0.34	0.44	1.00	0.19	0.60	0.52	0.52	0.45	0.27
SG	0.45	0.29	0.26	0.40	0.27	0.30	0.38	0.38	0.24	0.62	0.34	0.30	0.35	0.35	0.40	0.37	0.19	1.00	0.32	0.37	0.31	0.40	0.35
ES	0.52	0.53	0.66	0.54	0.58	0.53	0.78	0.77	0.44	0.41	0.57	0.70	0.31	0.74	0.39	0.57	0.60	0.32	1.00	0.71	0.65	0.67	0.53
SD	0.50	0.41	0.56	0.61	0.57	0.72	0.76	0.76	0.38	0.45	0.56	0.65	0.36	0.72	0.37	0.59	0.52	0.37	0.71	1.00	0.60	0.70	0.63
SW	0.50	0.52	0.73	0.49	0.53	0.44	0.72	0.73	0.44	0.33	0.57	0.58	0.29	0.75	0.39	0.53	0.52	0.31	0.65	0.60	1.00	0.67	0.49
UK	0.55	0.44	0.68	0.61	0.56	0.55	0.78	0.74	0.38	0.51	0.64	0.65	0.33	0.80	0.43	0.58	0.45	0.40	0.67	0.70	0.67	1.00	0.65
US	0.46	0.24	0.48	0.73	0.36	0.56	0.62	0.64	0.28	0.48	0.51	0.52	0.32	0.63	0.30	0.44	0.27	0.35	0.53	0.63	0.49	0.65	1.00

Table 1 -C: Correlation coefficients for the set of emerging countries

	AR	BR	CL	CH	CB	CZ	HN	IN	IS	KO	MY	MX	PK	PE	PH	PO	RS	SA	TA	TH	TK	VE
AR	1.00	0.51	0.45	0.17	0.16	0.18	0.26	0.13	0.25	0.17	0.12	0.56	0.06	0.33	0.22	0.20	0.29	0.28	0.20	0.25	0.21	0.18
BR	0.51	1.00	0.53	0.20	0.20	0.23	0.35	0.21	0.30	0.26	0.12	0.58	0.15	0.36	0.23	0.32	0.33	0.39	0.19	0.23	0.24	0.19
CL	0.45	0.53	1.00	0.24	0.21	0.19	0.29	0.19	0.28	0.21	0.17	0.49	0.12	0.41	0.22	0.27	0.29	0.39	0.24	0.30	0.27	0.19
CH	0.17	0.20	0.24	1.00	0.13	0.21	0.16	0.17	0.12	0.27	0.34	0.20	0.04	0.17	0.30	0.17	0.18	0.29	0.24	0.35	0.13	0.12
CB	0.16	0.20	0.21	0.13	1.00	0.22	0.20	0.08	0.18	0.07	0.12	0.21	0.13	0.16	0.14	0.13	0.17	0.24	0.09	0.17	0.17	0.10
CZ	0.18	0.23	0.19	0.21	0.22	1.00	0.51	0.20	0.15	0.25	0.18	0.22	0.13	0.12	0.21	0.46	0.25	0.38	0.18	0.23	0.25	0.17
HN	0.26	0.35	0.29	0.16	0.20	0.51	1.00	0.22	0.26	0.27	0.11	0.33	0.14	0.19	0.19	0.52	0.36	0.45	0.14	0.20	0.33	0.18
IN	0.13	0.21	0.19	0.17	0.08	0.20	0.22	1.00	0.21	0.25	0.18	0.17	0.23	0.08	0.21	0.25	0.18	0.24	0.27	0.23	0.18	0.09
IS	0.25	0.30	0.28	0.12	0.18	0.15	0.26	0.21	1.00	0.16	0.11	0.31	0.05	0.21	0.15	0.24	0.25	0.30	0.22	0.19	0.22	0.16
KO	0.17	0.26	0.21	0.27	0.07	0.25	0.27	0.25	0.16	1.00	0.18	0.22	0.04	0.14	0.27	0.24	0.15	0.30	0.32	0.34	0.20	0.07
MY	0.12	0.12	0.17	0.34	0.12	0.18	0.11	0.18	0.11	0.18	1.00	0.13	0.12	0.13	0.42	0.20	0.17	0.20	0.31	0.45	0.10	0.10
MX	0.56	0.58	0.49	0.20	0.21	0.22	0.33	0.17	0.31	0.22	0.13	1.00	0.12	0.37	0.22	0.31	0.33	0.37	0.22	0.25	0.23	0.21
PK	0.06	0.15	0.12	0.04	0.13	0.13	0.14	0.23	0.05	0.04	0.12	0.12	1.00	0.06	0.11	0.18	0.06	0.09	0.09	0.12	0.10	0.01
PE	0.33	0.36	0.41	0.17	0.16	0.12	0.19	0.08	0.21	0.14	0.13	0.37	0.06	1.00	0.18	0.15	0.21	0.29	0.12	0.20	0.12	0.13
PH	0.22	0.23	0.22	0.30	0.14	0.21	0.19	0.21	0.15	0.27	0.42	0.22	0.11	0.18	1.00	0.23	0.13	0.26	0.30	0.49	0.13	0.14
PO	0.20	0.32	0.27	0.17	0.13	0.46	0.52	0.25	0.24	0.24	0.20	0.31	0.18	0.15	0.23	1.00	0.22	0.40	0.22	0.24	0.25	0.12
RS	0.29	0.33	0.29	0.18	0.17	0.25	0.36	0.18	0.25	0.15	0.17	0.33	0.06	0.21	0.13	0.22	1.00	0.37	0.20	0.23	0.25	0.11
SA	0.28	0.39	0.39	0.29	0.24	0.38	0.45	0.24	0.30	0.30	0.20	0.37	0.09	0.29	0.26	0.40	0.37	1.00	0.27	0.35	0.29	0.15
TA	0.20	0.19	0.24	0.24	0.09	0.18	0.14	0.27	0.22	0.32	0.31	0.22	0.09	0.12	0.30	0.22	0.20	0.27	1.00	0.32	0.14	0.10
TH	0.25	0.23	0.30	0.35	0.17	0.23	0.20	0.23	0.19	0.34	0.45	0.25	0.12	0.20	0.49	0.24	0.23	0.35	0.32	1.00	0.22	0.13
TK	0.21	0.24	0.27	0.13	0.17	0.25	0.33	0.18	0.22	0.20	0.10	0.23	0.10	0.12	0.13	0.25	0.25	0.29	0.14	0.22	1.00	0.19
VE	0.18	0.19	0.19	0.12	0.10	0.17	0.18	0.09	0.16	0.07	0.10	0.21	0.01	0.13	0.14	0.12	0.11	0.15	0.10	0.13	0.19	1.00

Table 1 -D: Loadings from Factor Decomposition

	1	2	3	4	5	6	7	8
pe	0.36	0.07	0.05	0.04	0.03	0.02	0.02	0.02
BD	1.44	1.02	−0.35	−0.47	0.18	−0.16	0.23	−0.26
FR	1.41	1.17	−0.25	−0.76	0.28	−0.19	0.24	−0.47
NL	1.40	1.20	−0.30	−0.71	−0.21	−0.04	0.43	−0.14
UK	1.36	0.77	−0.34	−0.91	−0.15	0.24	0.62	−0.10
ES	1.35	0.91	0.35	−0.44	−0.50	−0.57	−0.58	−0.83
SD	1.35	0.80	0.09	−0.80	1.18	0.31	0.23	−0.33
CN	1.26	−0.23	0.37	−0.92	0.42	0.40	0.71	0.59
NW	1.24	0.27	0.13	0.79	−0.04	0.42	0.29	0.15
IT	1.23	0.89	0.13	−0.59	−0.07	−0.88	−0.17	−0.51
SW	1.21	1.30	−0.66	−0.12	−0.81	−0.23	−0.17	−0.00
BG	1.20	1.62	−0.63	0.05	−1.38	−0.03	0.18	−0.22
IR	1.18	0.65	−0.43	0.19	−0.81	−0.19	0.95	0.46
US	1.17	0.12	0.42	−2.04	0.95	0.28	1.45	0.40
AU	1.17	−0.48	−0.59	−0.35	−0.85	−1.17	−0.91	1.58
DK	1.11	1.02	−0.55	0.76	−0.27	−0.68	0.16	−0.37
FN	1.09	0.49	0.25	−0.90	2.36	0.51	−0.04	−0.37
SA	1.08	−0.58	0.40	1.05	−0.43	0.76	−0.43	1.13
HK	1.04	−1.50	−0.97	−1.21	0.36	1.03	−0.18	−1.14
OE	1.02	0.73	−0.52	1.88	−1.72	−0.07	0.05	0.35
PT	1.00	1.19	−0.24	1.16	−0.09	−0.20	−1.29	−1.03
HN	0.99	−0.07	0.78	2.14	0.88	1.25	−0.70	−0.07
NZ	0.95	−0.62	−0.83	0.23	−1.50	−0.83	−1.78	1.80
BR	0.94	−0.89	2.18	−0.25	−0.22	−0.65	−1.01	−0.58
SG	0.93	−1.64	−1.75	−0.42	−0.04	−0.11	0.80	−0.13
MX	0.93	−0.91	2.32	−0.77	−0.14	−0.34	−0.33	−0.79
GR	0.92	0.43	−0.12	1.51	−0.35	−0.12	−0.38	0.44
PO	0.89	−0.34	0.40	1.64	2.09	0.40	−0.55	−0.84
CL	0.87	−1.09	1.93	−0.23	−1.07	−0.58	−0.30	0.20
CZ	0.85	−0.01	−0.10	2.51	0.81	1.57	−0.06	−2.02
IS	0.81	−0.12	0.90	−0.54	0.95	−0.30	1.50	1.58
JP	0.79	−0.51	−0.63	−0.60	0.69	0.52	−1.10	1.90
AR	0.78	−0.94	2.26	−0.85	−0.65	−0.21	−0.19	−1.23
TH	0.76	−1.93	−1.15	0.01	−0.82	−0.09	0.49	0.19
KO	0.76	−1.01	−0.81	−0.17	0.95	0.84	−2.43	0.88
TA	0.71	−1.19	−0.94	−0.80	1.36	−0.54	−0.35	0.49
RS	0.70	−0.82	1.20	0.78	0.33	1.71	0.78	0.92
PH	0.66	−1.76	−1.30	0.28	−1.03	−0.80	−0.08	−0.75
TK	0.65	−0.46	0.77	1.32	0.72	1.14	1.05	2.62
CH	0.62	−1.55	−1.08	−0.22	−0.64	1.97	0.29	−2.32
IN	0.59	−0.83	−0.51	1.05	2.04	−2.46	−0.74	0.76
PE	0.52	−1.10	1.98	−0.14	−1.77	−0.41	−1.29	−0.48
MY	0.52	−1.90	−1.65	0.06	−0.38	−0.10	1.28	−0.94
CB	0.45	−0.64	0.76	1.15	−1.35	0.18	3.56	0.76
PK	0.23	−0.75	0.22	1.53	1.52	−4.03	1.56	−1.21

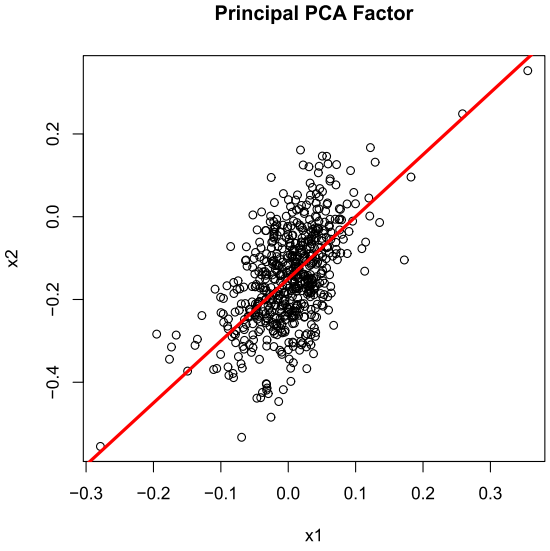
Table 1 -E: OLS regression of C_t measure on a linear trend

	DM	EM
(Intercept)	0.17*	0.15*
	(0.01)	(0.01)
X	0.39*	0.08*
	(0.04)	(0.03)
N	59	59
R^2	0.64	0.14
adj. R^2	0.63	0.13
Resid. sd	0.05	0.03

Standard errors in parentheses

* indicates significance at $p < 0.05$

Figure 1 -A: Cloud of points and estimation of PCA factors



Confidence intervals (From Bai and Ng (2006))

Based on a factor representation for the data X_{it}, such that:

$$X_{it} = \sum_{k=1}^{K} \Lambda_{ik} F_{kt} + e_{it}$$

where F_{kt} is an unobserved factor, Λ_{ik} is the sensitivity of asset i to factor k (or factor loading) and e_{it} is an error term.

Factors and loadings are estimated with the method of principal components. The estimated variables are denoted $\widetilde{F}_{kt}$ and $\widetilde{\Lambda}_{ik}$. We can write $\widetilde{\Lambda} = (\widetilde{\Lambda}_1, ..., \widetilde{\Lambda}_N)'$ as the matrix of factor loadings and $\widetilde{F} = (\widetilde{F}_1, ..., \widetilde{F}_T)$ the matrix of estimated factors. Then:

$$\widetilde{e}_{it} = X_{it} - \widetilde{\Lambda}_i' \widetilde{F}_t$$

Letting $\lambda = (\lambda_1, ..., \lambda_N)$ be the vector of eigenvalues of the matrix $\frac{X'X}{NT}$, $\widetilde{V} = diag(\lambda^k)$ is defined as the diagonal matrix containing the k largest eigenvalues.

The matrix $\widetilde{\Gamma}_t$ is a covariance estimator defined as:

$$\widetilde{\Gamma}_t = \frac{1}{N} \sum_{i=1}^{N} \widetilde{e}_{it}^2 \widetilde{\Lambda}_i \widetilde{\Lambda}_i'$$

when e_{it} is cross-sectionally uncorrelated. If $E[e_{it}^2] = \sigma_e^2, \forall t$ and $\forall i$, then $\widetilde{\Gamma}_t = \widehat{\sigma}_e^2 \frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{N}$ where $\widehat{\sigma}_e^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{e}_{it}^2$

The observed variable is denoted G_t. And the OLS estimate from the regression of G_{jt} on F_t is:

$$\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$$

Letting the measurement error be $\widehat{\varepsilon}_{jt} = G_{jt} - \widehat{G}_{jt}$, can be defined the term s_{jt}^2 estimated as:

$$\widehat{s}_{jt}^2 = \frac{1}{T}\widetilde{F}_t{}'\widetilde{F}_t\widehat{\sigma}_\varepsilon^2(j) + \frac{1}{N}\widehat{Avar}(\widehat{G}_{jt})$$

when ε_{jt} is conditionally homoscedastic.

The term $\widehat{Avar}(\widehat{G}_{jt})$ can be estimated as:

$$\widehat{Avar}(\widehat{G}_{jt}) = \hat{\gamma}_j'\widetilde{V}^{-1}\widetilde{\Gamma}_t\widetilde{V}^{-1}\hat{\gamma}_j'$$

And finally, the confidence intervals at a 95% confidence level are defined as:

$$CI : [\widehat{\varepsilon}_{jt} - 1.96\widehat{s}_{jt}, \widehat{\varepsilon}_{jt} + 1.96\widehat{s}_{jt}]$$

Chapter 2

Forecasting Equity Returns

2.1 Introduction

The predictability of equity returns has been for years a subject of intense discussion in financial economics. It is still one of the favorite questions debated in finance by academics and practitioners.

Traditionally, the exercise of forecasting uses one of the following approaches: the projection of a times series with only its past information or the cause and effect approach using related variables that explain the behavior of the forecasted series. Nowadays, recent econometric techniques allow to forecast a time series using a combination of its past information as well as past and contemporaneous information of related variables. We use this later approach using a Error Correction Model (ECM) for a set of selected variables in order to predict equity risk premia within a group of emerging markets.

In addition to the technical decisions relative to the forecasting model, there are several economic questions that we have to care about. The most important, the choice of the economic and financial variables that are able explain equity risk premia and the way each variable would influence the returns.

For several years, research in empirical finance has reported evidence in favor of return predictability using several economic variables and fundamental financial ratios. Following

the work of Ratio et al. (1988) or Fama and French (1989), many authors such as Ferson and Harvey (1999) or Hodrick (1992) among others have suggested to predict equity excess returns with a set of variables including dividend yields (or dividend yields in excess of the short term risk free rate), a credit (or default) spread, a term spread (long term minus short term rates) and the change in the short term risk free rate. Other authors have included the price to earnings ratio (P/E) ratio, the GDP, a ratio of consumption to wealth and other variables to forecast expected stock returns. For years, academics were satisfied predicting models using a set of lagged variables. Until recently, when two econometric challenges have been raised renewing the interest in this kind of studies. The first one, concerning the bias in the significance of the coefficients due to the presence of autocorrelation or non-stationary. The second one, referring to the poor results out-of-sample.

Each predictive variable presents different features and captures different components of the equity risk premium. The dividend to price ratio is considered a better forecast for long term horizons[1] and variables, such as the short term interest rate, provide better results at shorter terms. We have found in the literature a general conclusion stating that longer horizons are more forecastable, it is common to use in empirical studies quarterly or annual time series and test for forecasting horizons of several years. The exercise of forecasting shorter horizons is a more difficult task.

Equity return predictability has been largely studied for developed markets , particulary the US stock market. Authors as Ratio et al. (1988), **?** or Fama and French (1989) have focus on Developed Markets. More recently, some papers have studied Emerging equity markets such as the work of Harvey (1991), Bekaert and Harvey (1995), or van der Hart et al. (2003). They have found some evidence of predictability for those markets but the opposite has been argued by **?** when transaction cost are taken into account.

The aim of this chapter is to analyze the predictability of equity risk premia in several

[1]See **?** or **?** for more details

emerging equity markets using actual econometric techniques, we also include the US market for comparative purposes. The challenges to accomplish this task are multiple. First of all, we wish to follow a fundamental approach which means that we limit our framework to financial theory and from it we make the choice of our variables. Second, we have to face several econometric inflictions. Third, we have to deal with the imperfections of emerging markets data as well as with the fact that only short samples are available. And finally, we have to consider that the economic conditions have changed substantially during the last years making more difficult to find patterns that are true for every sample.

We focus on an equilibrium relation of the combined selected variables and we analyze the impact of this error component on the dynamics of the equity risk premia. Our work presents several innovations to the literature. We combine a set of fundamental variables with variables that, to our best knowledge, have rarely been studied as forecasts of excess returns such as the implied volatility[2]. And we focus on the changes of equity risk premia rather than the risk premia variable itself in order to include the maximum possible information and to respect the econometric restrictions of the model.

The chapter proceeds as follows. Section 2.3 presents a brief literature review on the topic of forecasting of equity risk premia. We focus on the most important studies presented in the academic literature. Section 2.5 describes the different theoretical models that we use in this study. We considered an error correction model (ECM) where the endogenous variables are the dividends to price ratio or dividend yield (DY), the growth rate of dividends (G) and the implied volatility index (VIX) from the S&P 500 index. Before starting the cointegration analysis, a description of the data was presented and the stationarity of the series was analyzed with standard tests of stationarity in section 2.6. After analyzing stationarity, we tested in section 2.7 if there are cointegration relations with the method of Engle and Granger (1987) and with the rank tests developed by **?**. The cointegrating vectors were analyzed and an ECM model was estimated. Section 2.8

[2]See Bollerslev and Zhou (2007a)

presents some conclusions for this chapter.

2.2 Literature Review on Comovements

2.3 Studies on the prediction of equity risk premia

Looking back to Graham and Dodd (1934) or Gordon's dividend growth model (1962), we remind the long history of predictability of equity returns based on earnings and dividend figures. However, after more that 70 years of financial research about the predictability of excess returns, academics and practitioners have changed their minds countless times.

Fama and French (1989) and Ratio et al. (1988) were pioneers in quantifying the relation between valuation ratios and future returns, finding that the ratios of price to dividends or price to earnings were positively related with equity excess returns and had some predictive power. Fama and French found that lagged earning or dividend yield are useful variables to forecast long term returns. Campbell and Shiller suggest that lagged moving average of earning should be used as return forecasts.

Later, a number of studies have been published arguing that other variables, different from financial ratios, can also account for some predictability. **?** and Hodrick (1992) showed that some functions of the short term risk-free rate can help to forecast equity returns. They have used the change in short term interest rates as a predictive local variable. They have use a stochastic detrending method on this variable to ensure stationarity. **?**, **?** and Fama and French (1989) used different yields of short and long term government or corporate bonds instead. Ferson and Harvey (1999) introduced the credit risk premium variable as a predictor and finally, **?** found that the consumption to wealth ratio could also be used for the same purposes.

The validity of all these studies can be questioned if we consider the results of the work by **?** or **?** who showed that the persistence in some variables has been the reason why they found statistically significant coefficients. More recently, the work of Lewellen (2004), Tourous, Valkanov and Yan (2004) or Campbell and Yogo (2004) suggested methods to

correct this bias. Other econometric techniques such as Newey and West (1987) statistics can also be used to correct the bias caused by autocorrelated or heteroscedastic regressors. A number of studies have therefore used multivariate regressions estimated with robust standard errors as the work of **?**, **?** or Hansen and Rodrick (1980).

Even if the dividend to price ratio has been reported in many studies as the most successful variable to forecast long run returns, authors such as **?** and Goyal and Welch (2003) have reported little robustness in the predictive power of dividend yields.

In the previous chapter, we have studied the relevance of including global or local factors to explain equity returns. This has been the concern of the asset pricing literature based on the CAPM model and its international version for foreign equities. As we showed in Chapter 1 and as it was evidenced by Harvey (1991) or Rouwenhorst (1999), Emerging equities are not explained totally by global factors, this observation being strongly related with the lack of integration of Emerging Markets. We believe that the inclusion of local factors is therefore required. Moreover, **?** analyzing a set of international developed markets have found that US instruments could be strong predictors of foreign equity returns. Here, we will explore if some US factors could be used as predictors for emerging equity markets as well.

The analysis of the results contained in all these years of academic research in financial economics makes us conclude that two points have to be carefully treated. First of all, the selected variables to be part of the model have to be consistent with financial theory. Second, the variables have to respect a given number of statistical properties in order have results that are not questionable by the econometric theory.

Our study follows the framework of the classic dividend discount model but we are also influenced by modern studies in the sense that we include other financial variables that we think should capture the variation in equity risk premia. Concerning the econometric literature referenced in this study, we follow the cointegration tests presented in Engle and Granger (1987), nonetheless, we focus mainly in the error correction model of **?**. We also

use standard test, such as Augmented Dickey-Fuller test (ADF) and Philipps and Perron test (PP), to determine the stationarity of the variables.

2.4 Methodology

2.5 Methodology

2.5.1 Explaning Expected returns

Our starting point is the dividend discount model (DDM) of **?** with constant discount and dividend growth rates, which states that the price of a stock is a function of the future dividend D, the expected growth rate of these dividends G and a discount rate R, defined as:

$$P = \frac{D}{R - G} \tag{2.1}$$

This model is derived from the premise that the fundamental value of a security can be defined as the present value of all the future cash flows associated with that security which is a usual consideration in financial asset pricing theory.

According to this present value model, if we write the equation for the returns, we have:

$$R = \frac{D}{P} + G \tag{2.2}$$

Returns then emanate from a capital gain component (D/P) and a growth component G, or both. This approach extracts the expected return implied in a DDM by solving the model at present price of the equity instrument and its dividends.

Many variables have been used in the financial literature to capture the effects of capital gains and growth. Practitioners have longly debated whether the income component or the growth component command returns during different economic cycles. For the capital component, the most common variables used to explain equity returns are the dividend yield or the earnings yield. The growth component can be estimated using an accounting statement. The sustainable growth rate of dividends should be the firms return rate of

reinvested capital (r^*) multiplied by its retention rate (equivalent to (1-dividend payout))[3]. The return on equity (ROE) can be used as an estimate of rate of reinvested capital and this way the growth component can be calculated by[4]:

$$G = ROE(1 - \frac{D}{E}) \tag{2.3}$$

The retention rate being the part of earnings not distributed in the form of dividends implies that the growth rate of dividends should be a function of the part of capital reinvested . A firm that distributes large amounts of dividends to its shareholders should not have high perspectives of growth.

The gross return for the period t to $t + H$ is defined as $R_{t,t+H} = \frac{(P_{t+H}+D_{t+H})}{P_t}$ and the continuously compounded return as $r_{t,t+H} = log(R_{t+H})$. The equity risk premia π is therefore calculated substracting the risk free rate, such that:

$$\pi_{t,t+H} = log(P_{t+H} + D_{t+H}) - log(P_t) - r_t^f \tag{2.4}$$

where the risk free rate should correspond to the rate of a debt instrument available for the same period t to $t + H$. The equity risk premia π is the variable of interest in our forecasting study.

If we use one observation every period of size H, we will need a long sample to test our model. However, it would be desirable to use all the information available at each date, and for this reason, we find inappropriate to limit the information to the dates corresponding to the chosen forecasting horizon. In the case of emerging equity markets , fundamental data is not very informative before the year 2000 making difficult to use old information. For these reasons, we suggest to use overlapping returns as a way to increase the number of observations but keeping a reasonable forecasting horizon. Some problems

[3]The retention rate has previously been used as a predictor of equity returns in the work of Lamont (1998). The dividend payout being the proportion of earnings distributed in the form of dividends, here denoted D/E

[4]Detailed calculation of growth rate is provided in the Appendix of this chapter

arise from this approach because if the frequency at which we calculate the returns is shorter than the return horizon, we have the presence of autocorrelation. To manage this difficulty, we suggest to axplain the change in the equity risk premia for an horizon H, calculated at a frequency h, where $h < H$. We can define this change in the risk premia variable as:

$$\Delta\pi_{t,h,H} = \pi_{t+h,t+h+H} - \pi_{t,t+H} \tag{2.5}$$

The risk premia variable (or its variation) has much higher risk than the dividends or growth figures. The part of variation explained by these two factors is then limited and therefore it would be relevant to include other variables that could explain the variations in the risk premia in the short run. Introducing a volatility variable (V), the model will have the following form:

$$R = \frac{D}{P} + G + V \tag{2.6}$$

? presents similar results assuming that the logarithm of dividend yield and the logarithm of the growth rate are random variables distributed normally.

2.5.2 Cointegration

Engel and Granger tests

A first approach to analyze cointegration is to use the two steps method of Engle and Granger (1987), the EG test. According to this approach, if the variables are cointegrated of the same order, then for a couple of variables y_t and z_t integrated of order one (or I(1) variable)[5] can find a cointegration relation of the form:

$$y_t = a + bz_t + u_t$$

[5]the first difference of these variables being stationary

where the $\hat{u_t}$ term, estimated with OLS, has to be stationary. The test of stationary in this case is an ADF test with critical values tabulated by **?** as the test is operating on the residuals of an estimated model. The stationarity of the variables y_t and z_t is evaluated with standard Dickey-Fuller (ADF) and Phillips-Perron (PP) tests which have the null hypothesis non-stationarity.

The drawback of the EG method is that for cointegration relations of more than two variables, it cannot determine if there is only one possible cointegration relation or if other linear relations are possible to find among the cointegrating variables. To remedy this problem, we use the test of Johansen (1988) and **?** allowing for multiple cointegration relations.

Cointegration VECM

The second approach to evaluate the cointegration of a set of variables is the Vector Error Correction Model (VECM) of **?**. A VECM representation is considered when a set of k variables integrated of order one (or I(1) variables) could be combined linearly to create a stationary variable, meaning that in the long term the variables in levels have a stable relation described by the cointegrating vectors. Transitory shocks have no permanent effect on the cointegrating combinations of variables.

A Vector Auto Regressive (VAR) model with p lags composed of k I(1) variables could be set up as:

$$y_t = \Pi_1 y_{t-1} + ... + \Pi_p y_{t-p} + u_t \tag{2.7}$$

where $y_t = (y_{1t}, ..., y_{kt})'$ is a vector of k time series at time t, $u_t = (u_{1t}, ..., u_{kt})'$ is a white noise process with covariance matrix Σ_u and the Π_i (kxk) matrices contain the coefficients of the lagged endogenous variables.

The model presented in equation 2.7 has an equivalent specification, according to the Granger representation theorem (See Engle and Granger (1987)), if the variables from the

VAR are cointegrated, having the following form:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + ... + \Gamma_p \Delta y_{t-p+1} + \Pi y_{t-p} + u_t \quad (2.8)$$

$$\Gamma_1 = -(I - \Pi_1 - ... - \Pi_i) \text{ for } i = 1, ..., (p-1) \quad (2.9)$$

$$\Pi = -(I - \Pi_1 - ... - \Pi_p) \quad (2.10)$$

where I is the $(k \text{x} k)$ identity matrix, the Γ_i matrices correspond to the coefficients measuring the effects of the lagged differenced variables.

The specification in equations 2.8 to 2.10 correspond to a long term form because the levels of y_t enter the VECM with p lags, as opposed to a transitory form where y_t enters with one lag and the Π_i matrices capture transitory effects and not long run impacts.

Before estimating the VECM, it is important to determine the number of lags to include in the regressions and to determine whether or not to include a linear trend in the model. The number of lagged differences of the endogenous variables, $p-1$, are chosen with the help of model selection criteria applied to the VAR model. The information criteria that can be used are AIC, HQ, SC and FPE[6]. A likelihood ratio test is used in order to determine if a linear trend should be included in the VAR (See **?**). The null hypothesis of not including a trend is tested against the alternative hypothesis stating that a combination of linear trend slopes should be zero.

We consider the maximum eigenvalue test as well as the trace tests statistic described in **?** in order to determine the number of cointegrating relations among the variables in vector y_t. The rank of the matrix Π is evaluated to determine the number of eigenvalues that are different from zero. These cointegration tests are formulated in terms of the

[6]FPE: Final prediction error, AIC: Akaike information criteria, SC: Schwartz information criteria, HQ: Hannan-Quinn information criteria

estimated eigenvalues $\hat{\lambda_i}$ of the matrix Π as:

$$\lambda_{trace}(q) = -T\Sigma_{i=q+1}^{k} ln(1 - \hat{\lambda_i}) \tag{2.11}$$

where q is the number of cointegrating vectors under the null hypothesis.

The trace tests determine whether the number of cointegrating vectors is less than or equal to q. The maximum eigenvalue tests if the number of cointegrating vectors is q against a number $q + 1$.

$$\lambda_{max}(q, q + 1) = -Tln(1 - \hat{\lambda_{q+1}}) \tag{2.12}$$

The critical values for both statistics are presented in **?**. There would be cointegration in the cases where the rank of the matrix Π is a number between zero and k. If the rank of Π is zero, no linear combination exists to make Πy_{t-p} stationary. And if the rank of Π is k (Π has full rank), then the variables in y_t are stationary. Only when the rank of Π is a value q (rank(Π)=q) and $0 < q < k$, there are q cointegration relations. In this case, Π could be written as $\Pi = \alpha\beta'$, where β would be a (kxq) matrix with q linear independent columns representing the cointegrating vectors. Each cointegrating vector β would transform y_{t-p} into a stationary process $\beta_i y_{t-p}$. The term α would be interpreted as the amount of each cointegrating vector entering the VECM. The elements of α are the speed of adjustment terms to the long run equilibrium, α is referred to as the adjustment matrix.

2.5.3 Financial and econometric challenges

Our approach to forecast equity premia uses a combination of the fundamental discount model presented in section 2.5 and the econometric approach presented in section 2.5.2.

Equation 2.6 suggests a relation that explains equity returns by three different components: income, growth and a risk component. This equation cannot be validated em-

pirically but we can study a linear relation based on it that includes all these variables as well as a coefficient associated to each variable. Assuming that the relation is respected at each date up to an error term and considering each variable as a random variable, the model that we would like to test, applied to the risk premia variable, would be:

$$\pi_{t,t+H} = \delta_1 DY_t + \delta_2 G_t + \delta_3 V_t + \varepsilon_t \tag{2.13}$$

This equation can be estimated with OLS if all the variables are stationary (or integrated of the same order). The results for the tests of stationarity, in the next section, show that none of the three explaining variables are stationary[7]. We analyze the relation between these variables as an equilibrium relation that can only be validated in the long term. That is why, we rely on a cointegration model that would let us analyze long term as well as short terms dynamics of the equity risk premia.

The variables $\pi_{t,t+h}$, DY_t, G_t and V_t should be stationary. Economically, it has no sense to observe some of these variables with changing statistical features over time. Empirically, none of this variables is stationary, which means that during the observed sample, the distribution of their moments are time varying. The analysis of section 2.5 leads to the conclusion that there should be a relation among the dividend yield, the growth component of dividends and the volatility of equity returns. We analyze this relation with an error correction model, where we consider equation 2.13 as a long term relation and we study whether the correction term calculated from these variables has an impact on the differenced variable $\Delta\pi$.

In the ECM model, the variable $\Delta\pi_{t,h,H}$ would be explained by the lagged levels of the variables DY_t, G_t and V_t as well as their lagged differences. Therefore, we introduce the risk premia variable in levels. For this, we build a index representing the levels from the risk premia differences, that we note Π_t. The term $\Delta\pi_t$ from equation 2.5 can be transformed into an index series such that $\Pi_t = RP_{t-h}(1 + \Delta rp_{t,h,H})$, the initial level Π_0

[7]The results are presented in the appendix.

can be fixed arbitrarily to one. The variable Π_t should be integrated of order one, it would present a unit root but its differenced variable should be stationary. The error correction model would be studied by analyzing the first equation of a VECM.

We estimate the VECM with different input vectors. The basic model is estimated with an input vector y_t defined as $y_t = [\Pi_t, DY_t, G_t, V_t]$. Then, other specifications are analyzed and compared to the initial model.

2.6 Data set description

2.6.1 Samples

All the data used in our empirical analysis are downloaded from DATASTREAM[8]. We study the period starting on January of 2000 to the end of June 2006. We have DATASTREAM Index prices in local currencies for a set of Emerging Markets and the S&P500 Index for the US market which we have included for comparative purposes. We have chosen a set of Emerging Markets from the Latin American region composed of five Latin American countries: Argentina, Brazil, Chile, Colombia and Mexico. For each country in the sample, we have the variables dividend yield, growth, interest rate, high yield and implied volatility. For the empirical analysis we focus only in one emerging country, we use Chile as an example and we compare the results to the US market.

2.6.2 Equity Returns and risk premia

We calculate every week the changes in the risk premium variable for a return horizon of 3 months as described in equation 2.5. The local risk free rate used to calculate the risk premium is specific to each country, we use the available deposit corporate rate for emerging markets and for the US risk free rate we use the three months constant maturity US Treasury Bill rate. The exercise of forecasting equity markets has been documented to have more success for long term horizons. The challenge with emerging equity markets is that the fundamental financial data has a very short history (usually the data set start in 1990) or sometimes the series available have very doubtful quality. For this study we have a data set starting in 2000 to mid 2006, before 2000 the series are blemished. This makes a sample 26 observations of quarterly returns, 66 observations of monthly returns or of 324 observations of weekly returns[9]. As presented in section 2.5, it would be important

[8]Datastream international

[9]We have chosen a forecasting horizon of 3 months (Longer than 3 months would be difficult to test with the data available and shorter period have been reported to be very difficult to forecast) to calculate the returns but 22 observation are not enough.

to use all available information. Therefore, we have analyzed time series of three months returns calculated every week which means a series of 324 observation. And in order to eliminate the problems of autocorrelation due to the fact of using overlapping returns, we have focused on the changes of this series instead of its level. In the following paragraphs, we describe briefly the main feature of each type of variable in our data set.

Dividend yield

The dividend yield of a stock represents how much a company pays out in dividends for a year relative to its share price. It measures the return in the form of cash flow for every investment in this equity security. For a country index the dividend yield is calculated from all the stocks composing the index. A variable taht is often used instead of the dividend yield is the earnings yield which also contains information that could be used by the investors to valuate equity securities or indexes, this variable measures the global return of an equity investment not only in the form of dividends distributed to stockholders but also in the form of capital reinvested in the company. The average dividend over the last 30 years for the S&P 500 index is close to 6.7% and the average earnings yield close to 7.8%. But the last years the average is close the dividend yield was close to 2%. Whereas for Emerging Markets this figure is close to 2.5%. It is probable that for a short term horizon, the dividend yield would have more importance than the earnings yield because the changes in the equity risk premia could be explained by the changes in the profitability of the company as well as the management decision to distribute capital.

Growth rate of dividends

The growth component in the forecasting model could be calculated from historical dividends or estimated with the accounting approach presented in equation 2.3 and developped in the Appendix. In the case of emerging equities the estimation is usually difficult as there is sometimes a lack of information. Some studies use a function of the country

GDP to estimate the future growth of local companies. And important consideration is that actual growth should have poor relevance in explaining future returns but expected growth measures would be important in determining equity returns.

Volatility

In addition to fundamental components, we have introduced a measure of volatility capturing the shorter term variation in equity returns. Two options have been considered: a realized volatility and an implied volatility variables. The lagged realized volatility stock market volatility is found to be positively and significantly related to the ex ante equity premium. An estimate of the realized volatility is the historical return volatility of the S&P500 index from daily returns[10] calculated over trailing intervals of 90 days. **?** has suggested to use implied volatility of options on a market portfolio as a direct forward looking estimate of the market variance, and hence of its risk premium.
The implied volatility has been studied intensively as a predictor of the realized volatility.[11] But few studies have used it as a predictor of expected returns. In a recent study, Bollerslev and Zhou (2007b) have used a similar approach, they have introduced the difference of volatilities to forecast US equity returns. Whaley (1993, 2000) refers to the implied volatility variable as the "investor's fear gauge". In fact, the implied volatility represents an ex ante market assessment of risk that captures the investor's sentiment on the market, it should therefore be related to the expected equity returns. We use the Chicago Board Option Exchange (CBOE) implied volatility index which corresponds to the volatility implied from S&P Index option prices[12].

We compare these two measures of volatility but also some functions or combinations of these measures of dispersion. We analyze the behavior of the ratio of implied to historical volatility over time, expecting that it could be a proxy of the risk premia for the S&P500

[10]Calculate as net log returns with closing prices for the index.
[11]See Corrado and Miller (2005), Christensen and Prabhala (1998), Blair, Poon, and Taylor (2001) or Fleming (1998)
[12]Hypothetical option with 30 days to maturity

index. We observe that the implied volatility for the S&P500 index is always higher than a historical 90-days volatility, the ratio of these two variables is on average equal to 1.3. We also analyze a logarithmic transformation of the implied volatility variable that will smooth the jumps in the level of the implied volatility. The different measures of volatility are presented in the Appendix. We decide to use the implied volatility to introduce a market component to the model.

Other potential variables

Another relevant component in explaining equity returns variation is the leverage component. Companies can borrow to finance their growth and for this reason the variation in interest rates can have an impact on equity returns.

The credit risk premium variable is very often used in asset pricing models. It could be measured as the spread between a high yield corporate bond index and a lower yield bond or government bond index[13]. As stated by Chen, Guo, and Zhang (2006) Chen et al. (2006), debt and equity are financial claims written on the same corporation, and therefore they must share the same systematic risk. The credit yield premium should therefore be related to the risk premia for equities. We tested a variable constructed as the high yield for corporate bonds over a US treasury bond spread as measure of credit risk premium. The credit risk premium should be negatively related to the equity returns. This spread also measures the sentiment of the market as investors would price very differently companies with different levels of credit risk during periods of higher risk aversion. Mody and Taylor (2003) have found evidence that the high yield spread can predict the real economic activity of the US. In our sample, it is difficult to see the benefit from including this variable because it is highly correlated with the implied volatility variable. Multicollinearity can result in numerically unstable coefficient, for this reason we exclude this variable from our model even though financially it has a meaning to include it.

[13]Some authors use the difference between Moody's Baa and Aaa corporate bond as credit spread (See Ferson and Harvey (1999) Ferson and Harvey (1999))

2.6.3 Descriptive statistics

We have first calculated descriptive statistics of the excess returns for the set of Emerging Countries Indexes and the US market index. Some Latin American markets present very low excess returns due to the very high levels of interest rates during the years 1998-2000, it is the case for Brazil, Mexico and Colombia where nominal interest rates have reached levels higher than 30% during this period. Other emerging countries in other regions also suffered with this problem. Those markets are therefore unattractive during these periods of high interest rates and very low excess (or even negative) returns in local currencies. Before 2000, it would be difficult to test a fundamental model as the one presented here. We reduce the sample to focus on the most recent periods, we analyze the six last years period (years 2000 to 2006) and the three last years period (years 2003 to 2006). The results are presented for the period 2000 to 2006 in table 2 -A in the appendix. The tables also present skewness, kurtosis as well as ADF statistics for each series.

We observe in figures 2 -A and 2 -B the input variables for the US and Chile. The variables price level P, dividend yield DY, earnings yield EY, growth rate of dividends G and short term interest rate I are specific to each country and the variables high yield spread HY and implied volatility VIX are measured for the US markets and could be considered as global factors having an impact on international returns. The first variable on the top left corner of these figures is the country index. All emerging countries present an upward trend for the whole period and it is the case for Chile that is the country that we used as an example. While for the US market, the period of study include the technology bubble with its rally until 2000 and the crash until 2003, the period 2004 to 2006 has been a continuous bullish period for the US market.

We also observe dividends and earnings figures. For the US market the average dividend yield for the period 2001 to 2006 is 1,6% and the average earning yield is 5,7% making an average payout ratio for the US companies of 29%. For the set of emerging countries, the same variables are observed at levels of 3,2% and 11,6% so that the divi-

dend payout of nearly 28%. For this sample, the earning and dividends paid by emerging companies are higher than the US.

We have also calculated descriptive statistics and correlations for the set of factors of each country in the appendix of this chapter. The results are presented in table 2 -B and 2 -D for the US and Chile. We note that the correlation coefficient is very high between the volatility variable and the high yield variable, this observation has no economic interpretation but can produce multicollinearity problems if both variables are used as predictors. The same is true for dividends and earnings. We keep the volatility variables and the dividend yield in our models, but alternative configuration are possible analyze with earning or high yield figures.

2.6.4 Stationarity tests

The price level process of an equity market is not stationary, however its returns usually are. Normally, for shorter horizons of returns, the tests of stationarity are easily confirmed. The earnings yield and dividend yield variables should also be stationary. In the dynamic growth model of Campbell and Shiller (1988) and all the following studies on the subject, the dividend yield is assumed to be stationary. Even the growth rate of dividends should be considered stationary as there is no logical economic reason for a rate of growth to have changing distributional features. For the implied volatility, it is usual to observe clusters or jumps in the process but the means have usually a mean reverting nature ensuring that the series remains stationary. Theoretically all these variables should be stationary, nevertheless in practice the validity of this hypothesis would depend on the sample used for testing it.

In our data set, we analyze the stationarity of the variables risk premia Π, dividend yield DY, growth G and implied volatility VIX. The series are shown in levels and differences in figure 2.1. We observe the persistent behavior of the risk premia Π variable. This variable in differences corresponds to the change in the equity risk premia for an

horizon of three months and it is stationary. The results for the ADF tests of stationarity for the US market are presented in table 2.1. For the ADF tests we used the Akaike Information Criterion (AIC) to determine the relevant number of lags to include in the autoregressive model. The analyzed series are integrated of order one according to this test.

Figure 2.1: Levels and differences: RP, DY, G and VIX for the US (2001-2006)

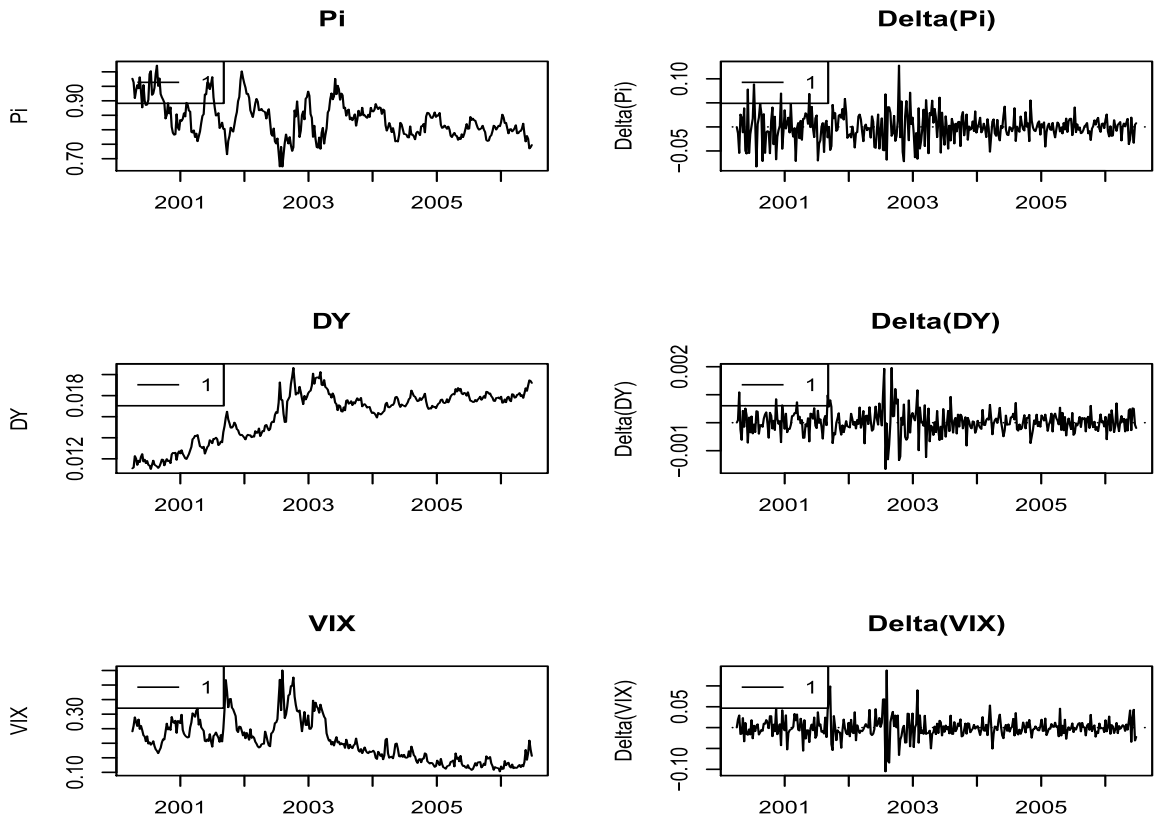


Table 2.1: Stationarity

	Lev	Dif
Π	−0.79	−5.33
DY	0.78	−7.48
G	−1.02	−8.48
VIX	−1.12	−11.12

The same analysis has been run for Chile. The series are presented in figure 2.2 and

the ADF test in table 2.2. We observe the same features in the case of Chile than the US market.

Figure 2.2: Levels and differences: RP, DY,G and VIX for Chile (2001-2006)

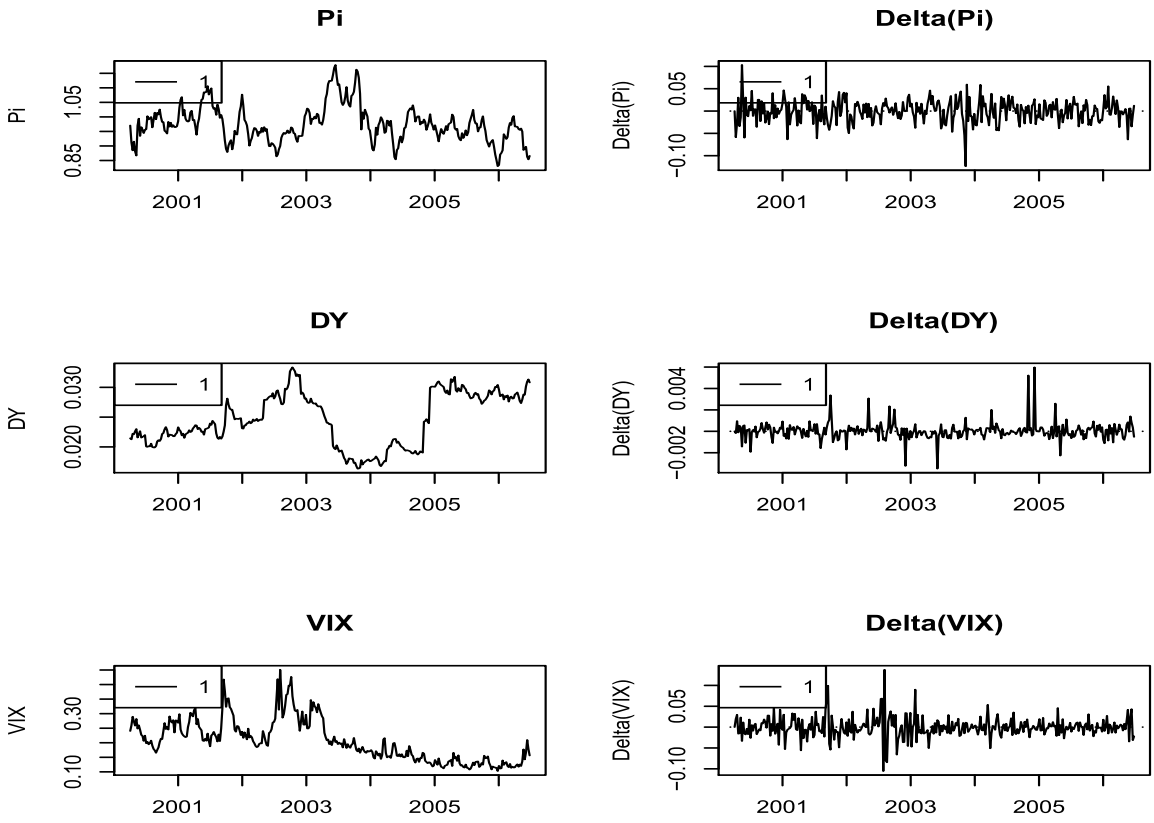


Table 2.2: Stationarity

	Lev	Dif
Π	−0.28	−8.25
DY	0.35	−16.92
G	0.31	−8.20
VIX	−1.12	−11.12

We conclude from this analysis that the different time series in levels are all non-stationary but their first differences can be considered stationary as they reject the null

hypothesis of the presence of a unit root.

2.6.5 Non Stationarity of stock valuation indicators

The present value model of As the dividend yield took on extreme and historically unprecedented values several researchers argued that the long-run relationship between prices and dividends had broken down such that the dividend yield would exhibit non-stationarity. This inturn would imply that the present value model that underlies the belief in a stationary dividend yield is no longer valid. However, several researchers have argued that this conclusion only holds from a linear perspective and that once non-linear dynamics are allowed within the price and dividend relationship then stationarity is reported.

several researchers have identified non-stationarities within the log dividend yield (see, for example, Froot and Obstfeld, 1991; Lamont, 1998; Balke and Wohar, 2002), which indicates possible non-cointegration of prices and dividends.

bubbles

2.7 Main Results

This section presents the main results of the cointegration analysis, we have taken as an example only one emerging country, Chile and we have compared the results with the US market case. For this two countries we develop a complete analysis of cointegration, then we present the results from the estimation of the model for other emerging countries.

For each country we have determined the number of lags to include in the VAR, we have tested the rank and maximum lambda statistics to determine the number of cointegration relations within the data set and we have calculated likelihood ratio statistics, before estimating the VECM, to determine if a linear trend should be included in the model. We present the results for the US market and Chile. The results for other countries are presented in the Appendix.

To estimate the VECM, we used the model defined in equation 2.8 with the vector y_t including the variables in levels for the risk premia Π, the dividend yield DY, the growth rate of dividends G and the implied volatility VIX. Then we have compared the results for this model with different specifications that exclude some of its components. From the comparisons we can see that the error correction term from the basic model plays an important role in explaining the changes in the equity risk premia. We also test the robustness of the basic model with the estimation of the model for different samples.

2.7.1 Cointegration analysis

Long term relation

To begin with, we explore the existence of cointegration relations among the variables. A first test of cointegration has been done running the EG-test with the set of variables in our system. For the set of US variables, the EG-test confirms the presence of a cointegration relation at 1% with a specification that includes a linear trend. Moreover, running the rank tests from **?** we also find one cointegration relation among the variables.

We have calculated different criteria (AIC, SC and HQ) to determine the number of lags to include in the corresponding VAR model, the details are presented in the Appendix in table 2 -C. While AIC and FPE writeria suggest to include 6 lags in the VAR, SC suggests 1 lag and HQ suggests 3 lags. We decide to estimate the US model with one lag after testing for autocorrelation, heteroscedasticity and normality of the residuals of different lag specifications[14].

We present the trace and the maximum eigenvalue statistics to determine the cointegration rank. Table 2.3 in this section presents the results. According to the maximal eigenvalue statistic (lambda max), the hypothesis of no cointegration can be rejected at 1%. The trace test statistic indicates a cointegration rank of 1. Therefore, there is one cointegration relation within the set of variables according to these rank tests.

Table 2.3: Rank Tests

	Trace test	10pct	5pct	1pct	Lambda max	10pct	5pct	1pct
r <= 3 |	2.43	6.50	8.18	11.65	2.43	6.50	8.18	11.65
r <= 2 |	14.05	15.66	17.95	23.52	11.62	12.91	14.90	19.19
r <= 1 |	33.25	28.71	31.52	37.22	19.20	18.90	21.07	25.75
r = 0 |	58.36	45.23	48.28	55.43	25.11	24.78	27.14	32.14

We conduct a likelihood ratio test to determine if a linear trend should be included in the VAR. In this test, the null hypothesis is for not including a linear trend. The test statistic is distributed as χ^2 with $(k-r)$ degrees of freedom. The null hypothesis is not rejected for this model. We can therefore estimate the VECM without any deterministic terms.

The estimated cointegrating vector for the US model, displayed in table 2.4, suggest that the variables, dividend yield DY, growth rate G and implied volatility VIX, are

[14]We used different tests on the residuals: Goldfeld-Quandt and Breusch-Pagan tests for heteroscedasticity, Durbin-Watson and Breusch-Godfrey for autocorrelation and Jarque Berra test for normality.

differently related to the levels of the equity premium variable Π. The coefficient related to the dividend yield is positive and has a value close to 12.5 for the US market. The coefficient for the implied volatility is negative and has a value close to -0,35. That means that a positive change in the level of the implied volatility would have an negative impact of one tenth in the level of the risk premium variable. And a positive change in the level of the dividend yield would have an impact of 12 times on the equity risk premium. We expected to have a positive relation for the dividend yield, a higher dividend yield has to be reflected in a higher excess return, however a negative relation for the volatility variable is not economically consistent as investor should demand for higher compensation as the level of risk increases. However, our volatility variable is an implied volatility variable which captures the risk aversion of investor's feelings about the market, it could be considered as an ex-ante measure of risk. At this step of the analysis, we care about the long term relation of these variables observed at the same moment, and not the impact of the level of a variable on subsequent levels of the other variable. The estimated coefficient is negative but close to zero which means that the negative relation between these variables is not very strong.

Table 2.4: Cointegrating vectors

	Chile	U.S.A
Π	1.00	1.00
DY	7.92	12.52
G	−1.07	0.03
VIX	−0.35	−0.11

Model estimation

The estimates of the VECM model are found in table 2.5. We observe from this table that there are significant coefficients for the error correction term, the constant term

and the lagged value of the first difference risk premia variables. The coefficient for the error correction term is negative with a value of -0.128, this ensures that there would be an adjustment to the long term equilibrium. The coefficient for the lagged differenced risk premia variable is also negative and significant but only at 5%. The coefficients for the other lagged variables in differences are not significant. We have also analyzed the adjusted R^2 of this regression which is close to 5.1%. This result seems unsatisfactory but looking at the results of similar studies about explaining and forecasting equity markets returns, we find similar levels in the coefficients of determination. In **?** the R^2 for yearly data regressions for the US are close to our results including only price to dividend figures. It is important to note that we use a smaller horizon for the calculation of returns and the task of explaining and forecasting equity returns becomes harder with more frequent returns even if the number of observation increases.

Table 2.5: VECM for Chile and US (2000-2006)

	Chile	US
ect1	-0.091^{***}	-0.128^{***}
	(0.022)	(0.030)
constant	0.089^{***}	0.130^{***}
	(0.022)	(0.031)
$\Delta\Pi_{t-1}$	0.017	-0.163^{*}
	(0.057)	(0.072)
ΔDY_{t-1}	-3.840^{*}	4.504
	(1.745)	(5.463)
ΔG_{t-1}	-0.069	0.137
	(0.201)	(0.301)
ΔVIX_{t-1}	-0.050	-0.099
	(0.055)	(0.088)
N	324	324
R^2	7.2%	6.9%
adj. R^2	5.4%	5.1%
Resid. sd	0.024	0.028

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

The same analysis has been run for different emerging countries. We have focused our analysis on Latin American equity markets and here, we present the results for Chile, the country that we have chosen as an example. The economy of Chile has presented high growth rates during the 90's, a recession period after 97 and a recovery after 2001. During the period of analysis, this economy performed better than the average in the region, becoming one of the most competitive countries in Latin America. And the stock exchange of Santiago has shown active trading during the last years.

We proceed in the same way than for the US market, first we determine the presence of cointegration relations and then we focus in the estimation of the model. The VECM for Chile has to be estimated with one lag according with our analysis, presented in table 2 -E in the Appendix and based on the HQ and SC criteria. No linear trend has to be included according to the likelihood ratio tests. We have also applied the Johansen procedure to determine the number of cointegration relations. From the rank tests, we can

see that only one cointegration relation is identified. The maximum eigenvalue statistic let us reject the null hypothesis at 1% significance and the trace test confirms that there is only one cointegration relation. This results are found in table 2 -G.

Table 2.6: Rank Tests

	Trace test	10pct	5pct	1pct	Lambda max	10pct	5pct	1pct
r <= 3 |	1.71	6.50	8.18	11.65	1.71	6.50	8.18	11.65
r <= 2 |	7.53	15.66	17.95	23.52	5.82	12.91	14.90	19.19
r <= 1 |	27.54	28.71	31.52	37.22	20.00	18.90	21.07	25.75
r = 0 |	53.22	45.23	48.28	55.43	25.68	24.78	27.14	32.14

The coefficients of the cointegrating vector are presented in table 2.4. Similarly to the US market case, there is a positive coefficient for the dividend yield and a negative one for the implied volatility variable. However, the coefficient for the implied volatility is close to zero and the coefficient related to the dividend yield variable has a value of 7.9 which is lower than the US market figure. For the growth variable, the coefficient is negative with a value of -1.07, we didn't expect a negative coefficient for this variable as the growth rate should be positively related the equity risk premia.

The estimated coefficients for the VECM for Chile are also presented in table 2.5. We observe again a negative coefficient for the error correction term that is significant at 1%. The constant term is also significant as it was for the US. However, the lagged value of the differenced risk premia is not significant in the case of Chile and the coefficient for the lagged differenced dividend to price ratio is significant at 5% which was not the case for the US model. The adjusted R^2 for this model is 5.4%. We haven't found any similar study to compare our results for Chilean example.

2.7.2 Analysis for different emerging markets

From the previous results for Chile and the US market, we found some similarities in the VECM estimation results. We have run the model for other countries from the Latin American region where we found available data. We included the results for Argentina, Brazil and Mexico. We compare the results for the cointegration vectors for the other Latin American countries in table 2.7. The dividend yield coefficient is also positive for Mexico but unexpectedly, it is negative for Argentina and Brazil. However, for these two countries the coefficients corresponding to the growth and implied volatility variables are all positive. As a result, we can not identify a clear pattern for the coefficients of the cointegration relations for the countries of Latin America.

Table 2.7: Cointegrating vectors

	ARGENTINA	BRAZIL	CHILE	MEXICO	U.S.A
P.l2	1.00	1.00	1.00	1.00	1.00
DY.l2	−8.97	−7.71	7.92	7.12	12.52
G.l2	0.81	2.65	−1.07	2.68	0.03
VIX.l2	1.93	2.86	−0.35	0.54	−0.11

The results from the VECM estimation are presented in table 2.8. We observe that all the coefficients for the error correction term are negative which means that the model is stable across countries, this coefficient represents the speed of adjustment to the long run equilibrium and we observe that it is stronger for the US with a value of -0.129. The coefficients of the error correction term are significant at 1% for all countries with the exception of Brazil for which the associated coefficient is only significant at 5%. The constant term in the VECM has a positive coefficient with a positive value close to 0.08 and significant for all the countries. The results are mixing for the lagged differenced variables.

The lagged differences of the risk premia variable $\Delta\Pi_{t-1}$ is negative for all countries but only significant for the US. The coefficients for the lagged differences of the dividend yield variable ΔDY_{t-1} is negative and significant for Argentina and Chile, slightly positive for Brazil and highly positive but not significant for Chile. And the coefficients for the lagged differences of the volatility variables ΔVIX_{t-1} are slightly negative for all Latin American countries and positive for the US but not significant for Brazil. The adjusted R^2 for all countries are close to 5% with the exception of Brazil which has a R squared close to zero.

Table 2.8: VECM for different countries (2000-2006)

	Argentina	Brazil	Chile	Mexico	US
ect1	-0.054^{***}	-0.023	-0.091^{***}	-0.122^{***}	-0.128^{***}
	(0.015)	(0.014)	(0.022)	(0.029)	(0.030)
constant	0.049^{***}	0.037	0.089^{***}	0.189^{***}	0.130^{***}
	(0.014)	(0.023)	(0.022)	(0.046)	(0.031)
$\Delta\Pi_{t-1}$	-0.024	-0.010	0.017	-0.079	-0.163^{*}
	(0.062)	(0.060)	(0.057)	(0.065)	(0.072)
ΔDY_{t-1}	-3.006^{*}	0.128	-3.840^{*}	-0.688	4.504
	(1.307)	(0.985)	(1.745)	(2.875)	(5.463)
ΔG_{t-1}	0.038	0.299^{*}	-0.069	-0.110	0.137
	(0.099)	(0.144)	(0.201)	(0.191)	(0.301)
ΔVIX_{t-1}	-0.108	-0.104	-0.050	-0.072	-0.099
	(0.111)	(0.083)	(0.055)	(0.075)	(0.088)
N	324	324	324	324	324
R^2	0.062	0.028	0.072	0.054	0.069
adj. R^2	0.044	0.010	0.054	0.036	0.051
Resid. sd	0.048	0.034	0.024	0.031	0.028

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

2.7.3 Comparison to alternative models

The variable of interest in the model presented until now is the change in the equity risk premia for a given market. We have estimated a VECM indicating that the change in the equity risk premia variable $\Delta\Pi_t$ is explained by an error correction term and the lagged differences of the variables $\Delta\Pi_{t-1}$, ΔDY_{t-1} and ΔVIX_{t-1}. We suggest to compare this model with a simple autoregressive model and an autoregressive model augmented with the lagged differenced variables. We observe the results in table 2.9:

Table 2.9: AR US and Chile (2000-2006)

	US	US	Chile	Chile
(Intercept)	−0.001		−0.000	
	(0.002)		(0.001)	
$\Delta\Pi_{t-1}$	−0.102	−0.085	0.098	0.063
	(0.055)	(0.071)	(0.055)	(0.057)
ΔDY_{t-1}		6.831		−2.832
		(5.533)		(1.762)
ΔVIX_{t-1}		−0.100		−0.065
		(0.090)		(0.056)
N	324	324	324	324
R^2	1.0%	1.6%	1.0%	2.2%
adj. R^2	0.7%	0.7%	0.7%	1.3%
Resid. sd	0.029	0.029	0.024	0.024

Standard errors in parentheses

* indicates significance at $p < 0.05$

We can see that any of the variables in these autoregressive specification is significant and that the R squared and adjusted R squared of each regression are very low, even close to zero in many cases. The VECM is therefore a better model to explain the changes in equity risk premia and we can see that the error correction term plays an important role

in this model.

2.7.4 Robustness

In order to test if the results are robust to different samples, we analyzed the sample starting in January 2003 for the same set of countries. The results are presented in table 2.10. We observe that all the coefficients for the error correction term are negative and significant for the US at 0,1%, for Chile at 5% and only significant at 10% for Argentina, Brazil and Mexico. The coefficient associated to the constant term is close to zero for all the five countries and significant again for all countries except for Brazil. The coefficients for the constant term from the model estimated with the sample starting in 2000 were all positive contrary to this case. For the lagged differenced variables, there is no recognizable pattern in the results so we can not conclude anything about the effect of this variables on the differences of the risk premia variable. The adjusted R^2 of the regression is close to 6.4% for the US and 4.3% for Chile but close to zero for the other countries.

The main result from tables 2.8 and 2.10is that the error correction term is negative and significant for all the countries. The estimation of the VECM for this shorter sample confirms this results but with lower levels of significance. The model is in this way rather stable across the different samples. The negative coefficient for the error correction term confirms the presence of cointegration among the variables. Consequently, there is a long run equilibrium among the variables in the system which indicates that there is no presence of rational bubbles during the period of analysis.

Table 2.10: VECM for different countries (2003-2006)

	Argentina	Brazil	Chile	Mexico	US
ect1	$-0.030^{\dagger}$	-0.006	$-0.028^{\dagger}$	-0.034	-0.018^{*}
	(0.016)	(0.004)	(0.015)	(0.023)	(0.007)
constant	$-0.010^{\dagger}$	-0.063	-0.018^{*}	0.009	0.049^{*}
	(0.006)	(0.043)	(0.009)	(0.007)	(0.019)
$\Delta\Pi_{t-1}$	0.047	-0.114	0.095	-0.148	$-0.186^{\dagger}$
	(0.084)	(0.085)	(0.078)	(0.094)	(0.097)
ΔDY_{t-1}	-0.912	-0.854	$-4.095^{\dagger}$	-0.649	13.182
	(4.458)	(1.671)	(2.321)	(5.667)	(9.777)
ΔG_{t-1}	0.007	0.482^{*}	0.327	0.421	0.890
	(0.112)	(0.221)	(0.340)	(0.301)	(0.617)
ΔVIX_{t-1}	-0.144	0.034	-0.038	-0.142	-0.404^{*}
	(0.242)	(0.241)	(0.142)	(0.204)	(0.173)
N	167	167	167	167	167
R^2	3.9%	4.6%	7.8%	4.5%	9.8%
adj. R^2	0.3%	1.1%	4.3%	1.0%	6.4%
Resid. sd	0.045	0.042	0.025	0.034	0.025

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

2.8 Summary and Conclusions

We have analyzed a cointegration model for a set of Latin American countries and the US market in order to see if there is some evidence of predictability of the equity risk premia. Facing the difficulties of the imperfect data sets in emerging countries, we suggest a way to transform the usual equity premia into differenced series from overlapping excess returns in order to eliminate any problem of autocorrelation in the series. The suggested approach was not found in previous literature, therefore we have no very similar studies to compare our results. However, there is a large number of studies testing prediction models applied to non-overlapping series of returns. Staring from a fundamental approach, we selected a specification that includes the dividend to price ratio and the implied volatility of the S&P 500 index as forecasting factors. We relied on an error correction model as the input variables present unit roots.

The main result of our analysis is the finding of a cointegration relation within the selected set of variables, this relation is present in several markets. We arrive to this conclusion after running the rank tests of Johansen which are confirmed with the estimation of an error correction model having negative and significant coefficients for the correction term for some of the countries that we have analyzed. From the estimated model we can conclude that there is some evidence of predictability for this set of countries. We also tested a specification including a growth variable but even if the rank tests suggest the presence of cointegrating relations, the estimated models have no significant coefficient and poor adjusted R^2. The introduction of the implied volatility variable in our model helps to improve the prediction. This variable has rarely been used in the literature as a forecasting variable for equity returns.

Our study has some limits. First, we have imposed a given forecasting horizon and we haven't tested the results if we change this parameter. Second, we haven't analyzed the out-of-sample results of this prediction models and as presented by Goyal and Welch (2003), good results in-sample could vanish when tested out-of-sample.

2.9 Appendix

Growth and payout ratio relation

Let define:

$$\begin{aligned} E_t &= \text{Earnings at time } t \\ D_t &= \text{Dividends at time } t \\ I_t &= \text{Investments at time } t \\ d &= \text{Dividend payout ratio} \\ r^* &= \text{return rate of reinvested capital} \end{aligned}$$

The amount of earnings can be decomposed into the amount distributed in the form of dividends and the amount reinvested in the company:

$$E_t = D_t + I_t$$

This quantities can be described in terms of the dividend payout ratio of the company as:

$$E_t = E_t d + E_t(1-d)$$

The part reinvested will have a return of:

$$\begin{aligned} I_t &= I_t(1+r^*) \\ I_t &= E_t(1-d)(1+r^*) \end{aligned}$$

And considering that the only component of earnings that will provide a return is the part of reinvested capital, we can write:

$$\begin{aligned} E_{t+1} &= I_t(1+r^*) + D_t \\ E_{t+1} &= E_t(1-d)(1+r^*) + D_t \\ E_{t+1} &= E_t + E_t(1-d)(r^*) \\ E_{t+1} &= E_t[1+(1-d)(r^*)] \\ E_{t+1} &= E_t(1+G) \end{aligned}$$

Then, the growth of earnings, that would also be the growth of dividends can be calculated as the retention rate multiplied by the return of the reinvested capital:

$$G = (1-d)r^*$$

G is considered as the sustainable growth rate of the company as it corresponds to the growth due to its internally generated funds. The variable G could be used to forecast earning figures.

Table 2 -A: EM and US market excess returns statistics (2001-2006)

	Mean	Std	Skew	Kurt	adf
ARGENTINA	15.5%	41.2%	0.470	4.631	- 4.397
BRAZIL	-0.5%	22.7%	- 0.341	2.380	- 3.235
CHILE	10.4%	13.2%	0.271	2.988	- 1.808
CHINA	16.3%	27.0%	- 0.181	3.358	- 2.586
COLOMBIA	26.6%	22.4%	- 0.360	4.490	- 1.813
CZECH.REPUBLIC	22.4%	20.7%	- 0.706	3.649	- 2.354
HUNGARY	10.0%	20.6%	- 0.434	2.852	- 2.988
INDIA	19.4%	26.1%	- 0.260	2.481	- 3.141
INDONESIA	12.9%	29.3%	- 0.467	2.915	- 3.436
ISRAEL	7.5%	18.6%	- 0.010	2.713	- 2.799
KOREA	12.7%	26.8%	- 0.058	2.929	- 2.706
MALAYSIA	6.8%	14.5%	- 0.031	2.473	- 2.831
MEXICO	11.7%	18.3%	- 0.884	3.120	- 3.203
PAKISTAN	26.1%	30.0%	0.024	3.245	- 3.135
PHILIPPINES	2.0%	21.2%	- 0.491	3.627	- 3.933
POLAND	5.7%	22.0%	- 0.668	3.161	- 4.145
RUSSIA	33.3%	30.5%	- 0.220	2.167	- 2.230
SOUTH.AFRICA	8.6%	19.3%	- 0.881	3.475	- 2.892
TAIWAN	1.0%	28.2%	0.335	4.309	- 4.181
THAILAND	14.7%	23.1%	0.295	3.451	- 2.698
TURKEY	3.8%	35.4%	0.284	3.465	- 3.219
U.S.A	-0.4%	15.0%	- 0.771	3.981	- 2.836

Figure 2 -A: Input variables for the US market (2001-2006)

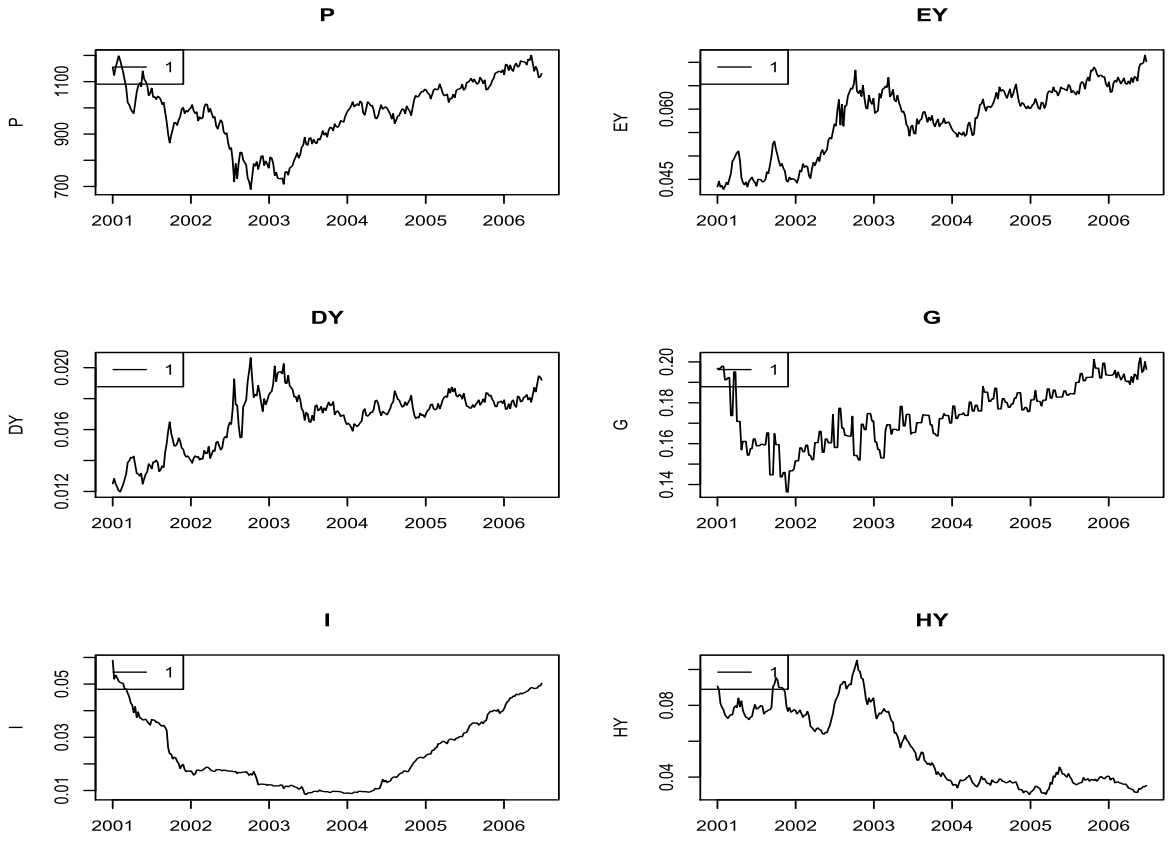


Table 2 -B: Descriptive statistics for the US-w3m

	Mean	Std	Skew	Kurt	P	EY	DY	G	I	HY	VIX
P	−0.004	0.150	−0.766	0.951	1.000	−0.018	−0.045	0.214	−0.084	−0.474	−0.591
EY	0.058	0.007	−0.627	−0.726	−0.018	1.000	0.905	0.763	0.339	−0.474	−0.288
DY	0.017	0.002	−0.698	−0.232	−0.045	0.905	1.000	0.534	0.070	−0.304	−0.056
G	0.173	0.014	−0.155	−0.697	0.214	0.763	0.534	1.000	0.538	−0.777	−0.692
I	0.023	0.012	0.665	−0.883	−0.084	0.339	0.070	0.538	1.000	−0.288	−0.403
HY	0.055	0.021	0.515	−1.247	−0.474	−0.474	−0.304	−0.777	−0.288	1.000	0.900
VIX	0.199	0.075	0.980	0.343	−0.591	−0.288	−0.056	−0.692	−0.403	0.900	1.000

Table 2 -C: Diagnostics for different lags in VAR

	k	GQ	p-GQ	BP	p-BP	DW	p-DW	BG	p-BG	JB	p-JB
AIC(n)	6.00	0.23	1.00	74.68	0.00	1.99	0.40	0.74	0.39	43.08	0.00
HQ(n)	3.00	0.24	1.00	65.28	0.00	2.01	0.45	0.20	0.65	33.96	0.00
SC(n)	1.00	0.24	1.00	63.23	0.00	2.05	0.60	0.33	0.57	23.92	0.00
FPE(n)	6.00	0.23	1.00	74.68	0.00	1.99	0.40	0.74	0.39	43.08	0.00

Figure 2 -B: Input variables for Chile (2001-2006)

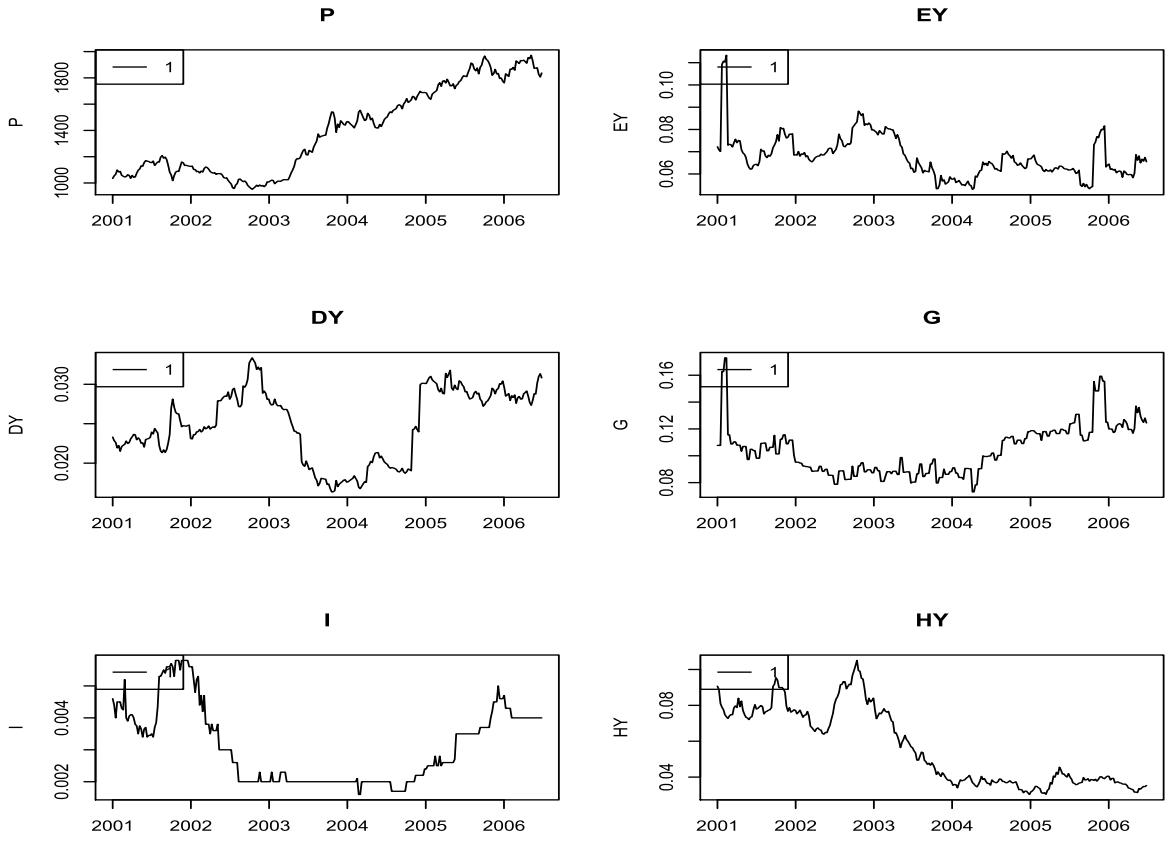


Table 2 -D: Descriptive statistics for Chile-w3m

	Mean	Std	Skew	Kurt	P	EY	DY	G	I	HY	VIX
P	0.104	0.132	0.269	−0.035	1.000	−0.372	−0.425	−0.091	−0.354	−0.301	−0.305
EY	0.067	0.008	0.480	−0.479	−0.372	1.000	0.429	−0.050	0.134	0.732	0.688
DY	0.025	0.005	−0.314	−1.185	−0.425	0.429	1.000	0.440	0.303	0.197	0.108
G	0.103	0.017	0.731	0.336	−0.091	−0.050	0.440	1.000	0.481	−0.461	−0.533
I	0.003	0.001	0.773	−0.536	−0.354	0.134	0.303	0.481	1.000	0.218	0.015
HY	0.055	0.021	0.515	−1.247	−0.301	0.732	0.197	−0.461	0.218	1.000	0.900
VIX	0.199	0.075	0.980	0.343	−0.305	0.688	0.108	−0.533	0.015	0.900	1.000

Table 2 -E: Diagnostics for different lags in VAR for Chile

	k	GQ	p-GQ	BP	p-BP	DW	p-DW	BG	p-BG	JB	p-JB
AIC(n)	2.00	0.67	0.98	28.26	0.00	2.04	0.55	1.98	0.16	34.71	0.00
HQ(n)	1.00	0.71	0.97	15.05	0.00	1.63	0.00	10.05	0.00	46.26	0.00
SC(n)	1.00	0.71	0.97	15.05	0.00	1.63	0.00	10.05	0.00	46.26	0.00
FPE(n)	2.00	0.67	0.98	28.26	0.00	2.04	0.55	1.98	0.16	34.71	0.00

Table 2 -F: RankTest

	Trace test	10pct	5pct	1pct	Lambda max	10pct	5pct	1pct
r <= 2 |	2.66	6.50	8.18	11.65	2.66	6.50	8.18	11.65
r <= 1 |	13.69	15.66	17.95	23.52	11.03	12.91	14.90	19.19
r = 0 |	41.84	28.71	31.52	37.22	28.15	18.90	21.07	25.75

Table 2 -G: RankTest

	Trace test	10pct	5pct	1pct	Lambda max	10pct	5pct	1pct
r <= 2 |	2.30	6.50	8.18	11.65	2.30	6.50	8.18	11.65
r <= 1 |	5.65	15.66	17.95	23.52	3.35	12.91	14.90	19.19
r = 0 |	31.73	28.71	31.52	37.22	26.09	18.90	21.07	25.75

Chapter 3

Portfolio construction and investment strategies

3.1 Introduction

We have studied in chapter 1 the behavior of comovements for a large set of international equity markets for a period of twelve years and we have seen how these comovements change constantly with time. In fact, when markets go though particular time periods as the TMT bubble, an energy crisis, or a terrorist attack, correlation measures spike and few diversification opportunities are left for investors. High correlation comes out during bearish phases of the market when diversification effects are the most needed. Longin and Solnik (2001) among other authors[1] have documented this asymmetric behavior of the comovements. We use a measure of integration as a basis for a time varying risk budgeting rule, that protects portfolio strategies from important losses in down trending markets. This measure could be considered as a proxy of market integration as we have presented it in chapter 1.

We are interested in constructing optimal portfolios focusing not only on the diversification across assets, but also across time. Many authors have studied the question of portfolio optimality over time. In Merton's Portfolio Problem (see Merton, 1969) the expected lifetime utility for an investor is maximized with a finite lifetime in a continuous

[1] See section 1.2.1 in Chapter 1

form.

This type of models has a closed form solution for a restricted number of cases. Campbell and Viceira (1999), Bansal et al. (2004), Brandt and Santa-Clara (2006) and others have studied the multi-horizon asset allocation problem and have suggested approximations of the solution to the multi-period problem. Here, we suggest an alternative empirical way of taking the time varying risk aversion of the investor into account. We propose a two step process where optimal weights are determined first by suitable optimization techniques and then the weights are adjusted according to the levels of integration. The advantages of our model are its simplicity and therefore its ease of implementation.

We use in the chapter a risk decomposition based on the CAPM[2] model. In fact, a portfolio can be seen as managing two separated components: the beta component and the alpha component[3]. The beta component corresponds to the part exposed to market risk, and the alpha component is at the heart of a long financial debate. Here, we consider alpha as the portfolio's return generated by a risk exposure unrelated to the market risk, where the market portfolio is an international portfolio. It is thus similar to Jensen alpha, the only difference is the definition of the market index. We study a pure alpha strategy, more precisely a portfolio strategy not related to the market risk and delivering abnormal returns. We focus on a long-short market neutral value strategy[4] based on low price-to-earnings ratios (P/E). If global factors determine strongly the excess returns, then less alpha opportunities are left for the long-short investor. It is then straightforward to use market integration as a guide for risk budgeting decisions. Intuitively, one would expect that an investment manager would allocate more risk to those decisions that he has most confidence in (high conviction bets), rather than to those he feels less certain about[5].

[2]Capital Asset Pricing Model from Sharpe (1964) and Lintner (1965).

[3]Jensen (1967) used the term alpha to define abnormal returns based on a CAPM framework.

[4]Value investors select stocks or portfolios with lower than average book-to-market ratios and/or price-to-earnings ratios or with higher than average dividend yields.

[5]We recall that alpha (long-short) investors do not use a beta-timing approach to generate alpha since they invest in a zero-beta portfolio. The investment process is defined as signals that are implemented as cross-sectional bets. Therefore, the only source of performance at a given time is the reward of the specific risk exposure across assets and it is not related to the common factor (beta) of excess returns.

In order to construct better and more realistic portfolios, we use a number of techniques used in portfolio optimisation that would be presented in the next sections. We want to build an optimal portfolio strategy that would generate alpha only using the information contained in varying comovements. A crucial point in portfolio optimization is to estimate the input data, expected returns and the covariance matrix have to be carefully estimated. Many methods are available today to built better estimates of those input variables. Even thought Chopra and Ziemba (1993) state that the sensitivity of the optimization to the error in expected returns is higher than to the errors in the covariance matrix, we decide to use robust techniques only for the covariance. The reason is that we do not have accurate estimates for every country in our data set, as we found in chapter 2, and therefore we prefer to leave the basic estimated without any correction. An alternative to use more robust estimates for the expected returns would have been for example the method of Black and Litterman Black and Litterman (1990) and include some prior beliefs to adjust the expected returns estimates. From the number of possible techniques to adjust the covariance matrix we decide to use a shrinkage technique.

We build strategy for two sets of indices separately: emerging and developed. We analyze the different strategies with a set of performance measures carefully selected. We also consider transaction costs and evaluate their impact.

The aim of this chapter is twofold. First, we set a framework that decomposes the portfolio risk into a market component and an alpha component and we show how integration is related to the market risk component. And second, we suggest an innovative way of managing risk through time, introducing integration as a key variable to manage long-short portfolios. We consider a simple long-short portfolio selection process based on forward earnings yields[6] and we build a zero-beta optimized portfolio using a covariance matrix corrected with a shrinkage procedure. We measure integration for every time period and we derive a scaling factor for our long-short weights at each date, that modifies

[6]Earnings per share divided by the actual Stock/Index Price.

the optimal weights in a way such that when markets are strongly integrated we take a more conservative position. With this procedure, we protect the strategy from the downside risk that a pure alpha investor would face during periods of high integration, when markets tend to exhibit a strong directional trend.

In the following section, we present the different methodological and theoretical aspects used in this chapter. Then, we present the data used in the empirical analysis and the main results. We finish with concluding remarks and some ideas for future research.

3.2 Prior Research on Portfolio Construction

In this section we present a brief review on relevant studies related to the estimation of expected returns, the construction of market neutral portfolios and the evaluation of portfolio performance. Some technical and practical issues are also included: the shrinkage methods, the budgeting of risk and the estimation of transaction costs.

3.2.1 Estimation of expected returns

The implementation of the portfolio problem implies a mean-variance optimization using a set of assets' expected returns and a covariance matrix. In order to achieve the best portfolio results in an ex-ante mean-variance model, it is needed to use the most accurate forecasts of returns and covariances. The most difficult part of this process is to forecast returns. A large number of studies have been dedicated to the prediction of returns with no definite answer[7]. We have dedicated the second chapter of these dissertation to study this difficult subject. There's evidence that equity returns are related to some economic variables such as short term interest rates, spreads between short and long term bonds, payout-ratios, yield spreads between high-grade and low grade bonds and could therefore be predicted. But there is no consensus on what variables to use and in which kind of relation to establish. Based on classical asset pricing theories as the CAPM and APT and valuation formulas as the Gordon model[8] many authors have suggested methods to forecast returns. Campbell and Hamao (1992) find that stock returns can be predicted using the dividend yields and interest rates variables. Campbell and Shiller (2001) suggested that the trailing earnings over price ratio is a good predictor of future returns.

[7]Some important studies related to this subject are Ang and Bekaert (2001), Campbell and Shiller (1998), Stambaugh (1999).

[8]From Gordon (1962).

3.2.2 Market neutral portfolios

Asset's returns could be decomposed into an exposure to the market and alpha, where by definition, alpha returns are uncorrelated with market returns[9]. Therefore, to focus on alpha it is necessary to eliminate the exposure to the market. That means that we have to build portfolios uncorrelated with the market (market neutral portfolios) delivering positive returns[10]. Through the use of short sales, one can engineer optimal long-short portfolios with the property of being neutral to market risk. These market neutral portfolios could be dollar neutral, beta neutral or both. A dollar neutral portfolio has zero net investment and a beta neutral portfolio targets a zero total portfolio beta.
We focus on the second kind of portfolios where the balanced long and short positions neutralize the portfolio's exposure to the underlying benchmark; the portfolio does not incur systematic risk nor does it earn the market return.

The zero-beta CAPM of Black (1972) predicts that: $E[R_i] = \gamma_0 + \gamma_1 \beta_i$ where $E[R_i]$ is the expected return on portfolio i, $\beta_i = \frac{Cov(R_i, R_M)}{Var(R_M)}$ is the beta of portfolio i, γ_0 is the expected return on the portfolio having zero covariance with the market portfolio, and γ_1 is the expected risk premium of the market portfolio. His paper and the one of Black et al. (1972) or Black and Treynor (1973) set the foundations for markets neutral investing and alpha strategies.

One of the uses of market neutral investing is to enhance the performance of any index portfolio combining an alpha strategy with any other equity index or asset. This process of transporting alpha has brought out "portable alpha" strategies. Here, we study the performance of the alpha strategy alone but it would be possible to combine it with some index portfolio or some other kind of asset delivering some performance related to other factors.

Our goal is to build a zero-beta portfolio and for this we have to determine optimal

[9] Alpha is the essence of alternative investments (See Amenc and Martelini (2003) for details about alternative investments and performance).
[10] The CAPM could not hold in this case.

long and short positions that satisfy the constraint setting the portfolio's beta to be equal to zero. We have studied in section ?? that the principal orthogonal linear factor commanding equity returns is highly correlated to an EWI. The use of an EWI as a market portfolio is then a reasonable choice in the context of alpha strategies. Using an EWI as the market portfolio is thus consistent with the idea of using a PCA to analyze the data.

3.2.3 Shrinkage

Portfolio optimization in practice present a lot of drawbacks related with the input estimation. To make optimized portfolios more realistic, the classical optimization problem has to be modified with one of the following tools: use of constraints as in Frost and Savarino (1988), use of bayesian estimators as the shrinkage methods proposed by Stein ? or Ledoit and Wolf (2003), use of portfolio resampling as proposed by Michaud (1998) or use of robust optimization programming.

Here, we decide to use shrinkage to make a more robust optimization. The basic idea behind bayesian techniques is to include prior information to produce new estimates. In deed, the shrinkage technique corrects a statistical estimator by pushing it to be closer to a certain expected value. It could be used for simple estimators as the mean, variance or regression coefficients Maddala et al. (2001).

An example could be the CAPM model's coefficients estimates alpha and beta estimated using OLS. This estimates are often inaccurate. Nevertheless with shrinkage estimation, the resulting estimate is a weighted average of two estimates. A structured estimate and a sample estimate. A popular approach to shrinkage estimation is the Stein-rule estimator. The Stein-rule estimator θ_S is defined as follows:

$$\theta_S = (1 - \delta)\theta_i + \delta\theta_p$$

where θ_i denotes the estimate parameter and θ_p the value of the parameter using prior

information or believes (it could also be an estimate with different information) and the constant δ defines the degree of shrinkage (or shrinkage constant).

In the case of the CAPM example, θ_i would be the OLS estimator for the parameters of the ith cross-section, θ_p denotes the pooled OLS estimator. The accuracy of an estimator could be measured through the estimator's root mean square error (RMSE). Maddala et al. (2001) have discussed this subject more deeply and have made comparisons of different estimators finding that shrinkage estimators are substantially more accurate than standard OLS estimates.

These estimators can be used for fund returns and give more accurate estimations of fund alphas. Fund alphas tend to fluctuate around a common mean, close to zero. As a result, large positive alpha estimates are likely to be overestimated, while large negative alphas tend to be underestimated. Shrinking the estimators towards their common mean reduces this problem improving the estimator's accuracy.

Shrinkage estimators can also be computed for the returns' variance as in the work of Daniels and Kass (2001); Ledoit and Wolf (2003). The standard statistical method is to gather a history of past stock returns to compute a sample covariance matrix. The core of the shrinkage method is that those estimated coefficients in the sample covariance matrix that are extremely high tend to contain a lot of positive error and therefore need to be pulled downwards. Similarly, to compensate for the negative error that tends to be embedded inside extremely low estimated coefficients they have to be pulled upwards. It is a shrinkage of the extremes towards the center. A choice has to be made to set a prior variance covariance matrix.

3.2.4 Risk Budgeting

We use an innovative technique to allocate risk through time. As we mentioned, during periods of high integration the strategies uncorrelated with the market deliver very poor performance but during periods of low integration there are higher opportunities to gen-

erate alpha. Hence, we construct a signal that alerts investors for periods when they need to be more conservative.

We got inspired from traditional risk budgeting techniques (Rahl (2000), Sharpe (2001)) that drive the investor to achieve the most desirable risk-return combination over a set of instruments. The goal is to allocate the budget across investments in an optimal manner. Once a risk budget is in place, the portfolio components can be monitored to assure that risk positions do not diverge from those stated in the risk budget by more than the pre-specified amounts.

This process corresponds to a cross sectional risk budgeting because risk is allocated across asset classes optimally. Our approach consists in allocating risk through time. To our best knowledge, no one has already presented risk budgeting through the time dimension and for this reason we consider that the method we present is groundbreaking. The idea is that the risk position could vary according to the state of the economy or the possibilities offered by the market.

3.2.5 Transaction costs estimation

A very sensitive problem in portfolio optimization is the estimation of transaction cost. Very often strategies are presented without taking this important variable into account which makes portfolios to be attractive in theory but unfeasible in practice. Transaction costs may include brokerage fees, bid ask spreads, taxes, fund loads and other visible or hidden costs. Their different nature makes that they can have different forms, they can be fix costs or proportional to the amount invested, all these features should be taken into account in the optimization problem. Modeling transaction costs exactly is a difficult task and practitioners use to approximate these impact of these costs in different ways. Very often transaction cost are estimated by the average transaction cost and an estimate of the turnover. This is the approach we have chosen to implement. Other alternatives are to integrate a penalty term for transaction costs into the objective function

of the optimization problem, the difficulty of this approach is that the term containing the transaction costs makes use of an absolute value function that transforms the quadratic optimization problem into a new optimization problem with no analytic solution. This problem has also some solution but a number of intermediate steps are necessary to evaluate the program.

Apart from the problem of deciding how to include transaction costs, there is the problem of quantifying these costs. Many authors have calculate this value for different kinds of assets. We are interested in estimating the costs of international equity indexes, and of course, these costs could be very different from one country to another making more difficult the estimation. For this reason in the literature we find only average transaction costs and not exact transaction costs for each asset. We follow other authors such as Guo (2002) who estimates transaction cost to 25 basis points or Berkowitz, Logue, and Noser (1988) who estimate the one-way transaction cost to be around 23 basis points for institutional investors in NYSE stocks. A study from Goldman Sachs (2004) estimates total transaction costs for S&P index futures close to 25 basis points, we take this value as a basis for all the markets. The use of futures or ETF can reduce transaction costs substantially to around 2 basis points for a round-trip transaction. For very illiquid assets the costs can climb to 200 basis points per complete transaction.

3.2.6 Performance measurement

The first portfolio performance measures formally defined are those found on the work of Jensen (1967) Jensen (1967), Sharpe (1964) Sharpe (1964), and Treynor (1965) Treynor (1965) where they defined the Jensen's alpha, the Sharpe ratio and the Treynor Ratio still used by academics and practitioners to evaluate the performance of portfolio managers. These measures are based on several assumptions about the returns distributions and their relation to the returns of the market.

Other methods have been developed to determine more precisely the levels of perfor-

mance for different returns' distributions. Information ratio is a risk-adjusted measure of a portfolio out-performance with respect to a benchmark. The Sortino ratio (See Sortino and Price, 1994) is the ratio of the excess return over the downside semi-volatility. The Omega ratio ,introduced by Keating and Shadwick (2002), corresponds ratio of the adjusted probabilities of gains and losses relative to a given threshold return, it is a measure that uses information about the whole distribution of returns and is not limited to some of their moments[11].

Very often analysts report other types of information about a fund or asset such as its worst loss over history or over the year or its best performance also for a given time period. The Maximum Drawdown which is biggest loss of the investment after reaching a peak (See Acar and James, 1997) is also reported very often.

We consider three risk indicators adequate to evaluate our strategies. The Information ratio is a common risk adjusted measure that determines the return we have earned per units of risk engaged. The Sortino ratio is important to evaluate this strategy that has been designed to protect the investor from losses during some difficult periods, thus if our strategy has equal Information ratio than other strategies, the Sortino ratio could indicate which one is better if one of them has lower downside risk. And finally the maximum drawdown let us control that our strategies do not result in extreme losses at a given point in time despite the good ratio calculated for the other criteria.

[11]See Sancetta and Satchell (2004) for details on those measures

3.3 Methodology

We structure two different portfolio strategies that we evaluate for two different sets of equity markets. To generate the first strategy we go through the following steps: first, we define the window of estimation; second, we estimate the expected returns for each asset; third, we estimate and historical covariance matrix and its shrinkage estimate; fourth, we run a beta neutral portfolio optimization to calculate optimal weights for each asset; fifth we determine the optimal portfolio and finally we move to the next window. The process is repeated for each window in the period of study. The resulting strategy could be defined a set of optimal portfolios, one for each date of evaluation.

The strategy could be defined in a formal way a set of optimal weights vectors w_t defined as $(w_{1t}, ..., w_T)$ corresponding to the different trading times $t = (0, 1, ..., T)$. The weight w_{it} corresponds to the weight of asset i at time t. The associated price vector $(P_0, P_1, ...P_T)$ contains the asset prices for the set of indices at each date where the vector P_t contains the prices for N available assets at time t. The portfolio value P_{pt} at time t can be defined as:

$$P_{pt} = \sum_{i=1}^{N} w_{i(t-1)} P_{it} \tag{3.1}$$

The resulting payoff for the strategy corresponds to the set of portfolio values P_{pt} for each trading time t.

The performance of the optimal portfolio is then evaluated with the different criteria described above. The second strategy consist of the same procedure plus an additional step before the calculation of the optimal portfolio value. We re-scale the optimal weights determined with the optimization program using the information about comovements and then we evaluate the value of the portfolio. In the following subsections we present some

details about the different choices and techniques used to construct the strategies.

3.3.1 Estimation of expected returns

In chapter 2 we have analyzed different prediction models for developed and emerging markets. We suggest to use earning yields to estimate expected returns. Here, we use a proxy for the internal rate of return which is the one year (12 months) forward earnings yield estimate $E[EY_{12}]$ minus the risk free rate r_f, we subtract the risk free rate as we want to forecast equity markets risk premia:

$$\hat{r_{t+1}} = E[EY_{12}] - r_f \tag{3.2}$$

We use $\hat{r_{t+1}}$ as an input for the optimization problem.

3.3.2 Portfolio Optimization

We rely on a modern portfolio theory setting where the investor wants to maximize his utility $U(w)$ generated by a portfolio p composed of N assets finding the optimal weights w^* for each asset. For this, we set the following optimization problem:

$$\max U_w = \max E[r_p] - \lambda Var[r_p] \tag{3.3}$$

Subject to the constraints:

$$w'\beta = 0$$

$$Var[r_p] = \sigma_0^2$$

Where $E[r_p]$ is the expected return of the portfolio defined as $E[r_p] = w'R$, R being the vector of expected excess returns of the assets and w is the vector of weights for each asset. The term $Var[r_p]$ is the variance of the portfolio defined as $V[r_p] = w'\Sigma w$ with Σ the estimated covariance matrix of returns and σ_0^2 the target portfolio's variance. The coefficient λ represents the investor's level of risk aversion and β the vector of asset's betas.

The zero-beta constraint makes that by construction the optimal portfolio would not be exposed to the market risk. Using an Equally Weighted Index, we can calculate the vector of betas as :

$$\beta = \frac{\Sigma e}{e'\Sigma e} \tag{3.4}$$

so that the zero beta constraint becomes:

$$w' \frac{\Sigma e}{e'\Sigma e} = 0 \tag{3.5}$$

As long-short strategies are not linked to market movements they have low correlation with traditional investments and provide excellent diversification benefits.

3.3.3 Shrinkage estimators

We start with the sample covariance matrix S calculated over a window and a highly structured estimator of the covariance matrix denoted by F. The shrinkage method consist in finding a compromise between these two covariances matrices by computing the shrunk covariance matrix Σ_S as a convex linear combination of S and F:

$$\Sigma_S = \delta F + (1 - \delta)S$$

where δ is a number between 0 and 1. An optimization method is used to determine the optimal δ^* [12]. The sample covariance matrix is shrunk towards the structured estimator. An example of a shrinkage target could be to use a constant correlation covariance matrix as defined in Elton and Gruber (1979). They have suggested, in another application of portfolio optimization, to use a variance-covariance matrix with equal correlation among every pair of stock or indexes[13]. The average of all the sample correlations is the estimator of the common constant correlation and together with the vector of sample variances the shrinkage target matrix, denoted by F, is constructed. Another possible shrinkage target could be a variance covariance matrix with correlations equal to zero[14]. Using the shrinkage method, we improve the robustness of the optimization procedure.

3.3.4 Risk Budgeting

To construct our portfolio strategy, we need to define a way to built the optimal portfolio at each date but allowing the risk taken to vary over time. We use the comovement measure presented in equation 1.3.3 to build a signal for the units of risk that we could take at each trading time. For low levels of integration we allow our portfolio to have the maximum risk

[12] See Daniels and Kass (2001) for complete development and the appendix for a brief review of this method.

[13] Ledoit and Wolf (2003) also use this type of covariance matrix in a shrinkage application.

[14] This matrix is often called a Sharpe matrix as optimal weights of the optimal portfolio in this case would be the Sharpe ratios.

exposure which is equivalent to use the optimal weights from the optimization. We reduce the risk by reducing the weights of every position in proportion to its level of integration. The scaling[15] process we have defined after testing many possible scaling systems consist in adjusting the optimal vector of weights w^* by a fraction of w^* so that the scaled weights w_S would be:

$$w_S = w^*(1 - \frac{\widetilde{I}_t - I_{min}}{I_{Max}}), \text{ when } I_{Max} > I_t > I_{min} \tag{3.6}$$

where $\widetilde{I}_t$ at time t is the level of integration demeaned[16] by windows in order to use the same thresholds for the whole period of study[17]. I_{min} is a threshold level of integration that indicates when integration is low and therefore there are higher chances to generate alpha and I_{Max} is a limit level of integration indicating that above it there are very few opportunities to generate alpha. Equation (3.6) modifies the weights from the optimization for levels of integration between I_{min} and I_{Max} in a linear way. For levels above I_{Max} it sets the weights to zero and for levels below the threshold I_{min} the optimal weights are not changed. We could have designed a system that upscales weights for very low integration levels indicating a lot of confidence in the signals and more possibilities to generate alpha, but we left this for further research.

We have analyzed the distribution of the historical integration measure and defined different thresholds corresponding to different quantiles of its distribution.

3.3.5 Transaction costs estimation

Transaction cost are calculated as the absolute value of trades times the average transaction cost:

[15]Note that the scaled weights are also optimal but for a different level of risk as neutral portfolio weights could be multiplied by any constant and the portfolio continues to be optimal

[16]The mean of the window is substracted

[17]Then the variable is centered to 0.5 in order to have a measure that is defined between the values zero and one.

$$TC = \theta|\gamma_t - \gamma_{t-1}| \tag{3.7}$$

Where γ_t and γ_{t-1} are old and new positions and θ is the parameter representing the average transaction costs.

The turnover can be estimated based on the weight vectors of a given strategy as:

$$TO = \sum_{t=1}^{T} \frac{|w_t - w_{t-1}|}{|w_{t-1}|} \tag{3.8}$$

Annual transaction costs tc (in percentages) are calculated as the average annual transaction costs θ times the turnover TO:

$$tc = \theta * TO * \frac{h}{260} \tag{3.9}$$

where h is the return horizon. The turnover is the percentage of position changed compared to the position held the period before.

3.3.6 Performance measures

To study the performance of a portfolio strategy we use three different indicators: The Information ratio (IR), the Sortino ratio (SR) and the maximum drawdown (MDD). For all the three measures we define r_P as the portfolio's return, r_b as the benchmark's return and r_F as the return of a risk free asset. In our case, we are concerned with absolute returns and no benchmark is used.

The Information ratio is a measure of the excess annualized return of a portfolio r_p over the annualized returns of its benchmark r_b divided by the volatility of these excess returns:

$$IR = \frac{E[r_p - r_B]}{\sigma(r_p - r_b)} \tag{3.10}$$

The information ratio is proportional to the t-statistic used to evaluate if the the null hypothesis of zero excess return for the strategy. The Sortino ratio is a variation of the famous Sharpe ratio[18] but where the return is adjusted by the down side volatility (σ_{DS}) which is a volatility calculated only with negative returns (or below the minimum accepted risk). The Sortino ratio is defined as:

$$SR = \frac{E[r_p - r_f]}{\sigma_{DS}(r_p - r_f)} \tag{3.11}$$

And the maximum drawdown is defined as the maximum cumulated continuous loss over the period of study; the largest percentage loss for a given period.

[18]The Sharpe ratio is defined as the ratio of the excess return of an asset over the risk free rate and the volatility of this excess return.

3.4 Description of the Data

All the data have been extracted from DATASTREAM (DS)[19]. We have use for this chapter two different data sets. The first one is composed of 17 emerging markets (3 from Latin America, 4 from emerging Europe, 8 Asian countries, South-Africa and Turkey) and the second one is composed of 14 developed equity markets (US, Canada, 8 countries from Europe, 3 from Asia and Australia). We have included the two different types of markets in order to be able to compare the different results. We present detailed descriptive statistics and correlation coefficients for both sets of markets in the appendix. See tables 3 -A and 3 -B for emerging markets and 3 -C and 3 -D for developed ones).

We have collected closing prices for the main DS indexes of each country in our data set expressed in local currencies. The choice of local currencies implies that we consider that there is diversification in the use of different currencies or that we can hedge the remaining exposure to the currencies with the use of other instruments, but we are not concerned with this problem in this chapter. For purposes of comparison we have also made simulations with all the data is expressed in a common currency: the US Dollar. The time period starts in January 1993 and goes to August 2006 and data is available at a daily frequency. During this period all markets had gone through the IT-bubble, its burst and then a recovery period. It is then a good sample to study a long term portfolio strategy. We also use the one year forward earnings yield estimates from I/B/E/S[20] and risk free rates for every country in order to calculate the respective risk premia.

Descriptive statistics were calculated for each country index excess returns. The results are presented in Tables 3 -A and 3 -C.

For the set of emerging countries, the mean annualized excess return is 6.3% and the annualized mean volatility is 27%. Excess returns range from -1.73% for Taiwan to

[19]Datastream International. Datastream International Ltd.

[20]I/B/E/S stands for the Institutional Brokers'Estimate System. It gathers estimates made by stock analysts on the future earnings of publicly traded companies. Here we use an aggregate estimate for each equity index.

20.41% for China and volatilities from 19% for South Africa to 46% for Turkey. The same statistics were calculated for the set of developed markets. We observe that the the highest annualized average return is 9.4% for Sweden and the lowest is 0.7% for the United Kingdom. The average return is close to 4% for all the countries. If you look at the risk figures, volatilities range from 12.5% for Australia to 25.1% for the Hong Kong. During the period 1998-1999, the Hong Kong experienced high volatility in the risk free rates, this explains the levels of volatility in the excess returns. The table also presents figures for skewness and kurtosis, the skewness coefficients are positive for the three Latin American countries, Four asian markets (China, Philippines, Taiwan and Thailand) and Turkey. And negative for the rest of emerging countries as well as all developed markets, it corresponds to a very bullish period of the equity market. Every country indexs present high degree of kurtosis but the kurtosis coefficient is not far from 3 for developed countries excepting Hong Kong that presents some leptokurtosis and Italy and Japan that exhibit very low kurtosis coefficients.

Tables 3 -B and 3 -D present correlation coefficients for each group of markets. All markets are positively correlated. Average correlation for emerging markets is close to 0.21 with highest correlation coefficients close to 0.5 for LA markets and lowest for very distant couples of markets such as Turkey and Philippines or Indonesia and Argentina where correlation is only 0.06. For the set of developed markets correlation are much higher, 0.44 on average. European markets being correlated on average at a level close to 85%. The lowest correlation being for the US and the Canadian markets to Australia and Japan between 0.1 and 0.19.

All these statistics illustrate the different features of the two sets of markets and we expect very different results for the simulated strategies.

3.5 Main Results

The pure alpha equity investor (long-short and zero-beta) is simulated by applying the most classical selection process based on forward earnings yields. This is the simplest way of constructing a portfolio based on valuations. For each country the risk premium is defined as the excess earnings yield above the local risk free rate (See equation 3.2). Rational investors would prefer high premium countries to low premium countries and so does our optimization model. The final positions are derived through mean variance optimization in order to diversify risk and eliminate beta. The risk is controlled after each optimization by adjustment of the weights to a target volatility of 15%. As our optimized portfolio has long and short positions, it is possible to decrease (increase) the overall portfolio risk reducing (enlarging) every position by a given proportion.

The long-short zero-beta strategy is rebalanced daily. We compare the performance figures of portfolio strategies with and without trading cost for the two kinds of markets. Assuming that only equity index futures are used in the portfolio, trading cost can be considered as 25 basis points of the volume traded[21]. To estimate the annual cost of transactions, we estimate the annual turnover and multiply it by the average cost of transaction.

We construct an Equally Weighted Index (EWI) with the data of our sample that we use as a reference for the construction of the beta-neutral portfolio. This choice is reasonable because we base the risk budgeting decision on the level of variance accounted by the main factor and it will also give better signals than a value weighted index that usually under-performs the EWI during bearish times.

In the optimization process, we used the matrix $\hat{\Sigma_S}$ obtained from the historical 90 days sample covariance matrix adjusted with the shrinkage method described in equation (3.3.3). With this matrix we can compute betas for each stock using equation (3.4). We calculated optimal weights every day with a fixed rolling window procedure.

[21] See the study of Goldman Sachs (2004)

We build a zero beta portfolio at each date with the estimated expected returns[22] and covariance matrix $\hat{\Sigma_S}$ calculated from the matrix of excess returns for each window. We evaluate the performance of this resulting strategies. For the set of emerging markets the beta neutral strategy delivers an information ratio of 1.03, a Sortino Ratio of 1.64 and a maximum drawdown of 36.7%, these results are presented in the first panel of Table 3.1. For the set of developed markets, the strategy has an Information ratio of 0.84, a ratio of Sortino of 1.3 and a maximum drawdown of 37,7% (See Table 3.2).

Table 3.1: Results In local Currencies - Emerging Markets

Imin	IMax	IR	Sortino	MDD	Vol	TC	IRTC
Beta Neutral Strategy							
		1.026	1.641	- 0.367	17.13%	0.042	77.9%
Beta Neutral Strategy with Risk Budgeting							
23.48%	26.43%	1.142	1.521	- 0.193	0.139	3.40%	0.897
24.18%	26.43%	1.145	1.529	- 0.200	0.141	3.39%	0.905
24.82%	26.43%	1.144	1.529	- 0.207	0.143	3.39%	0.907
23.48%	26.79%	1.222	1.744	- 0.172	0.146	3.93%	0.953
24.18%	26.79%	1.225	1.753	- 0.177	0.148	3.92%	0.961
24.82%	26.79%	1.225	1.754	- 0.181	0.150	3.92%	0.964
23.48%	27.79%	1.107	1.679	- 0.261	0.153	4.04%	0.843
24.18%	27.79%	1.109	1.685	- 0.270	0.156	4.04%	0.850
24.82%	27.79%	1.106	1.682	- 0.278	0.158	4.03%	0.851
		1.158	1.653	- 0.215	0.148	3.78%	0.903

Note: *Imin* and *IMax* are the thresholds used to scale the optimal weights, the columns *IR*, *Sortino* and *MDD* report respectively the Information Ratio, the Sortino Ratio and the maximum draw down of the Strategy, the column *Vol* is the average annualized volatility of the strategy, the column *TC* is corresponds to the average annual transaction costs, and the *IRTC* is the Information Ratio taking the cost of transaction off.
The first panel reports the results for the strategy without scaling and the second panel the results for strategies with different levels of thresholds to scale the weights.

[22]We use forward earning yields as an estimate of expected returns.

We evaluate again the strategies, this time including a penalty for transaction costs. We estimate transaction costs based on the optimal positions and calculate the modified performance of the strategy. It has an Information ratio of 0.78 for emerging markets and 0.62 for developed ones, this indicator shows that transaction costs penalize hardly the performance of the strategies.

The payoff of these long-short strategies is negatively correlated with the levels of integration estimated at each date. Integration is calculated for rolling windows of one year length at a daily frequency following equation 1.3.3 from chapter **??**. As we have mentioned before, we have observed that high integration allows very few opportunities to generate alpha. We illustrate in figures 3.1 and 3.2, the graphical time-relation between our proxy of market integration and the standard long-short pure alpha strategy for the emerging and developed markets data set. We can see that the strategies suffer during periods of high correlations among equity markets. The most impressive period was the end of the TMT bubble where all equity markets exhibit a strong downside trend. For emerging markets, the long-short strategy has many short period of continuous negative performance also associated with levels of high integration.

Figure 3.1: Long-short beta neutral payoff and integration measure- Emerging Markets

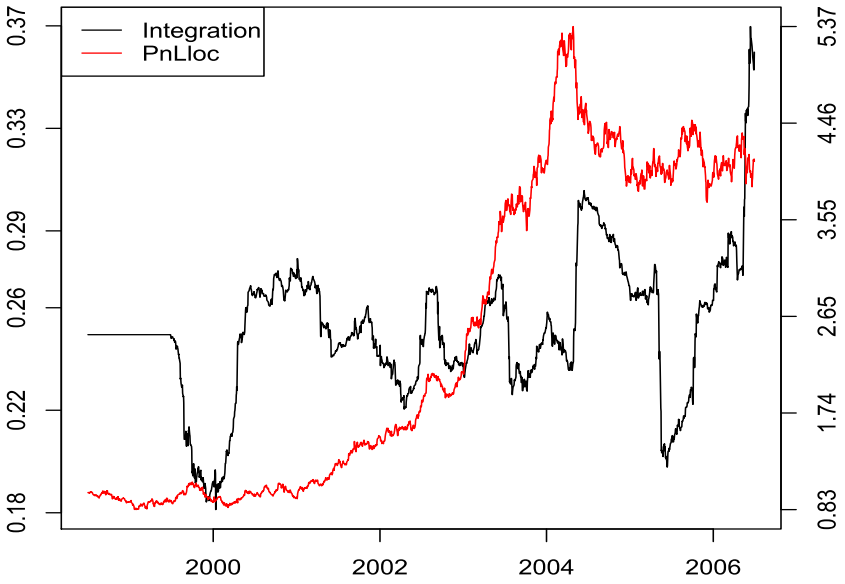
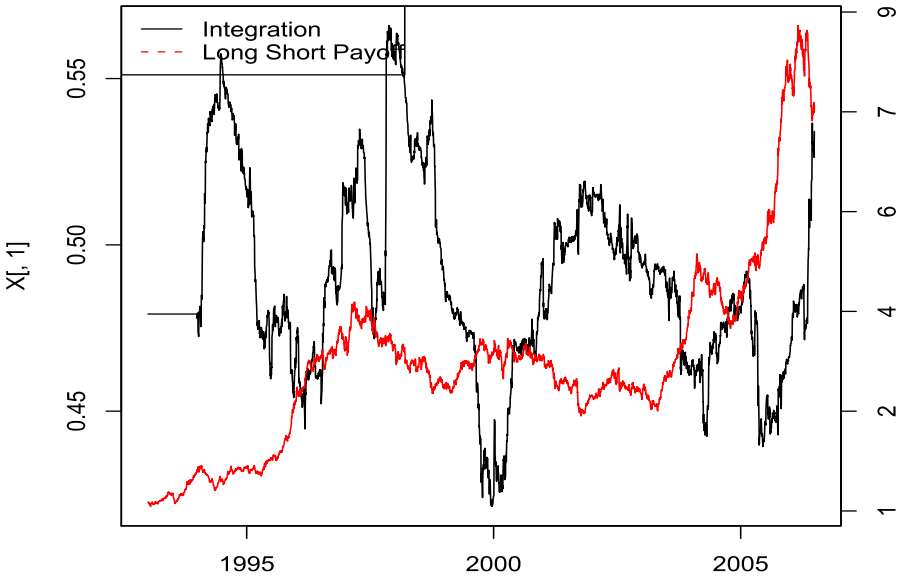


Figure 3.2: Long-short beta neutral payoff and integration measure- Developed Markets



We construct a scaling factor for the long-short weights based on the level of integration as described by equation (3.6). The level of market integration guides us through time. Its most valuable feature is that it provides us with a downside protection since it alerts us during periods of high integration when the market is very drifted. We let the integration thresholds used in equation (3.6) to vary and compute the performance of each strategy. As we do not have a definite way to determine what is a high level of integration we analyze the historical values. We test three different minimum thresholds corresponding to the second to fourth deciles of the distribution of the $\widetilde{I}_t$ variable combined with three different maximum thresholds corresponding to its fifth to seventh deciles.

The second panel of Table 3.1 presents the results of the simulations for the set of emerging markets. $\widetilde{I}_t$ has a value of 25.58% on average for these markets. We used thresholds corresponding to the indicated deciles having the values of 23,48% for the lowest I_{min} and 27.79% for the highest I_{Max}. The average value of $\widetilde{I}_t$ for the set of developed markets is 48% and the standard deviation is 9.3%. The minimum $\widetilde{I}_t$ used as I_{min} threshold is 46.7% and the maximum used as I_{Max} is 50.1%. We calculated 9 different strategies considering the different thresholds and the resulting strategies are all better in terms of the selected performance ratios. Performance ratios improve from 1.026 to 1.16 on average for the emerging countries and from 0.84 to 1.22 on average for developed, the Sortino ratio improved only slightly on average for emerging markets and from 1.3 to 1.54 for developed and the maximum drawdown has been reduced from 36.7% to 21.5% for emerging and from 37.7% to 26.8% for developed markets.

We have then estimated the transaction costs for these strategies and calculated new performance ratios taking transaction costs into account. The average transaction costs for the strategy without risk budgeting was 4.25% for emerging markets and 3.8% for developed ones. The average annual transaction costs for the strategies with risk budgeting are 3.78 % and 2.5%. Transaction costs impact the Information ratio from 1.16 to 0.9 for developed markets and from 1.22 to 1.03 on average for developed. In all cases the

strategies with risk budgeting perform better than the basic beta neutral strategy. Even if transaction cost reduce the Information ratio significantly, we still find an Information ratio close to 1 for both sets of markets, which is a good result for an alpha strategy in these market classes and considering that only very simple methods were adopted to estimate expected returns. All these results are detailed in Tables 3.1 and 3.2.

Table 3.2: Results In local Currencies-Developed Markets

Imin	IMax	IR	Sortino	MDD	Vol	TC	IRTC
Beta Neutral Strategy							
-	-	0.838	1.302	-37.76%	17.47%	3.79%	0.621
Beta Neutral Strategy with Risk Budgeting							
46.7%	48.3%	1.314	1.533	-25.77%	12.31%	2.13%	1.140
47.5%	48.3%	1.313	1.532	-25.97%	12.40%	2.13%	1.141
47.9%	48.3%	1.314	1.534	-26.03%	12.44%	2.13%	1.143
46.7%	49.1%	1.268	1.607	-27.04%	13.24%	2.47%	1.081
47.5%	49.1%	1.266	1.605	-27.26%	13.35%	2.47%	1.081
47.9%	49.1%	1.266	1.605	-27.33%	13.41%	2.47%	1.082
46.7%	50.1%	1.083	1.488	-27.05%	14.17%	2.88%	0.879
47.5%	50.1%	1.080	1.484	-27.26%	14.31%	2.88%	0.878
47.9%	50.1%	1.079	1.482	-27.33%	14.37%	2.88%	0.878

Note: *Imin* and *IMax* are the thresholds used to scale the optimal weights, the columns *IR*, *Sortino* and *MDD* report respectively the Information Ratio, the Sortino Ratio and the maximum draw down of the Strategy, the column *Vol* is the average annualized volatility of the strategy, the column *TC* is corresponds to the average annual transaction costs, and the *IRTC* is the Information Ratio taking the cost of transaction off.
The first panel reports the results for the strategy without scaling and the second panel the results for strategies with different levels of thresholds to scale the weights.

The impact of the risk scaling improves the Information ratio of the strategy from 1.026 to 1.16 and from 0.84 to 1.22 for the two kinds of markets. The other risk indicators have also improved. We conclude that the use of a time-varying risk budgeting based on an integration measure improves the payoff of a low P/E investor with the protection

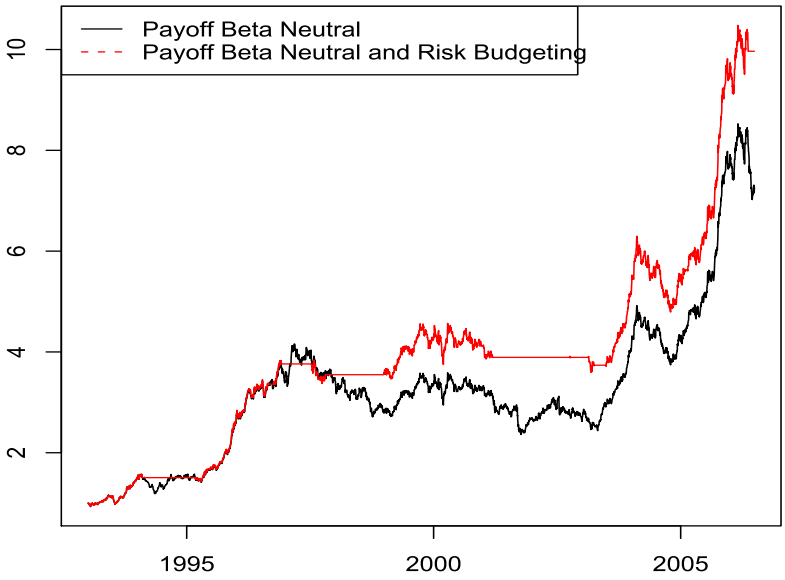
against loses during certain critical periods.

Figure 3.3 and 3.4 show the impact of using market integration as a guide for risk budgeting. These figures show the two payoffs with and without risk budgeting and we see that the risk budgeting strategy is immunized for risk during some periods.

Figure 3.3: Long-short beta neutral payoff and integration measure



Figure 3.4: Long-short beta neutral payoff and integration measure



Our results were based in a data set in local currencies. We can justify this choice in two ways. Firstly, we can consider that using a basket with different currency exposures would offset some of the total exposure by way of diversification. Here we consider a combination of 8 different currencies. Secondly, the investor can hedge his positions in foreign currencies by the use of derivatives contracts and for this reason we can leave this issue with no more analysis without invalidating our results. Nevertheless, we can compare the results to those calculated in US Dollars. We expect that the performance of the strategies would be reduced due to the lower correlation of returns in US Dollars[23] and the poor performance of the US Dollar, in particular during the last period of the study. All the tables and figures corresponding to the calculations in USD are presented in the Appendix.

We observe that for the set of emerging markets, in the first panel of Table 3.3, performance ratios are similar in local currencies or in Dollars. However, for developed markets, the Information ratio of the basic beta neutral strategy in US Dollars is 0.53, much lower than the 0.84 in local currency. Applying the correction to the weights according to the proposed methodology, we find and average information ratio of 0.81 considering the same deciles of the distribution of the integration measure. If we take transaction costs into account, the figures drop to 0.28 and 0.6. In all cases the use of risk budgeting improves the performance of the strategies in terms of the Information ratio and Sortino ratio but for some strategies we have lower maximum drawdown figures.

All the results for the performance criteria of the different strategies are summarized in Table 3.3. In all cases, with or without transaction costs, in local currencies or in US Dollars the use of a risk budgeting technique enhances the performance of the strategies.

[23]Considering that what we look for is alpha and it won't be possible to generate alpha if all the asset are highly correlated

Table 3.3: Performance measures for different strategies

		IR	Sortino	MDD	TC	IRTC
Emerging Markets						
Local	No RB	1.026	1.641	- 0.367	4.22%	0.779
	RB	1.158	1.653	- 0.215	3.78%	0.903
USD	No RB	1.093	1.724	- 0.264	4.25%	0.825
	RB	1.128	1.606	- 0.167	3.68%	0.880
Developed Markets						
Local	No RB	0.838	1.302	- 0.566	3.79%	0.621
	RB	1.220	1.541	- 0.268	2.49%	1.034
USD	No RB	0.531	0.793	- 0.536	3.86%	0.283
	RB	0.807	0.990	- 0.442	2.63%	0.605

Note: The columns *IR*, *Sortino* and *MDD* report respectively the Information Ratio, the Sortino Ratio and the maximum draw down of the Strategy, the column *TC* is corresponds to the average annual transaction costs, and the *IRTC* is the Information Ratio taking the cost of transaction off.
The first panel reports the results for developed markets and the second panel emerging markets. The lines *No RB* correspond to strategies built without Risk Budgeting and the lines *RB* for the strategies where the weights were scaled in order to drive the level of risk taken.

3.6 Summary and Conclusions

This chapter uses the information found in chapters 1 and 2 about the structure of international equity market returns and the way these markets share common risk factors in order to create portfolio strategies in different types of markets. We have shown that monitoring the importance of the first global risk factor, used here as a proxy of market integration, and its power explaining cross-sectional excess returns can be used as a guide for risk budgeting strategies. High explanation power of the first common risk factor indicates periods of strong trends that leave less opportunities for the pure alpha investor to capture portfolio performance. We have analyzed the different strategies with a set of selected performance indicators and we have shown that the use of a technique to manage the risk through time based on the levels of comovements is useful and enhance the performance of the strategies.

In fact, the use of such integration measures protects the zero-beta investor from extreme losses in periods of high correlations due to specific market phases. Our simulations illustrate that measuring and monitoring integration could protect the long-short strategy, based on valuations, during the burst of the TMT bubble and other similar episodes. Such techniques can be very useful for hedge funds or alternative investment strategies that seek alpha generation through zero-beta portfolios.

In this chapter, we have analyzed separately emerging and developed markets for the same period of study. First, we have built a beta neutral portfolio strategy and analyzed its performance indicators. We have used indicators such as the information ratio, the sortino ratio and maximum drawdown which are pertinent to evaluate this portfolio strategies and are commonly used by practitioners. Then, we have constructed a strategy that takes into account the actual levels of integration and reduces the risk taken when the masure of integration is higher than a given threshold. This strategy presents better performance indicators so we conclude that the management of risk over time based on the criteria we have defined could be very useful for portfolio managers. We have also observed that

for the set of emerging markets using our risk budgeting techniques enhances more the performance than for the set of developed markets .

In order to analyze the robustness of each portfolio strategy , we have compared the compared the results of a data set in US Dollars with a data set in local currencies. We have observed that for the set of emerging markets there is no significant change between the two data sets, however, for the set of developed markets the strategy calculated in US Dollars is penalized compared to the one calculated in local currencies. In addition to the robustness of the currency used, we have also analyzed the impact of transaction costs. For this, we have calculated an annual average cost per transaction and a measure of turnover based on the changes of weights for the instruments in each portfolio. We compute the performance of each strategy taking the measure of transaction costs into account and we observe that all the strategies are penalized but still deliver good estimates of performance letting us to conclude that the idea of considering the level of integration could be important for the management of equity portfolios and particularly for emerging markets .

From a theoretical point of view our method is an alternative to dynamic portfolio construction such as

An alternative to manage risk through time could be to change the investor's risk aversion parameter in the optimization process. Other more sophisticated techniques to estimate expected returns or to measure integration could be studied in future research. We think that it would be interesting to explore more deeply the idea of managing risk through time, with similar methods, and particularly the optimal way of defining thresholds of integration. Our results could be improved using dynamic thresholds for the scaling of the weights, and also computing ex-ante measures of integration.

Appendix

Table 3 -A: Descriptive Statistics - Emerging markets

	Mean	Std	Skew	Kurt
ARGENTINA	7.26%	27.56%	0.708	14.028
BRAZIL	-0.77%	23.83%	0.686	19.151
CHINA	20.41%	28.34%	0.245	6.054
CZECH.REPUBLIC	11.89%	20.75%	- 0.249	5.629
HUNGARY	-1.11%	25.27%	- 0.488	10.035
INDIA	11.80%	26.06%	- 0.654	7.673
INDONESIA	-4.74%	28.63%	- 0.016	8.723
KOREA	14.77%	33.54%	- 0.048	6.290
MALAYSIA	6.62%	22.51%	- 0.252	59.983
MEXICO	5.77%	20.49%	0.173	8.065
PHILIPPINES	-0.90%	20.33%	1.165	19.322
POLAND	-0.57%	24.65%	- 0.039	6.300
RUSSIA	13.26%	36.41%	- 0.296	7.237
SOUTH.AFRICA	5.36%	19.02%	- 0.541	7.966
TAIWAN	-1.73%	26.70%	0.024	5.507
THAILAND	7.84%	29.31%	0.157	11.437
TURKEY	11.95%	46.53%	0.125	7.978

Note: Statistics calculated for the period 01-1993 to 08-2006 in local currencies.

Table 3 -B: Correlation coefficients for the period 01-1993 to 08-2006-Emerging Markets

	ARG.	BRA.	CHI.	CZR.	HUN.	IND.	IDN.	KOR.	MAL.	MEX.	PHI .	POL.	RUS.	SAF.	TAI.	THA.	TRK.
ARGENTINA	1.00	0.47	0.11	0.16	0.18	0.07	0.06	0.09	0.06	0.40	0.10	0.10	0.20	0.20	0.10	0.13	0.12
BRAZIL	0.47	1.00	0.13	0.17	0.21	0.10	0.07	0.15	0.04	0.56	0.11	0.19	0.23	0.26	0.12	0.14	0.20
CHINA	0.11	0.13	1.00	0.18	0.20	0.19	0.22	0.27	0.23	0.13	0.20	0.23	0.15	0.27	0.21	0.29	0.16
CZECH.REP.	0.16	0.17	0.18	1.00	0.50	0.18	0.16	0.20	0.13	0.24	0.10	0.39	0.37	0.39	0.15	0.19	0.27
HUNGARY	0.18	0.21	0.20	0.50	1.00	0.19	0.21	0.29	0.15	0.27	0.15	0.47	0.36	0.44	0.12	0.20	0.29
INDIA	0.07	0.10	0.19	0.18	0.19	1.00	0.20	0.26	0.11	0.13	0.13	0.20	0.18	0.23	0.20	0.17	0.11
INDONESIA	0.06	0.07	0.22	0.16	0.21	0.20	1.00	0.25	0.19	0.05	0.26	0.22	0.16	0.19	0.18	0.29	0.11
KOREA	0.09	0.15	0.27	0.20	0.29	0.26	0.25	1.00	0.20	0.15	0.24	0.29	0.17	0.29	0.36	0.32	0.19
MALAYSIA	0.06	0.04	0.23	0.13	0.15	0.11	0.19	0.20	1.00	0.07	0.15	0.21	0.14	0.18	0.15	0.27	0.15
MEXICO	0.40	0.56	0.13	0.24	0.27	0.13	0.05	0.15	0.07	1.00	0.11	0.20	0.24	0.32	0.12	0.13	0.16
PHILIPPINES	0.10	0.11	0.20	0.10	0.15	0.13	0.26	0.24	0.15	0.11	1.00	0.22	0.08	0.19	0.17	0.26	0.06
POLAND	0.10	0.19	0.23	0.39	0.47	0.20	0.22	0.29	0.21	0.20	0.22	1.00	0.29	0.35	0.20	0.24	0.24
RUSSIA	0.20	0.23	0.15	0.37	0.36	0.18	0.16	0.17	0.14	0.24	0.08	0.29	1.00	0.37	0.13	0.18	0.25
SOUTH.AFRICA	0.20	0.26	0.27	0.39	0.44	0.23	0.19	0.29	0.18	0.32	0.19	0.35	0.37	1.00	0.19	0.27	0.22
TAIWAN	0.10	0.12	0.21	0.15	0.12	0.20	0.18	0.36	0.15	0.12	0.17	0.20	0.13	0.19	1.00	0.22	0.14
THAILAND	0.13	0.14	0.29	0.19	0.20	0.17	0.29	0.32	0.27	0.13	0.26	0.24	0.18	0.27	0.22	1.00	0.20
TURKEY	0.12	0.20	0.16	0.27	0.29	0.11	0.11	0.19	0.15	0.16	0.06	0.24	0.25	0.22	0.14	0.20	1.00

Table 3 -C: Descriptive Statistics - Developed markets

	Mean	**Std**	**Skew**	**Kurt**
AUSTRALIA	3.7%	12.5%	- 0.370	5.192
CANADA	7.2%	13.8%	- 0.664	6.218
FRANCE	5.0%	18.8%	- 0.166	3.065
GERMANY	4.1%	18.3%	- 0.377	3.268
HONG.KONG	3.7%	25.1%	- 0.041	10.036
ITALY	4.2%	20.3%	- 0.174	2.822
JAPAN	1.6%	18.8%	- 0.085	2.801
NETHERLANDS	5.5%	18.4%	- 0.264	4.459
SINGAPORE	2.3%	18.4%	- 0.010	6.409
SPAIN	7.7%	18.5%	- 0.275	3.051
SWEDEN	9.4%	22.4%	- 0.016	3.855
SWITZERLAND	8.2%	16.4%	- 0.341	4.770
UNITED.KINGDOM	0.7%	15.0%	- 0.272	3.338
U.S.A	4.5%	16.5%	- 0.126	4.300

Note: Statistics calculated for the period 01-1993 to 08-2006 in local currencies.

Table 3 -D: Correlation coefficients for the period 01-1993 to 08-2006-Developed Markets

	Austra.	Canada	France	Germ.	HongK.	Italy	Japan	Netherl.	Singap.	Spain	Sweden	Switzer.	UK-	USA
Austra.	1.00	0.19	0.27	0.31	0.47	0.22	0.39	0.30	0.38	0.26	0.31	0.28	0.28	0.09
Canada	0.19	1.00	0.45	0.46	0.21	0.37	0.16	0.42	0.19	0.42	0.42	0.40	0.44	0.68
France	0.27	0.45	1.00	0.80	0.31	0.72	0.24	0.84	0.29	0.78	0.72	0.76	0.80	0.43
Germ.	0.31	0.46	0.80	1.00	0.36	0.66	0.23	0.80	0.30	0.72	0.70	0.74	0.72	0.46
HongK.	0.47	0.21	0.31	0.36	1.00	0.25	0.37	0.33	0.57	0.28	0.31	0.30	0.33	0.12
Italy	0.22	0.37	0.72	0.66	0.25	1.00	0.18	0.68	0.23	0.69	0.60	0.65	0.65	0.35
Japan	0.39	0.16	0.24	0.23	0.37	0.18	1.00	0.24	0.33	0.19	0.25	0.23	0.24	0.10
Netherl.	0.30	0.42	0.84	0.80	0.33	0.68	0.24	1.00	0.29	0.74	0.69	0.80	0.80	0.41
Singap.	0.38	0.19	0.29	0.30	0.57	0.23	0.33	0.29	1.00	0.25	0.28	0.28	0.29	0.13
Spain	0.26	0.42	0.78	0.72	0.28	0.69	0.19	0.74	0.25	1.00	0.65	0.70	0.70	0.39
Sweden	0.31	0.42	0.72	0.70	0.31	0.60	0.25	0.69	0.28	0.65	1.00	0.64	0.67	0.37
Switzer.	0.28	0.40	0.76	0.74	0.30	0.65	0.23	0.80	0.28	0.70	0.64	1.00	0.74	0.38
U.K.	0.28	0.44	0.80	0.72	0.33	0.65	0.24	0.80	0.29	0.70	0.67	0.74	1.00	0.41
U.S.A	0.09	0.68	0.43	0.46	0.12	0.35	0.10	0.41	0.13	0.39	0.37	0.38	0.41	1.00

Figure 3 -A: Long-short beta neutral payoff and integration measure

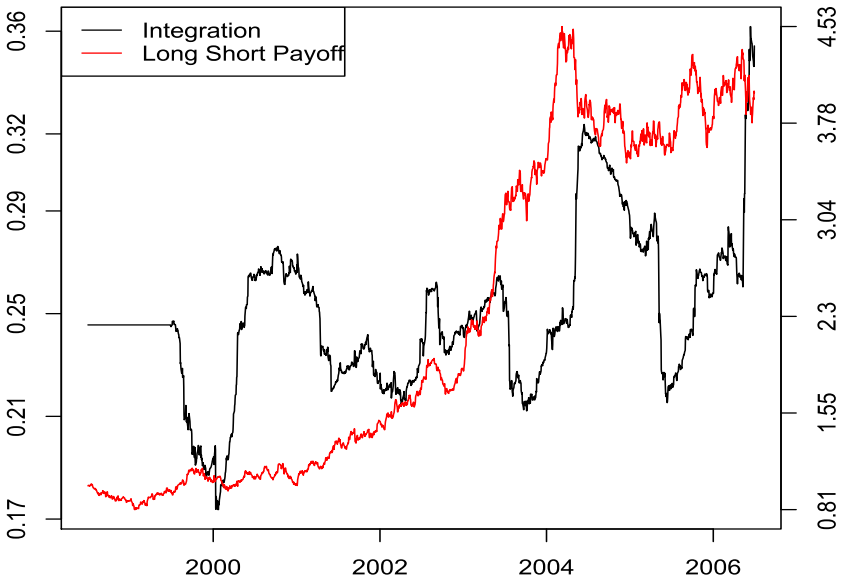


Figure 3 -B: Long-short beta neutral payoff and integration measure

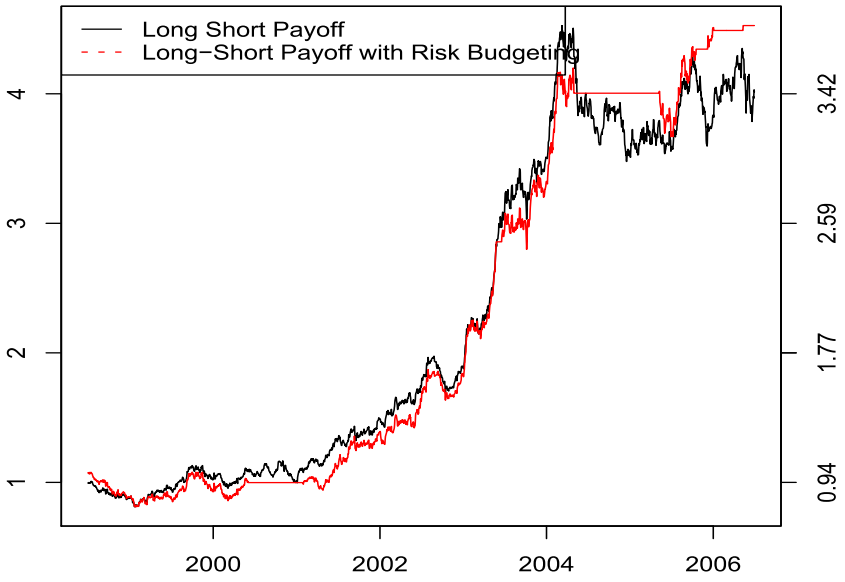


Table 3 -E: Results In USD - Emerging Markets

Imin	IMax	IR	Sortino	MDD	Vol	TC	IRTC
Beta Neutral Strategy							
		1.093	1.724	- 0.264	0.158	4.25%	0.825
Beta Neutral Strategy with Risk Budgeting							
22.5%	26.1%	1.226	1.624	- 0.131	0.137	3.33%	0.984
23.4%	26.1%	1.231	1.637	- 0.134	0.140	3.33%	0.994
24.3%	26.1%	1.235	1.647	- 0.138	0.143	3.32%	1.002
22.5%	26.7%	1.022	1.471	- 0.176	0.145	3.81%	0.760
23.4%	26.7%	1.022	1.475	- 0.183	0.149	3.80%	0.766
24.3%	26.7%	1.020	1.477	- 0.190	0.151	3.80%	0.769
22.5%	28.7%	1.132	1.705	- 0.178	0.154	3.93%	0.877
23.4%	28.7%	1.133	1.710	- 0.185	0.157	3.92%	0.883
24.3%	28.7%	1.132	1.712	- 0.191	0.160	3.92%	0.887
		1.128	1.606	- 0.167	0.148	3.68%	0.880

Note: *Imin* and *IMax* are the thresholds used to scale the optimal weights, the columns *IR*, *Sortino* and *MDD* report respectively the Information Ratio, the Sortino Ratio and the maximum draw down of the Strategy, the column *Vol* is the average annualized volatility of the strategy, the column *TC* is corresponds to the average annual transaction costs, and the *IRTC* is the Information Ratio taking the cost of transaction off.
The first panel reports the results for the strategy without scaling and the second panel the results for strategies with different levels of thresholds to scale the weights.

Table 3 -F: Results In USD - Developed Markets

Imin	IMax	IR	Sortino	MDD	Vol	TC	IRTC
Beta Neutral Strategy							
-	**-**	0.531	0.793	-53.62%	15.58%	3.86%	0.283
Beta Neutral Strategy with Risk Budgeting							
45.0%	46.9%	0.973	1.111	-30.88%	11.75%	2.20%	0.7852
45.7%	46.9%	0.974	1.113	-30.88%	11.83%	2.20%	0.7877
46.3%	46.9%	0.976	1.117	-30.88%	11.88%	2.20%	0.7910
45.0%	47.5%	0.829	1.041	-43.34%	12.83%	2.65%	0.6224
45.7%	47.5%	0.828	1.041	-43.72%	12.94%	2.65%	0.6235
46.3%	47.5%	0.829	1.042	-44.05%	13.01%	2.65%	0.6250
45.0%	48.3%	0.621	0.819	-57.22%	14.01%	3.05%	0.4032
45.7%	48.3%	0.619	0.816	-57.96%	14.14%	3.05%	0.4030
46.3%	48.3%	0.617	0.813	-58.62%	14.24%	3.05%	0.4027

Note: *Imin* and *IMax* are the thresholds used to scale the optimal weights, the columns *IR*, *Sortino* and *MDD* report respectively the Information Ratio, the Sortino Ratio and the maximum draw down of the Strategy, the column *Vol* is the average annualized volatility of the strategy, the column *TC* is corresponds to the average annual transaction costs, and the *IRTC* is the Information Ratio taking the cost of transaction off.
The first panel reports the results for the strategy without scaling and the second panel the results for strategies with different levels of thresholds to scale the weights.

General Conclusion

This dissertation analyzed the features of different input variables for a portfolio construction problem applied to a set of emerging equity markets as well as an application to the construction of portfolio strategies. We focused on determinant input variables for the portfolio problem such as the expected return and the matrix of covariances. We focused on this set of markets for many reasons. First of all, emerging economies are growing at faster pace then developed ones and their financial markets are developing very fast. Second, these markets have suffered from continuous changes in the recent period leading us to question the validity of previous studies. Third, these types of markets have attracted less empirical analysis than the set of developed markets.

This study has been structured into three chapters. The first one focused on the study of comovements of equity market indices and the second on some models to forecast the returns. The third chapter focuses on portfolio construction, we used the elements founded in the second chapter and combined them to actual methods of portfolio construction. The result was very satisfactory in the sense that we were able to build original portfolio strategies.

In the first chapter, we have analyzed the comovements of a large set of equity markets. We decomposed the set of markets into developed and emerging and compared the behavior of the comovements over time. We concluded from this chapter that the comovements are time varying with a tendency to increase over time.

We analyzed the average level of comovements and showed that it is closely related to

the levels of volatility. This first analysis of the different emerging equity markets gave us important information about the structure of those markets. Decomposing the returns into principal components and evaluating the levels of variance explained by each principal factor let us understand the internal structure of equity returns. We have also studied the impact of the comovements on financial integration and diversification. A world of integrated markets is a world of better risk sharing, lower cost and larger opportunities of investments. But highly integrated markets could lead to lower diversification opportunities. These two issues were analyzed as direct consequences of higher comovements.

We have shown that the principal factor is responsible for the largest part of the comovements so that it drives its changes and that the structure of equity returns is different for emerging or developed markets. Additionally, we have determined that the first factor is highly correlated with an equally weighted index composed of the variables in the data set while the second factor is useful to account for episodes of crises such as the Asian and the Argentinean crisis.

A main result is that there is an upward trend in the global comovements and also that this trend is stronger for developed than emerging countries. In addition, taking advantage of the large cross-section and time series dimensions, we have estimated the number of relevant factors over a rolling window in the lines of Bai and Ng. Thus, we have concluded that for emerging countries, this number is constantly decreasing and that it is a number close to 4 for emerging markets and 3 for the developed markets on average. We have also analyzed the correspondence between the measure of comovements an the number of factors indicated with the Bai and Ng criteria by comparing for the same dates the behavior of this two indicators. We expected to have a diminishing number of factors while observing an increasing level of comovements, but the data does not confirm this statement. We conclude that a decreasing number of factors should not be interpreted as a sign of integration. We have shown that the measure of comovements could be related to important economic and financial issues such as the process of inte-

gration of equity markets and the time varying opportunities of diversification offered by international markets.

It is really important for the portfolio construction problem to analyze the predictability of the returns. The resulting forecast could be included in portfolio optimization problems. Forecasting equity returns has been widely debated as a main topic of the financial literature, many approaches are possible, we tried to analyzed different basic options of forecasting and also suggest something new to the literature. Based on an economic reasoning with a fundamental financial approach, we suggested an innovative model of forecasting that we tested for a set of emerging markets.

The main result of our analysis is the finding of a cointegration relation within the selected set of variables, this relation is present in several markets. We arrive to this conclusion after running the rank tests of Johansen which are confirmed with the estimation of an error correction model having negative and significant coefficients for the correction term for some of the countries that we have analyzed. From the estimated model we can conclude that there is some evidence of predictability for this set of countries. We also tested a specification including a growth variable but even if the rank tests suggest the presence of cointegrating relations, the estimated models have no significant coefficient and poor adjusted R^2. The introduction of the implied volatility variable in our model helps to improve the prediction. This variable has rarely been used in the literature as a forecasting variable for equity returns. Our study has some limits. First, we have imposed a given forecasting horizon and we haven't tested the results if we change this parameter. Second, we haven't analyzed the out-of-sample results of this prediction models and as presented by Goyal and Welch (2003), good results in-sample could vanish when tested out-of-sample.

The third chapter of this dissertation suggested a methodology to construct a portfolios strategy taking into account the specificities of emerging markets.

We have analyzed separately emerging and developed markets for the same period of

study. First, we have built a beta neutral portfolio strategy and analyzed its performance indicators. We have used indicators such as the information ratio, the sortino ratio and maximum drawdown which are pertinent to evaluate this portfolio strategies and are commonly used by practitioners. Then, we have constructed a strategy that takes into account the actual levels of integration and reduces the risk taken when the measure of integration is higher than a given threshold. This strategy presents better performance indicators so we conclude that the management of risk over time based on the criteria we have defined could be very useful for portfolio managers. We have also observed that for the set of emerging markets using our risk budgeting techniques enhances more the performance than for the set of developed markets . In order to analyze the robustness of each portfolio strategy , we have compared the compared the results of a data set in US Dollars with a data set in local currencies. We have observed that for the set of emerging markets there is no significant change between the two data sets, however, for the set of developed markets the strategy calculated in US Dollars is penalized compared to the one calculated in local currencies. In addition to the robustness of the currency used, we have also analyzed the impact of transaction costs. For this, we have calculated an annual average cost per transaction and a measure of turnover based on the changes of weights for the instruments in each portfolio. We compute the performance of each strategy taking the measure of transaction costs into account and we observe that all the strategies are penalized but still deliver good estimates of performance letting us to conclude that the idea of considering the level of integration could be important for the management of equity portfolios and particularly for emerging markets .

Bibliography

Adjaouté, K., Danthine, J.-P., 2004. Equity returns and integration: is europe changing? Oxford review of economic policy 20 (4).

Amenc, N., Martelini, L., 2003. Portfolio optimization and hedge fund style allocation decisions. Edhec Risk and Asset Management Research Center.

Ang, A., Bekaert, G., 2001. Stock return predictability: is it there? NBER Working Paper (8207).

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. Journal of Financial Economics 63 (3), 443–494.

Babilis, S., Fitzgerald, V., October 2005. Risk appetite, home bias and the unstable demand for emerging market assets. International Review of Applied Economics 19 (4), 459–476.

Baele, L., Ferrando, A., Hördahl, P., Krylova, E., Monnet, C., 2004. Measuring financial integration in the euro area. ECB (14).

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135–171.

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

Bai, J., Ng, S., 2003. Confidence intervals for factor forecasts with many predictors. Manuscript.

Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131, 507–537.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics 25, 52–60.

Bansal, R., Dahlquist, M., Harvey, C. R., 2004. Dynamic trading strategies and portfolio choice. NBER Working Paper (10820).

Bekaert, G., Harvey, C. R., 1995. Time varying world market integration. The Journal of Finance 50 (2).

Bekaert, G., Harvey, C. R., 1997. Emerging equity markets volatility. Journal of Financial Economics 43, 29–77.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45 (3), 444–455.

Black, F., Jensen, M. C., Scholes, M., 1972. The capital asset pricing model: Some empirical tests. Michael C. Jensen, Studies in the theory of capital markets, Praeger Publishers Inc.

Black, F., Litterman, R., 1990. Asset allocation: Combining investor views with market equilibrium. Sachs & Co., Fixed Income Research.

Black, F., Treynor, J. L., 1973. How to use security analysis to improve portfolio selection. Journal of Business 46 (1), 66–86.

Bollerslev, T., Zhou, H., 2007a. Expected stock returns and variance risk premia. Finance and Economics Discussion Series - Federal Reserve Board, Washington 11.

Bollerslev, T., Zhou, H., 2007b. Expected stock returns and variance risk premia. Finance and Economics Discussion Series Divisions of Research and Statistics and Monetary Affairs Federal Reserve Board.

Brandt, M. W., Santa-Clara, P., 2006. Dynamic portfolio selection by augmenting the asset space. The Journal of Finance 61 (5).

Breden, D. T., 1979. An intertemporal asset pricing model with stochastic consump-. tion and investment opportunities. Journal of Financial Economics 7, 265–296.

Brooks, R., Negro, M. D., 2003. Firm-level evidence on international stock market co-movements. Federal Reserve Bank of Atlanta.

Brown, S. J., 1989. The number of factors in security returns. The Journal of Finance 44, 1247–1262.

Burmeister, E., McElroy, M. B., 1988. Joint estimation of factor sensitivities and risk premia for the arbitrage pricing theory. The Journal of Finance 43, 721–733.

Campbell, J. Y., Hamao, Y., 1992. Predictable stock returns in the United States and Japan: a study of long-term capital market integration. The Journal of Finance 47 (1), 43–69.

Campbell, J. Y., Lettau, M., Malkiel, B. G., Xu, Y., 2001. Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. The Journal of Finance 56 (1).

Campbell, J. Y., Shiller, R. J., 1998. Valuation ratios and the long-run stock market outlook. The Journal of Portfolio Management Winter.

Campbell, J. Y., Shiller, R. J., 2001. Valuation ratios and the long-run stock market outlook: An update. Cowles Foundation Discussion Papers.

Campbell, J. Y., Viceira, L. M., 1999. Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics 114, 433–95.

Carrieri, F., Errunza, V., Hogan, K., 2005. Characterizing world market integration through time. Manuscript.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and the mean-variance analysis on large asset markets. Econometrica 51 (5), 1281–1304.

Chen, L., Guo, H., Zhang, L., 2006. Equity market volatility and expected risk premium. Research Division Federal Reserve Bank of St. Louis Working Paper Series 007-A.

Chopra, V. K., Ziemba, W. T., 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. The Journal of Portfolio Management Winter.

Connor, G., Korajczyk, R., 1986. Performance measurement with the arbitrage pricing theory a new framework for analysis. Journal of Financial Economics 15, 373–394.

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

Daniels, M. J., Kass, R. E., 2001. Shrinkage estimators for covariance matrices. Biometrics 57, 1173–1182.

de Santis, G., Gerard, B., 1997. International asset pricing and portfolio diversification with time-varying risk. The Journal of Finance 52 (5), 1881–1912.

Elton, E. J., Gruber, M., 1979. Simple criteria for optimal portfolio selection: The multi-index case. ed.

Engle, R. F., Granger, C. W. J., 1987. Cointegration and error-correction: Representation, estimation, and testing. Econometrica 55.

Errunza, V., Hogan, K., wei Hung, M., 1999. Can the gains from international diversification be achieved without trading abroad? The Journal of Finance 54, 2075–2107.

Errunza, V., Losq, E., 1985. International asset pricing under mils segmentation: Theory and test. The Journal of Finance 40 (1), 105–124.

Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics November.

Fama, E. F., French, K. R., 1992. The cross-section of expected returns. The Journal of Finance 47 (2).

Ferson, W. E., Harvey, C. R., 1999. Conditioning variables and the cross section of stock returns. The Journal of Finance 54 (4), 1325–1360.

Forbes, K. J., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. Journal of Finance 57 (5), 2223–2261.

Frost, P., Savarino, J. E., 1988. For better performance constrain: constrain portfolio weights. Journal of Portfolio Management Fall.

Goetzmann, W. N., Li, L., Rouwenhorst, K. G., 2005. Long-term global market correlations. Journal of Business 78 (1).

Gordon, M. J., 1962. The investment, financing, and valuation of the corporation. The American Economic Review 52 (5).

Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Management Science 49 (5), 639–654.

Guo, H., 2002. On the out-of-sample predictability of stock market returns. Federal Reseve Bank of Saint Louis - Working Papers Series (008C).

Harvey, C. R., 1991. Predictable risk and returns in emerging markets. Review of financial studies 8 (3).

Heston, S. L., Rouwenhorst, K. G., Wessels, R. E., 1995. The structure of international stock returns and the integration of capital markets. Journal of Empirical Finance 2, 173–197.

Hodrick, R. J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Review of Financial Studies 5 (3), 357–386.

Jensen, M., 1967. The performance of mutual funds in the period 1945-1964. The Journal of Finance 23 (2).

Karolyi, A., Stulz, R., 2002. Are financial assets priced locally or globally? NBER (8994).

Karolyi, G. A., Stulz, R. M., July 1996. Why do markets move together? An investigation of U.S.-Japan stock return comovements. Journal of Finance 51 (3), 951–86, available at http://ideas.repec.org/a/bla/jfinan/v51y1996i3p951-86.html.

Korajcyk, R., 1996. A measure of stock market integration for developed and emerging markets. World Bank Economic Review 10 (2), 267–89.

Ledoit, O., Wolf, M., 2003. Honey, I shrunk the sample covariance matrix. Working Paper.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics 47 (1), 13–37.

Longin, F., Solnik, B., 1995. Is the correlation in international equity returns constant: 1960-1990? Journal of International Money and Finance 14 (1), 3–26.

Longin, F., Solnik, B., 2001. Extreme correlations of international equity markets. The Journal of Finance 56 (2).

Lucas, R. E., 1978. Asset prices in an exchange economy. Econometrica 46 (6), 1429–1445.

Maddala, G., Li, H., Srivastava, V. K., 2001. A comparative study of different shrinkage estimators for panel data models. Annals of economics and finance 2, 1–30.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7 (1), 77–91.

Michaud, R. O., 1998. Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation. Harvard Business Press.

Mody, A., Taylor, M. P., 2003. The high-yield spread as a predictor of real economic activity: Evidence of a financial accelerator for the united states. IMF Staff Papers-International Monetary Fund 50 (3).

Rahl, L., 2000. Risk Budgeting: A New Approach to Investing. Risk Books.

Ratio, T. D.-P., of Future Dividends, E., Factors., D., 1988. Campbell, john y. , and robert j. shiller. The Review of Financial Studies 3 (1).

Ross, S. A., 1976. Return, Risk and Arbitrage.

Rouwenhorst, K. G., 1999. Local return factors and turnover in emerging stock markets. The Journal of Finance 54 (4).

Sancetta, A., Satchell, S. E., 2004. Calculating hedge fund risk: the draw down and the maximum draw down. Applied Mathematical Finance 11.

Sharpe, W. F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. The Journal of Finance 19 (3), 425–442.

Sharpe, W. F., September 2001. Budgeting and monitoring the risk of defined benefit pension fundsAvailable at http://www.stanford.edu/ wfsharpe/art/q2001/q2001.htm.

Solnik, B., 1974. An equilibrium model of the international capital market. Journal of Economic Theory 8 (4), 500–524.

Solnik, B., 2008. Equity home bias and regret: An international equilibrium model. HEC Working Paper.

Solnik, B., Roulet, J., 2000. Dispersion as cross-sectional correlation. Financial Analysts Journal 56, 54–61.

Stambaugh, R. F., 1999. Predictive Regressions. Journal of Financial Economics 54, 375–421.

Stock, J. H., Watson, M. W., 1988. Testing for common trends. Journal of the American Statistical Association.

Stock, J. H., Watson, M. W., 1998. Diffusion indexes. No. 6702.

Stulz, R., 1984. Pricing capital asset s in an international setting. Journal of international Business 55.

Stulz, R. M., 1981. On the effects of barriers to international investments. The Journal of Finance 36 (4), 923–934.

Treynor, J., January-February 1965. How to rate management of investment funds. Harvard Business Review.

Trzcinka, C., 1986. On the number of factors in the arbitrage pricing model. Journal of Finance 41.

van der Hart, J., Slagter, E., van Dijk, D., 2003. Stock selection strategies in emerging markets. Journal of Empirical Finance 10.