ESSAYS ON LIQUIDITY COMMONALITY IN EQUITY MARKETS

Riccardo Werther Borghi

Thesis submitted to Cass Business School in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Finance
Cass Business School
City, University of London

July 2018

Contents

Contents iii

List of Tables v

List of Figures vii

Acknowledgements ix

Abstract xi

Introduction 1

1 On the commonality in stock liquidity and its determinants 5

1.1 Introduction . . . 5
1.2 Data . . . 10
1.3 Market liquidity . . . 12
1.4 The sources of common factors in liquidity . . . 17
1.5 Model implied liquidity commonality . . . 22
1.6 The determinants of commonality in liquidity . . . 28
1.7 Conclusions . . . 34
Appendix 1.A Data description . . . 49
Appendix 1.B Deseasonalisation procedure . . . 51
Appendix 1.C Kalman Filter . . . 51
Appendix 1.D Robustness checks . . . 54

2 High-frequency quoting and liquidity commonality **57**
2.1 Introduction . . . 57
2.2 Institutional setting and data . . . 61
2.3 Estimating liquidity commonality with high-frequency data . . . 69
2.4 Testing if HFQ increases liquidity commonality . . . 75
2.5 Intraday liquidity commonality and HFQ . . . 79
2.6 Conclusions . . . 82
Appendix 2.A Data details . . . 95
Appendix 2.B Additional results . . . 99
Appendix 2.C Some comments on PCA . . . 100

3 The dynamics of systematic risk in a large equity portfolio **105**
3.1 Introduction . . . 105
3.2 Model . . . 109
3.3 Data and identification . . . 115
3.4 Estimation results . . . 119
3.5 The benefits of using time-varying factor loadings . . . 124
3.6 Global comovements of stock returns: new evidence . . . 128
3.7 The connection with the profile of the firm . . . 131
3.8 Conclusions . . . 133
Appendix 3.A Leverage for Financials . . . 155
Appendix 3.B Data cleaning . . . 157
Appendix 3.C Replication of Bekaert et al. (2014) . . . 158

Conclusions and further research **169**

Bibliography **173**

List of Tables

1.1 Summary statistics, trading data (daily) . . . 37
1.2 Summary statistics, exogenous variables (monthly) . . . 38
1.3 Static factor model estimates . . . 39
1.4 The determinants of total liquidity commonality . . . 40
1.5 The determinants of volatility-driven liquidity commonality . . . 41
1.6 The determinants of exposure-driven liquidity commonality . . . 42
1.A.1 Variables description . . . 50
1.C.1 Initialisation Kalman Filter . . . 52
1.C.2 Parameters estimates . . . 53
1.D.1 Unrestricted SUR, total commonality . . . 56

2.1 Summary statistics (order book) . . . 84
2.2 Summary statistics (HFT) . . . 85
2.3 Liquidity commonality . . . 86
2.4 Panel estimation . . . 87
2.5 Panel estimation (sample split) . . . 88
2.6 Panel estimation (bid-offer) . . . 89
2.B.1 Descriptive statistics (Mar 2010 - Jul 2010) . . . 99
2.B.2 Hourly patterns . . . 100

3.1 Universe of Securities . . . 135
3.2 Descriptive statistics . . . 136
3.3 Pearson correlation between factors and exogenous variables . . . 137

3.4 Mapping estimated factors and Fama-French 3 factors 138
3.5 Model estimates 139
3.6 Goodness of fit 140
3.7 Misspecification tests 141
3.8 Static variance decomposition 142
3.C.1 Table II Bekaert et al (2014) 162
3.C.2 Table III Bekaert et al (2014) 163
3.C.3 Table IV Bekaert et al (2014) 167

List of Figures

1.1 Deseasonalisation effect on the illiquidity indexes 43
1.2 Stock illiquidity and liquidity shocks 44
1.3 Sort-dependent liquidity correlation matrices 45
1.4 Time-varying factor significance 46
1.5 Some factor loadings estimates 47
1.6 Commonality in liquidity within region 48
1.D.1 52-week rolling correlation 54

2.1 Signature plot: number of factors 90
2.2 Signature plot: eigenvalues 91
2.3 Signature plot: autocorrelation dispersion 92
2.4 Intraday patterns in HFQ, commonality and BO 93
2.5 Intraday patterns in volatility and order flow 94

3.1 Estimated global and regional factors 143
3.2 The conditional factor exposure of two large firms 144
3.3 Average deviations from OLS per year 145
3.4 Time-varying variance decomposition 146
3.5 Time-varying variance decomposition (by region) 147
3.6 Time-varying variance decomposition (by region) - continued . . . 148
3.7 Time-varying variance decomposition (by sector) 149
3.8 Time-varying variance decomposition (by sector) - continued . . . 150
3.9 Time-varying variance decomposition (by sector) - continued . . . 151

3.10 Relationship with firm size 152
3.11 Relationship with firm leverage 153
3.12 Relationship with expected returns 154
3.A.1 Relationship with firm leverage 156
3.C.1 Goodness of fit interdependence model 165

Acknowledgements

I would like to express my sincere gratitude to the following people who have supported me during this journey:

My first advisor, Professor Giovanni Urga, for his continuous mentoring since the very beginning, and my second advisor, Professor Richard Payne, for fruitful discussions and straight-to-the-point feedback on my work

My PhD examiners, Professor Catherine Doz and Professor Giovanni Cespa, for very interesting discussions and important feedback of my work

Timberlake Consultants, for sponsoring my studies and for giving me the fantastic opportunity to visit the European Central Bank and the University of Cambridge multiple times

My colleagues and friends in the Faculty of Finance at Cass, Angela Gallo, Ian Marsh, Filippo Andrini, Michele Bergamelli and Riccardo Pianeti, for our discussions and their advice

My parents, Paola and Werther, for their unconditional love and immense support

Eleonora, for her constant moral support and understanding

My friends, Davide, Gianne, Josh, Matteo, Nick and Simone, for their positivity

Abstract

This thesis contributes to the emerging literature that investigates the drivers of liquidity commonality in equity markets. Our main contribution is threefold. First, we propose a new method to estimate liquidity commonality in equity markets to explore both supply-side (funding liquidity of intermediaries) and demand-side (trading behaviour of investors) determinants of its dynamics. The empirical analysis uses weekly data on 1909 stocks from the US, Japan, the UK and Euro zone countries, from January 2000 to January 2017. Second, we propose high-frequency quoting (HFQ), an activity carried out by high-frequency traders (HFT), as a new supply-side explanation of high-frequency liquidity commonality. Using the upgrade of the London Stock Exchange (LSE) trading system as an exogenous shock, we find that an increase in HFQ leads to an increase in liquidity commonality. Furthermore, we analyse the intraday patterns in high-frequency liquidity provision in relation to other microstructure variables, using tick-by-tick data for the FTSE100 stocks listed on the LSE, from September 2010 to July 2011. Finally, motivated by the crucial role of factor models in our research, we propose a two-level factor model with time-varying loadings that captures financial, global and regional risk in stock returns. We use it to investigate the dynamics of systematic risk in a large portfolio of firms from 54 countries, from January 2006 to January 2016.

Introduction

In the last ten years market liquidity, the ease to trade an asset, has played a pivotal role in the events that shaped financial markets, attracting the attention of regulators and academics. During the Flash Crash of May 2010, market liquidity vanished in a few seconds due to the selling pressure exercised by a single algorithm [Kirilenko et al. (2017)]. On January 2018, the European Union updated the Markets in Financial Instruments Directive (MiFID) with, among other things, a set of rules aimed at increasing pre- and post-trade transparency for investors that seek liquidity from intermediaries. Since the seminal papers by Kyle (1985) and Amihud and Mendelson (1986), several authors have analysed and modelled the market liquidity of a security. However, financial markets integration and continuous improvement in technology have made liquidity increasingly connected across securities and this has made the multivariate study of liquidity a hot research topic. The concept of commonality in liquidity, i.e. the co-movement in market liquidity of different stocks, was first introduced by Chordia et al. (2000) and Hasbrouck and Seppi (2001). Commonality in liquidity implies correlation in execution costs and this has important implications for asset pricers and portfolio managers. On the asset pricing front, Acharya and Pedersen (2005) proves that the covariance of stock-level liquidity with aggregate liquidity is a determinant of expected returns. On the portfolio management front, consider a situation where a trading desk must execute a basket of orders by the end of the day. If liquidity is positively correlated across stocks and there is an overall decline in liquidity, then trading becomes unavoidably expensive.

The purpose of this thesis is threefold. First, we investigate supply-side (funding liquidity of intermediaries) and demand-side (trading behaviour of investors) determinants of liquidity commonality. Our research on the determinants of commonality in liquidity builds on earlier work on demand factors by Brockman et al. (2009) and Koch et al. (2016) and work on supply side effects by Coughenour and Saad (2004) and Hameed et al. (2010). Earlier work which looks at both demand and supply side effects is sparse and their methods are limited. Second, we propose high-frequency trading (HFT) activity as a new supply-side explanation of liquidity commonality. HFTs are firms that use proprietary capital to act as market maker on multiple securities. Compared to banks, HFT's business model is much less diversified and consequently more exposed to funding constraints that could affect the liquidity of multiple stocks. In the limit, with only one HFT firm supplying liquidity, a reduction in its capital, driven by losses in one stock, might make the HFT less willing to put capital at risk and thus less willing to supply liquidity for all stocks. A few papers have proposed explanations for liquidity commonality that can be indirectly related to HFT activity [Coughenour and Saad (2004), Domowitz et al. (2005)], but the direct effect of HFT activity on liquidity commonality is still unexplored. In terms of methodologies, the econometric analysis of liquidity poses a series of challenges. Even though liquidity measures such as Amihud (2002) are integrated of order zero or fractionally integrated, most of the literature [e.g. Chordia et al. (2000), Brockman et al. (2009)] applies a first difference transformation to liquidity, which mechanically introduces negative autocorrelation in the time series [Hasbrouck and Seppi (2001)]. Furthermore, the literature measures stock-specific liquidity commonality by the R^2 of a one-factor model, which suffer from bias due to the changing volatility of underlying market factors [Forbes and Rigobon (2002)]. In Chapter 1, using portfolio liquidity, we control for this effect developing a new method to estimate time-varying correlations in financial markets. In Chapter 2, using stock liquidity, we take care of this bias controlling for time effect in a panel data model. These methodological challenges motivates the third objective of this thesis, which is to propose and analyse a two-level factor model with time-varying loadings, which we develop in Chapter 3. The model is flexible enough to accommodate changes in the underlying commonality structure among time series. We use it to estimate common components in stock returns and provide new evidence on stock returns comovements.

The thesis is divided in three chapters. Chapter 1 explores demand and supply drivers of the correlation between transaction costs for a group of stock portfolios and proposes a new method to estimate time-varying correlations in financial markets. We estimate the liquidity correlation matrix as implied by a factor model where we allow both the factor loadings and the factor volatilities to vary over time and measure liquidity commonality within a set of stocks as the mean pair-wise correlation (MPC) among the liquidity of the stocks in that set. As our model allows for factor heteroskedasticity, we can control for the bias pointed out by Forbes and Rigobon (2002). We estimate the *total commonality* (TC) as the conditional mean pairwise correlation and decompose this into *volatility-driven commonality* (VD) and *exposure-driven commonality* (ED). We propose time-series tests based on seemingly unrelated regression (SUR) models to evaluate the explanatory power of a set of exogenous variables that approximate the funding constraints of intermediaries and the behaviour of investors for commonality. The tests use either TC, VD or ED as the dependent variable to identify which part of liquidity commonality is driven by the supply and which part by the demand-side of liquidity. We use weekly data on the US, Japan, the UK and Euro zone countries, from January 2000 to January 2017. Furthermore, we introduce a modified version of the Amihud (2002) measure, adjusted for deterministic components of trading volume, motivated by Gallant et al. (1992), so that price impact at the beginning and end of the sample are comparable. The Amihud (2002) measure is the most widely used approximation of price impact because it is easily constructed from daily data and it provides a good approximation to high-frequency price impact [Goyenko et al. (2009)].

Chapter 2 investigates if the changing nature of liquidity providers - i.e. from banks to HFTs [Menkveld (2013)] - has increased the interconnectedness of equity markets, captured by liquidity commonality. We use the order-driven market offered by the London Stock Exchange (LSE) as laboratory test, focusing on the period between January 2010 and December 2011. Our data base comprises all trades and best quotes updates executed or posted on the LSE, for the stocks that entered the FTSE100 Index. We use the technological upgrade of 14 February 2011, when the LSE introduced the Millennium Exchange trading system, to identify variation in HFT activity. First, we compare the eigenvalues in two five-month period before and after the shock, extending the analysis of Hasbrouck and Seppi (2001). Second, we perform a panel data analysis to formally test an increase in the liquidity commonality, controlling for market conditions. We estimate stock-level liquidity commonality as the explanatory power of common liquidity factors on individual stock liquidity, measured by order book depth. We estimate the common factors by principal component analysis (PCA) as in Hasbrouck and Seppi (2001) and we additionally shed light on the impact of long memory and sampling frequency on the structure of a factor model for stock liquidity. We find empirical support for Hasbrouck and Seppi (2001)'s decision of using a three factor model when the data are sampled at 15-minute frequency. The PC_1 criteria of Bai and Ng (2002) suggests that a three factor model can be used for frequencies from eight to 15 minutes.

In Chapter 3, we formulate an international factor model where stock returns are assumed to be a function of two types of factors: global (one financial sector factor and one latent factor) and region-specific (one latent factor per region). The financial sector factor is observed while the global and region latent factors are estimated with PCA using the procedure of Breitung and Eickmeier (2015). In our model, we let the factor loading, i.e. the sensitivities of each return to the factors, vary over time as independent AR(1) processes, thus the structure of the model is an extension of Breitung and Eickmeier (2015) and Ando and Bai (2015), whereby the factor loadings are assumed time invariant. We analyse a panel of 2000 stock returns from six worlds regions to investigate if the importance of unobserved global and regional factors is time-varying and we give a straightforward visual interpretation of that. Furthermore, we relate the dynamics of systematic risk to the profile of the risk. We rank stocks by loading persistence and variance and connect this to size, leverage and insample expected returns.

Chapter 1

On the commonality in stock liquidity and its determinants*

1.1 Introduction

Commonality in liquidity has important implications for asset pricers and portfolio managers. On the asset pricing front, Acharya and Pedersen (2005) derive a liquidity-adjusted CAPM in which the average level of liquidity of a security is a determinant of its expected returns, but expected returns also depend explicitly upon the covariance of stock-level liquidity with aggregate liquidity [see, also, Amihud and Mendelson (1986), Pastor and Stambaugh (2003), Lee (2011)]. On the portfolio management front, consider the arrival of a shock which implies that the investor must sell a portion of her portfolio. If costs of trading vary across stocks but are invariant over time, the investor can always pick a set of positions to liquidate that has low expected trading cost. If however, liquidity covaries positively across assets and the arrival of shocks to the investor is associated with a decline in overall liquidity, then trading becomes unavoidably expensive exactly when the investor wishes to trade.

*A joint research paper with Prof. Richard Payne and Prof. Giovanni Urga, based on the results of this chapter, has been presented at the "Young Finance Scholars' Conference" (University of Sussex, 12^{th} June 2017), the "Macquarie Quantamentals Research Seminar" (Macquarie Securities Group, 11^{th} November 2016) and the "2015 PhD Research Days" (Cass Business School, 8^{th} June 2015).

This paper explores demand and supply drivers of the correlation between transaction costs for a group of stock portfolios and proposes a new method to estimate time-varying correlations in financial markets. We estimate the liquidity correlation matrix as implied by a factor model where we allow both the factor loadings and the factor volatilities to vary over time and measure liquidity commonality within a set of stocks as the mean pair-wise correlation (MPC) among the liquidity of the stocks in that set. Forbes and Rigobon (2002) show that Pearson's correlation coefficients are biased by the changing volatility of underlying market factors. Thus, studies of liquidity commonality that use R^2-based commonality measures, such as Hameed et al. (2010), Karolyi et al. (2012), Koch et al. (2016), Moshirian et al. (2017), suffer from bias due to the presence of heteroskedasticity in underlying factors. However, as our model allows for factor heteroskedasticity, our estimates do not suffer from this problem. We estimate the *total commonality* (TC) as the conditional mean pairwise correlation and decompose this into:

- *volatility-driven commonality* (VD), which is the component of TC driven by the time-varying covariance matrix of the factors, assuming that factor loadings are constant over time. VD captures the fact that correlation can increase (decrease) if underlying factors become more (less) volatile, even if stocks' exposures are constant;
- *exposure-driven commonality* (ED) which is the component of TC driven by time-varying factor loadings, holding the factor covariance matrix constant at its long-run value. ED captures the fact that correlation can increase (decrease) if some stocks become more (less) exposed to common liquidity factors, holding the variability of those factors constant

We propose time-series tests based on seemingly unrelated regression (SUR) models to evaluate the explanatory power of a set of exogenous variables that approximate the funding constraints of intermediaries and the behaviour of investors for commonality. The tests use either TC, VD or ED as the dependent variable to identify which part of liquidity commonality is driven by the supply and which part by the demand-side of liquidity.[1] Using weekly data on the US,

[1]Various authors have associated changes in factor loadings as permanent changes in co-movement statistics and changes in factor volatilities as temporary. For instance, Bekaert et al.

Japan, the UK and Euro zone countries, from January 2000 to January 2017, we address the following questions. What are the most important sources of liquidity risk? In addition to a global liquidity factor, do regional and sectoral factors explain individual stock liquidity? What are the underlying economic sources of liquidity commonality? Does liquidity commonality relate to variables that approximate the behaviour of the demand side of the market (i.e. trader behaviour) or the supply side (e.g. funding liquidity constraints)?

Our analysis yields three main contributions. Our first contribution is the identification of the sources of liquidity risk factors in a multi-country portfolio of stocks. While Karolyi et al. (2012) consider only country-specific commonality, in this paper we identify global, regional and sectoral factors and we use a data-driven approach to rank the importance of the three types of factors. We estimate one global, four regional and ten GICS sectoral factors as equally-weighted portfolios. Then, in order to identify the estimated factors, we orthogonalise them.

We find that the common drivers of the cross-section of liquidity are mainly global and region-specific, while liquidity of stocks in the same sector show little tendency to co-move. This evidence suggests that market liquidity is not a stock characteristic that is determined by the business profile of a company. Instead, it is determined by region-specific forces that make the liquidity of different firms more similar due to geographic proximity. For instance, the transaction costs of two Japanese stocks belonging to the Energy and IT sectors may covary strongly, while two IT stocks listed in US and Japan may have costs of trading that behave very differently. We postulate that both demand (behaviour of investors) and supply (behaviour of intermediaries) specific shocks might be the cause of this commonality and proceed to investigate both channels.

This investigation forms our second contribution. We investigate how demand and supply side factors contribute to commonality. This builds on earlier work on demand factors by Brockman and Chung (2002) and Brockman et al. (2009), and on supply side effects by Coughenour and Saad (2004) and Hameed et al. (2010). It is worth noting that earlier work which looks at *both* demand and supply side effects is sparse and also that earlier work only looks at total

(2014) and Dungey and Renault (2017) use changes in loadings to detect contagion across markets.

correlations. We have the advantage of being able to study VD and ED as well as TC. Furthermore, existing papers estimate liquidity factor models using a rolling-window estimation, with results depending on the window's length, a nuisance parameter[2]. In this paper instead, we consistently estimate the time-varying factor loadings via maximum likelihood (ML) estimation through the Kalman Filter. Our model is related to the solution proposed by Hallin et al. (2011), who use a dynamic factor model, in the spirit of Forni et al. (2000), featuring lags and leads of latent liquidity factors.

We find that both demand and supply shocks play a role in explaining liquidity commonality. On the supply side, we find that TC is positively related to the US commercial paper spread and the TED spread, in line with Brunnermeier and Pedersen (2009) and Hameed et al. (2010): when funding constraints are binding, liquidity commonality increases. We find that a negative shock to liquidity supply is positively related to VD but negatively associated to ED. Thus, when funding liquidity becomes constrained (which may be the case during a crisis), the underlying common factors become more volatile and they increase the total liquidity commonality measure, while at the same time the exposure driven component of commonality decreases, partially offsetting the positive effect on total liquidity commonality. Beaupain et al. (2010), who condition their analysis upon regimes of market return volatility, find that the magnitude of liquidity commonality among US large caps is typically lower in stressful markets. Our results on the ED component have a similar flavour to this, but our overall commonality results point in the opposite direction.

On the demand side, we find that index-related trading (proxied by ETF trading volume) is positively related to liquidity commonality, supporting the results of Kamara et al. (2008) and Koch et al. (2016): Thus correlated liquidity demand from institutional investors increases commonality. Conversely, global market sentiment (proxied by the US sentiment index of Baker and Wurgler (2006)) is negatively related to commonality, implying that, when people become more optimistic, price impact is less correlated across stocks. In contrast to the supply-side case, the coefficients of demand-side variables, when significant, are consistent across TC, VD and ED. In particular, ETF trading volume is

[2] Inoue et al. (2017) show that the length of the estimation window greatly affects also the forecasting performance of a model.

the only variable that is always significant and positive. We interpret this result as evidence that the correlated liquidity demand of institutional investors is the strongest regional economic force that makes stock liquidity co-move.

Our third contribution is the empirical measure of portfolio liquidity that we employ. Price impact is, together with tightness and depth, one dimension of stock liquidity defined by Kyle (1985) and it is the sensitivity of stock price to signed trade flow. We introduce a modified version of the Amihud (2002) measure, adjusted for deterministic components of trading volume, motivated by Gallant et al. (1992), so that price impact at the beginning and end of the sample are comparable. The Amihud (2002) measure is the most widely used approximation of price impact because it is easily constructed from daily data and it provides a good approximation to high-frequency price impact [Goyenko et al. (2009)]. Other authors who use this proxy include Bali et al. (2014), Acharya and Pedersen (2005) and Karolyi et al. (2012). Our measure is essentially a value-weighted index of liquidity, rebalanced yearly so that we use only active stocks (stocks that are members of large stock market indexes) and we avoid the risk of small-stock bias.[3]

We find that our measure of liquidity, in contrast to the unadjusted Amihud (2002), does not display any visible trend or seasonality. The unadjusted measure is particularly affected by a linear trend in trading volume for US stocks and holiday effect around half-day trading for UK stocks. Not controlling for these deterministic patterns would bias the inference of unit root tests towards concluding that the series are non-stationary [see, e.g. Ng and Perron (1995)]. On the contrary, our measure allows an unbiased comparison of price impact throughout the sample and across countries, facilitating the multivariate analysis of liquidity.

The remainder of the paper is organized as follows. Section 1.2 presents the data. Section 1.3 introduces our measure of liquidity. Section 1.4 describes the identification of the common liquidity factors. Section 3.2 presents the methodology to estimate liquidity commonality. Section 3.4 analyses the determinants of liquidity commonality. Section 1.7 concludes.

[3] Korajczyk and Sadka (2008) investigate the commonality between liquidity measures and show that different liquidity measures are proxies of the same underlying risk factor that is priced in the cross-section of returns.

1.2 Data

Our data cover the stock markets of 15 countries: United States, United Kingdom, Japan, Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. We classify stocks by currency region or by sector. We group our stocks into four currency "regions": United States (US), United Kingdom (UK), Japan (JP) and Euro zone (EZ). Thus, the stocks within each region are subject to a series of common financial and monetary shocks. Modelling EZ as one aggregate entity is motivated by Brooks and Del Negro (2005), who show that within-region country factors can be mostly explained by regional factors. Bekaert et al. (2009) follow the same procedure. Then, we use the Global Industry Classification Standard (GICS) ten-sector classification to separate stocks by business profile[4].

We collect trading data at daily frequency on the constituents of four large stock market indexes: S&P500, FTSE350, Nikkei225 and Euro STOXX[5] from Bloomberg. For each stock, we download last price, intraday high and low, trading volume, and number of shares outstanding. Trading volumes are composite, i.e. aggregate across exchanges, to avoid jumps in volumes due to relocations of primary listing across US exchange venues[6] and to consider all the transactions executed in fragmented markets[7]. We consider large cap indexes because our focus is on the time-series determinants of liquidity commonality, rather than the cross-sectional ones [such as in Brockman et al. (2009)]. The sample goes from 11^{th} January 2000 to 20^{th} January 2017. It includes multiple episodes of market turmoil that can be used to study how liquidity commonality behaves in normal and abnormal market conditions. These events comprise the burst of the dot-com bubble (2000-2001), the terrorist attacks of 9/11, the great finan-

[4]The ten GICS sectors are: Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services and Utilities.

[5]The Euro STOXX is a subset of the STOXX Europe 600 that is composed only of stocks listed in Euro zone countries. Source: `http://www.stoxx.com/indices/index_information.html?symbol=SXXE`.

[6]Examples include: Kraft Foods from NYSE to Nasdaq in 2012 and Imax from Nasdaq to NYSE in 2011.

[7]Indeed, price impact is venue-specific and the availability of intraday quotes allows to calculate the trade-by-trade impact on price. However, when high-frequency data is not available and markets are very fragmented, the primary exchange volume is a very partial information.

cial crisis (2007-2009) and the European sovereign-debt crisis (2011-). Consequently, the stock market indices we study have experienced a high turnover of their composition. To reduce survivorship bias, we use the index composition at the beginning of the sample and rebalance it every five years. In total we consider 1909 stocks that entered the stock market indexes during the sample and we use them to construct capitalisation-weighted indexes of liquidity rebalanced every year with the available data. For more information and details on database coverage, see Appendix 2.A. Furthermore, we build a set of global and region-specific variables capturing the state of capital markets, the demand for market liquidity by investors and the supply of liquidity by intermediaries. We use the VIX Index, TED spread, St. Louis FED financial stress indicator, Baker and Wurgler (2006)'s US sentiment index and the US commercial paper (CP) spread as global variables, while short-term rates, the returns on the banking sector, exchange rate fluctuations, net equity flow, market turnover and market volatility are region-specific variables. Table 1.A.1 in the Appendix describes the variables, how they are computed and their source.

Table 1.1 reports the summary statistics for returns, trading volume and market capitalisation divided by region. Over the sample period, US stocks have the highest average daily stock return and the largest return reversal, with a first autocorrelation coefficient (AR1) of -0.030. The return distributions have positive skewness and very fat tails for all regions. UK-listed stocks have, on average, a skewness very close to zero and the largest kurtosis of the sample, suggesting the presence of large outliers. The median market capitalisation of British stocks in our universe is \$2 billion, which is much lower than their mean capitalisation (\$10 billion) due to the presence of medium-sized companies in the FTSE350 index. We use the FTSE350 instead of the FTSE100 to consider stocks whose business is more related to shocks specific to the UK[8]. Trading volume is the largest for US stocks but it is comparable across countries. The maximum values of market cap correspond to Vodafone (UK) and Apple (US). The very low minimum are due to the inclusion of stocks that might have traded outside

[8]We compare the distribution of market capitalisation and trading volume of the stocks listed in the FTSE350 with those of the FTSE100 and we find that, they are comparable. Thus, we expect the illiquidity of the FTSE350 stocks to have a higher level than those in the FTSE100 but their dynamics to be very related. We use the FTSE350 to cover a larger universe of stocks.

the index, but these are going to be excluded from our liquidity index. The AR(1) coefficient indicates that the volume is very persistent.

[Table 1.1 about here.]

Table 1.2 reports time series summary statistics for the exogenous variables used to capture capital market conditions and the level of demand/supply of liquidity. Market volatility is the percentage standard deviation of daily market return and the VIX Index is the expected monthly volatility (annualised) of the S&P500 Index returns. The VIX has a mean of 20% and a median of 17% and US market volatility (after annualising it) is in the same scale. Market turnover is the average trading volume, expressed as percentage of shares outstanding, which results in an average between 0.980% (US) and 1.2% (JP). Exchange rate is in monthly percentage change, where a positive value indicates a depreciation of the currency relative to the SDR. The St. Louis FED financial stress indicator, whose positive values indicate stress, indicates that, on average, our period was calm (the median value of FED stress is lower than the mean) with a few very stressful episodes. ETF volume is dollar trading volume in local ETFs traded on US markets, calculated as a percentage of the total stock market capitalisation. Net equity flow is the difference between gross sales of foreign stocks by foreigners to US residents and gross purchases of foreign stocks by foreigners from US residents, scaled by the sum of of gross sales and purchases of foreign stocks by foreigners to/from US residents. A positive net (%) equity flow signals that US residents are net buyers of foreign stocks. All variables based on interest rates - short-term rates, CPs and TED spread - are expressed in percentage points.

[Table 1.2 about here.]

1.3 Market liquidity

In this section we introduce our measure of adjusted price impact that approximates the average market liquidity of a portfolio of stocks. We use it to capture the market liquidity of a region-sector portfolio. Our measure is a modified version of Amihud (2002)'s measure, adjusted for deterministic components of trading volume, so that market liquidity at the beginning and end of the sample are comparable.

1.3.1 Measuring stock and portfolio illiquidity

The Amihud (2002) measure is the most widely used proxy for liquidity because it provides a good approximation of price impact and it can be easily calculated with low-frequency data[9]. It is defined as:

$$AMIHUD_t = \frac{1}{D_t}\sum_{d=1}^{D_t}\frac{|r_{td}|}{\$V_{td}}10^6, \tag{1.1}$$

where t is the time window (year, month or week), r_{td} is the log-return (in percentage) on day d, $\$V_{td}$ is the \$-volume on day d and D_t is the number of days where the ratio $|r_{td}|/V_{td}$ is identified during time window t. Eq. (1.1) is an estimate of the daily stock price movement per \$1 million trading and it can be related to various spread-based measures of liquidity calculated with high-frequency data, such as effective spread, realised spread and price impact[10]. Several adjustments to Eq. (1.1) have been proposed. For instance, Acharya and Pedersen (2005) standardise (1.1) by the ratio of NYSE total market capitalisation at t and at the beginning of the sample, while Chordia et al. (2005), Hameed et al. (2010) and Karolyi et al. (2012) orthogonalise stock liquidity against a set of variables that capture deterministic components in liquidity. Motivated by the same idea, our measure is a logarithmic transformation of the ratio between average

[9] Kyle (1985) defines three dimensions of liquidity: tightness, depth and price impact (resilience). See Goyenko et al. (2009) for a comparison of low and high frequency measures of liquidity.

[10] A measure often used [see, e.g. Hendershott et al. (2011)] to calculate the price impact of transaction q is:

$$PI_q = I_q[\log(MID_{q+\Delta}) - \log(VAL_q)], \tag{1.2}$$

where q is a timestamp that identifies a single transaction, $I_q = \{1, -1\}$ is a trade direction indicator for buyer- or seller-initiated trade, $MID_{q+\Delta}$ is the midquote at time $q+\Delta$ and VAL_q is the security's underlying value at the time when the trade was submitted. In empirical research, VAL_q is approximated by MID_q Thus, Eq. (1.2) is simply the percentage change in the midquote over some Δ period of time, caused by trade q. $AMIHUD_t$ is linked to PI_q because the daily log-returns, r_d, is the cumulative effect of the unsigned price impacts of all transactions of that day d (i.e. the sum of Eq. (1.2) excluding I_q), hence the change in the value of the security due to the trading activity within the day.

daily stock return volatility and deseasonalised average daily stock trading volume. Thus, our main intuition is that seasonalities in $AMIHUD_t$ are caused by seasonalities in trading volume, and only the latter needs to be corrected. In the remainder of the section we formalise our measure of liquidity.

First, we calculate stock-by-stock average daily volatility and trading volume at weekly frequency as:

$$V_{k,t} = \frac{1}{D_{k,t}} \sum_{d=1}^{D_{k,t}} CV_{k,t,d} \tag{1.3}$$

$$\hat{\sigma}_{k,t} = \frac{1}{D_{k,t}} \sum_{d=1}^{D_{k,t}} \log\left(\frac{PH_{k,t,d}}{PL_{k,t,d}}\right) \tag{1.4}$$

for $k = 1, \ldots, N_K$ and $t = 1, \ldots, T$, where N_K is the total number of firms in the sample[11], $PH_{k,t,d}$ and $PL_{k,t,d}$ are the highest and lowest intraday stock prices, respectively, $CV_{k,t,d}$ is the trading volume in local currency and $D_{k,t}$ is the number of days with non-zero volume or non-stale price. We then express $CV_{k,t,d}$ in today's dollar terms, so that it is comparable across countries but its time-series dynamics are not affected by FX rate fluctuations[12]. We cap intraday volatility at 100% and we remove outliers in $CV_{k,t,d}$, defined as observations larger than the 99% percentile.

Second, we aggregate volatility and trading volume to calculate value-weighted indexes for sector s of region r as:

$$\hat{\sigma}_{s,r,t} = \sum_{k \in \mathcal{I}(s,r)_t} w_{k,t} \hat{\sigma}_{k,t} \tag{1.5}$$

$$V_{s,r,t} = \sum_{k \in \mathcal{I}(s,r)_t} w_{k,t} V_{k,t} \tag{1.6}$$

for $s = 1, \ldots, N_S$ and $r = 1, \ldots, N_R$, where $\mathcal{I}(s,r)_t$ is the set of indexes identifying the stocks in sector-region (s, r) at time t, defined as:

$$\mathcal{I}(s,r)_t = \{k \in \mathcal{K}_{st}\} \cap \{k \in \mathcal{K}_{rt}\}, \tag{1.7}$$

[11] The time index t also depends on k, i.e. $t_k = t_{1,k}, \ldots, t_{T,k}$, given that most of the stocks have time series of different length. For ease of notation, we assume t to be equal across stocks.

[12] The use of volumes in local currencies allows to keep FX rates exogenous and to test if exchange rates against the dollar is a demand-side explanatory variable of liquidity commonality.

where $\mathcal{K}_{st}$ ($\mathcal{K}_{rt}$) is the set of indexes of stocks belonging to sector s (region r) in week t. $w_{k,t}$ is the weight of stock k in the value-weighted average at time t. $\mathcal{I}(s,r)_t$ is updated yearly so that only active stocks are considered. We keep only the stocks with full time series.

Third, in order to calculate a price impact that is comparable across markets, we exclude deterministic components from trading volume, $V_{s,r,t}$, using the method proposed by Gallant et al. (1992) and Tauchen et al. (1996). In particular, we control for a time trend and holiday effects for all the weeks of December and January. We estimate these seasonalities in both conditional mean and conditional variance and construct an adjusted trading volume, $AV_{s,r,t}$, that is orthogonal to them. See Appendix 1.B for more details on the deseasonalisation procedure. The steps so far are essentially the calculation of a robust version of the numerator and denominator of Eq. (1.1).

Finally, we combine the volatility and volume to calculate the illiquidity of a portfolios of stocks in sector s of region r as:

$$\mathrm{ILL}_{s,r,t} = \log\left(1 + \frac{\hat{\sigma}_{s,r,t}}{AV_{s,r,t}}\right) \tag{1.8}$$

for $s = 1, \ldots, N_S$ and $r = 1, \ldots, N_R$, where $AV_{s,r,t}$ is the adjusted version of $V_{s,r,t}$. Eq. (1.8) has several advantages compared to the standard formula proposed by Amihud (2002). First, the double effect of cross-sectional and time-series aggregation reduces the estimation error of true market liquidity [see the discussion in Korajczyk and Sadka (2008)]. Second, the value-weighted indexes - which are rebalanced yearly - allow us to calculate continuous time series taking into account the historical composition of the stock market index[13]. Third, the log-range - the numerator of (1.8) - is a consistent estimator of intraday volatility [Alizadeh et al. (2002) prove that it is even more efficient than realised volatility] that does not suffer from the drawbacks of log-returns. For instance, log-returns overestimate large negative real price changes and they can be close to zero when closing prices are close to each other, even if the stock price experi-

[13]The continuous time series can be then deseasonalised using the full time series for the estimation of the deterministic components. We have also deseasonalised individual stocks' volume before calculating the index, however this poses an estimation issue because the panel of stocks is unbalanced. In addition, it is reasonable to expect that market trends and seasonalities are region-specific.

enced a lot of intraday variation. Finally, our method allows to deseasonalise trading volume alone, instead of adjusting the ratio of volatility over volume [as in Acharya and Pedersen (2005) and Karolyi et al. (2012)] where volatility does not need to be adjusted. Brandt et al. (2010), show that returns volatility does not have a time trend and it goes back to its long-run mean after occasional shocks.

Fig. 1.1 shows two examples of the results of the deseasonalisation procedure. The top panels plot the illiquidity indexes of US-Utilities and UK-Utilities, calculated using the unadjusted trading volume, while the bottom panels plot the same time series after the trading volume adjustment. In the US example, one can see a clear downward trend in illiquidity, caused by an increasing trading volume over the last 20 years. In the UK example, half-day trading activity around Christmas creates a predictable pattern in trading volumes, that is then inherited by Eq. (1.8). As one can see from Fig. 1.1, our adjustment removes region-specific predictable patterns, allowing the comparison across regions and simplifying the multivariate analysis of liquidity. In additional unreported tests we run the augmented Dickey-Fuller (ADF) test on the raw and adjusted illiquidity index. We find that we can not reject the null of a unit root in the unadjusted illiquidity index, while we reject the null for adjusted illiquidity. Thus, market (il)liquidity is a stationary process and it should not be differenced, with trends and seasonalities leading to spurious results due to the low power of ADF tests [see, e.g. Schwert (1989) and Ng and Perron (1995)][14].

[Figure 1.1 about here.]

The illiquidity measure in Eq. (1.8) is characterised by strong time series dependence (long memory). High price impact today leads to higher price impact tomorrow and vice versa, which suggests that information on stock liquidity is acquired slowly by the average investors [Bali et al. (2014)].

In the remainder of the section, we compute liquidity shocks, or news, which represent the part of liquidity more difficult to acquire by investors for its ellusive nature. Liquidity shocks will be our dependent variable.

[14] Part of the literature [e.g. Chordia et al. (2000), Brockman et al. (2009)] use the percentage change in liquidity as dependent variable. Hasbrouck and Seppi (2001) point out that this practice would artificially introduce negative autocorrelation in the time series.

1.3.2 Liquidity shocks as dependent variable

In line with the literature [Bali et al. (2014), Acharya and Pedersen (2005), Korajczyk and Sadka (2008), Karolyi et al. (2012)] we filter away the memory of liquidity and focus on its unexpected part, which we call "liquidity shocks". Bali et al. (2014) interpret this as new public information on liquidity. From an econometric point of view, liquidity shocks are covariance stationary and display weak (or zero) time series dependence, which simplifies substantially the multivariate analysis[15]. We estimate liquidity shocks as the residuals from an ARMA(1,1) model:

$$\mathrm{ILL}_{s,r,t} = \alpha_{0,s,r} + \alpha_{1,s,r}\mathrm{ILL}_{s,r,t-1} + \alpha_{2,s,r}\varepsilon_{s,r,t-1} + \varepsilon_{s,r,t}, \tag{1.9}$$

with normally distributed innovations. The model is estimated with ML and liquidity shocks are defined as:

$$\mathrm{L}_{s,r,t} \equiv -\hat{\varepsilon}_{s,r,t}, \tag{1.10}$$

for $s = 1, \ldots, N_S$ and $r = 1, \ldots, N_R$.

$\mathrm{L}_{s,r,t}$, is our variable of interest. In total, there are $N = N_R N_S$ time-series that we stack in vector $\boldsymbol{L}_t$. Fig. 1.2 plots the illiquidity index, $\mathrm{ILL}_{s,r,t}$, and liquidity shocks, $\mathrm{L}_{s,r,t}$ of our $N = 40$ region-sector portfolios.

[Figure 1.2 about here.]

1.4 The sources of common factors in liquidity

The presence of common factors in market liquidity is an established result in empirical finance. Most of the analysis focuses on US markets, with exception of Brockman et al. (2009) and Karolyi et al. (2012). When the analysis is extended to multiple markets, questions arise about which risk factors affect market liquidity. For instance, are global factors enough to explain the cross-asset dependence in liquidity? Are there significant regional factors? Do sector-specific

[15] Conversely, the direct modelling of price impact or volumes require the use of error distributions with non-negative supports or multiplicative error models. We leave this for future research.

factors play a role too and are they more important than regional ones? Is the importance of each factor stable over time? Brockman et al. (2009) answer some of these questions considering global and exchange-specific factors, while Karolyi et al. (2012) consider only country-specific commonality. We consider global, region and sector factors and we use a data-driven approach to rank the importance of the three factors, before extending the model to allow for time-variation in factor importance.

For convenience, we introduce the following notation. Let $i_{s,r}$ be the position of $\mathrm{L}_{s,r,t}$ in the dependent variable vector $\boldsymbol{L}_t$. Then:

$$\mathcal{I}_r = \{n \mid i_n = i_{s,r}, s \in \mathbb{N}_S\} \tag{1.11}$$

$$\mathcal{I}_s = \{n \mid i_n = i_{s,r}, r \in \mathbb{N}_R\} \tag{1.12}$$

are the sets of indices that identify the liquidity of all stocks in region r or sector s, with $\mathbb{N}_R = \{1, \ldots, N_R\}$ and $\mathbb{N}_S = \{1, \ldots, N_S\}$. Furthermore, we define:

$$\mathcal{I}_{SR} = \{n \mid i_n = i_{s,r}, ((i_{s,r})_{r\in\mathbb{N}_R})_{s\in\mathbb{N}_S}\} \tag{1.13}$$

$$\mathcal{I}_{RS} = \{n \mid i_n = i_{s,r}, ((i_{s,r})_{s\in\mathbb{N}_S})_{r\in\mathbb{N}_R}\} \tag{1.14}$$

$\mathcal{I}_{RS}$ ($\mathcal{I}_{SR}$) is the set of indices for the sector-region variables ordered by region (sector). Note that $\mathcal{I}_{RS} = \mathcal{I}_{SR} = \{1, \ldots, N\}$, although the indices correspond to a different order of the variables. Unless specified differently, the dependent variable vector is defined as:

$$\boldsymbol{L}_t \equiv [\mathrm{L}_{n,t}]_{n=1,\ldots,N,\, n\in\mathcal{I}_{RS}} \tag{1.15}$$

We estimate one global, four regional and ten sectoral factors (based on the GICS classification) using the cross-sectional average (CA) estimator, which corresponds to an equally-weighted portfolio. This approach has been followed by the literature on returns comovements [see, e.g., Bekaert et al. (2009)] and gives the advantage of estimating the factors as linear combinations of observable variables, as in a normal portfolio framework. In addition to being simple to compute, the CA estimator of the factors is equivalent to the principal component (PC) estimator when $N/T \to 0$ and factor loadings are static [Westerlund and Urbain (2015)]. Further results based on Monte Carlo simulations show that the CA estimator perform at least as well as the PC estimator [Chudik et al. (2011)]. In our setting we have that $T > N$ and recent literature shows that when factor loadings are unstable over time, the PC estimator consistently estimates the factor space, both in case of random walk and autoregressive process [Bates et al. (2013), Mikkelsen et al. (2017)].

Then, in order to interpret the factors and avoid the multicollinearity issue, we orthogonalise them to create N triplets of mutually uncorrelated factors. The problem of identifying the factors is tackled through the following procedure.

First, we standardise all variables to have zero mean and unit variance, such that they are comparable across countries, and we estimate the global factor as:

$$\hat{G}_t = \frac{1}{N} \sum_{n=1}^{N} \mathrm{L}_{n,t} \tag{1.16}$$

Second, all variables are orthogonalised against $\hat{G}_t$, defining: $\hat{Z}_{n,t} = \mathrm{L}_{n,t} - \hat{\gamma}_n \hat{G}_t$, for $n = 1, \ldots, N$, so that the new variables can be used to estimate the first-stage regional and sectoral factors as:

$$\hat{R}^{(1)}_{r,t} = \frac{1}{N_S} \sum_{\substack{n=1 \\ n \in \mathcal{I}_r}}^{N_S} \hat{Z}_{n,t} \tag{1.17}$$

$$\hat{S}^{(1)}_{s,t} = \frac{1}{N_R} \sum_{\substack{n=1 \\ n \in \mathcal{I}_s}}^{N_R} \hat{Z}_{n,t} \tag{1.18}$$

for $r = 1, \ldots, N_R$ and $s = 1, \ldots, N_S$. By construction, all regional and sectoral factors are uncorrelated with the global factor, but they are correlated with one another, which is a feature that we will take into account in formulating our model.

At this stage, we need to understand which group of factors is the most relevant, i.e. if region or sector drivers are the main determinants of liquidity correlation. This helps us both to answer some of our research questions and defines the order of orthogonalisation of the factors[16]. In order to achieve this goal we

[16]Given that we have three types of factors, there are two possible ways to proceed with the

inspect the correlation matrix of market liquidity and residual liquidity grouping stocks by region or sector. If variables in the same group are related by a common factor, we expect to find a block diagonal structure. Fan, Furger, and Xiu (2016) employ the same procedure to show that the CAPM is not enough to capture cross-dependence of returns and GICS-specific factors are omitted variables. Fig. 1.3 plots the correlation matrix of liquidity, $\boldsymbol{L}_t$, and residual liquidity, $\hat{\boldsymbol{Z}}_t$ (orthogonal to the global factor), using different ordering of the stocks.

[Figure 1.3 about here.]

The top (bottom) left panel reports the correlation matrix of $\boldsymbol{L}_t$ ($\hat{\boldsymbol{Z}}_t$) when the variables are grouped by sectors. The residual vector is arranged as:

$$\hat{\boldsymbol{Z}}_t \equiv [\hat{Z}_{n,t}]_{n=1,\ldots,N,\, n\in\mathcal{I}_{SR}}. \tag{1.19}$$

Similarly, the right panels of Fig. 1.3 report the correlation matrices for variables grouped by regions. The matrices' elements are gradually coloured if the correlation coefficients are higher than 0.1. The block-diagonal structure in the bottom right panel suggests that the liquidity of stocks in the same region tends to move together and some region-specific factors are omitted variables that must be added to the model. Thus, the regional factor is the second most important factor after the G_t, and its estimate is defined as:

$$\hat{R}_{r,t} \equiv \hat{R}^{(1)}_{r,t}, \tag{1.20}$$

for $r = 1, \ldots, N_R$. Finally, the sectoral factor is the remaining variation after controlling for both the global and regional factors:

$$\hat{S}_{s,t} \equiv \hat{S}^{(2)}_{s,t} = \hat{S}^{(1)}_{s,t} - \sum_{r=1}^{N_R} \hat{\gamma}_r \hat{R}^{(1)}_{r,t}, \tag{1.21}$$

for $s = 1, \ldots, N_S$. In Appendix 1.D we show that the orthogonality of the estimated factors holds also conditionally.

orthogonalisation. We could either regress each region factor against the ten sector factors and keep the residual variation as region factor, or regress each sector factor against the four region factors and keep the residual variation as sector factor.

Our findings are in line with the results on stock return comovements. Various researchers have concluded that global, country- and region-specific factors are more important than industry factors in explaining the cross-section of expected returns [e.g., Heston and Rouwenhorst (1995), Griffin (2002), Bekaert et al. (2009)]. Recently, Ando and Bai (2017) estimates the group membership of a stock and reach the same conclusion. This evidence suggests that market liquidity is not a stock characteristic that is determined by the business profile of a company (sector). Instead, it is determined by regional drivers that make the liquidity of different firms become more similar because of geographic proximity, which could be due to region-specific liquidity demand and/or supply shocks.

1.4.1 Time varying importance of factors and ARCH effect

The estimated factors can be used as regressors in a standard factor model of liquidity such as:

$$\mathrm{L}_{nt} = \beta_n^G G_t + \beta_n^R R_{n,t} + \beta_n^S S_{n,t} + \varepsilon_{nt} \tag{1.22}$$

for $n = 1, \ldots, N$, $n \in \mathcal{I}_{RS}$. Table 1.3 reports the least squares estimates of the factor loadings and the adjusted-R^2 for three model specifications: only global factor; global plus region; all three factors. The results show that all three factors add explanatory power to the model and that a one-factor model (such as the one used by Chordia et al. (2000), Karolyi et al. (2012) and many others) is not enough to capture the cross-sectional dependence in liquidity.

[Table 1.3 about here.]

The main limitation of specification (3.1) is that the parameters are assumed to be constant throughout the sample. To assess whether this is realistic assumption, we re-estimate the model using 1-year rolling windows. Fig. 1.4 plots the t-statistics of the three factor loadings over time in four examples, together with two straight lines at $\pm$ 1.96. The figure shows that t-statistics are highly volatile and the regional and sector factors are not always significant. Thus, even if regional factors are, on average, more relevant than sector ones, during certain periods their roles are inverted and it is important to include all sets of factors in a model for liquidity. This evidence finds support from a small literature that argues that the relative influence of industry and country factors depends on the sample period [Baca et al. (2000) and Cavaglia et al. (2000)].

[Figure 1.4 about here.]

Furthermore, a source of uncertainty in the estimation of common liquidity risk is the changing variance of the factors. As pointed out by Forbes and Rigobon (2002) and, recently, Dungey and Renault (2017), time-varying factor volatility is a feature that corrupts correlation coefficients, which complicates the identification of genuine changes in relations between variables. The R^2 measure of liquidity commonality used by Hameed et al. (2010) and Karolyi et al. (2012) suffers from the same drawback. We use Engle's Autoregressive Conditional Heteroscedasticity (ARCH) test to test if factor volatility is time-varying and we reject the null of homoscedasticity at the 95% confidence level for all factors[17].

Thus, we formulate a model that takes into account these data features and we use it to estimate the correlation matrix of market liquidity. First, the model features global, region and sector liquidity common factors. Second, the factor loadings are allowed to change over time to capture time-varying importance of the sources of liquidity risk. Third, the covariance matrix of the factor is also allowed to change over time.

1.5 Model implied liquidity commonality

In this section, we introduce our factor model of liquidity that we use to calculate liquidity commonality within region r, TC, defined as the mean pair-wise correlation among the ten GICS sectors in that region. Then we decompose TC into VD and ED.

[17]The critical values for the test distribution are 3.8415 (95%) and 6.6349 (99%). The test statistics for the 14 factors are: 64.461 (EZ), 62.945 (JP), 46.733 (UK), 38.686 (US), 29.869 (Consumer Discretionary), 13.762 (Consumer Staples), 35.113 (Energy), 43.332 (Financials), 62.341 (Health Care), 57.760 (Industrials), 9.470 (Information Technology), 3.944 (Materials), 204.624 (Telecommunication Services), 76.036 (Utilities).

1.5.1 Model

We formulate a factor model where sector-region liquidity, L_{nt}, is a linear function of global, sector and region factors. Sector liquidity in region r is measured by Eq. (1.10). International factor models are useful to analyse time-series comovements and they have been used, for instance, to test for market integration [Flood and Rose (2005)], to test for contagion across countries and asset classes [Dungey and Martin (2007), Belvisi et al. (2016)] and to analyse return comovements in equity markets [Bekaert et al. (2009), Bekaert et al. (2014)]. In our model, factor loadings change over time to capture the time-varying importance of factors: we assume they evolve as independent autoregressive (AR) processes of order one[18]. The liquidity of sector-region n at time t, L_{nt} is modelled as:

$$L_{n,t} = \beta^G_{n,t} G_t + \beta^R_{n,t} R_{n,t} + \beta^S_{n,t} S_{n,t} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, (\sigma^{\varepsilon}_n)^2) \tag{1.23}$$

$$\beta^G_{n,t} = (1-\phi^G_n)\bar{\beta}^G_n + \phi^G_n \beta^G_{n,t-1} + u^G_{n,t}, \quad u^G_{n,t} \sim iidN(0, (\sigma^{u,G}_n)^2) \tag{1.24}$$

$$\beta^R_{n,t} = (1-\phi^R_n)\bar{\beta}^R_n + \phi^R_n \beta^R_{n,t-1} + u^R_{n,t}, \quad u^R_{n,t} \sim iidN(0, (\sigma^{u,R}_n)^2) \tag{1.25}$$

$$\beta^S_{n,t} = (1-\phi^S_n)\bar{\beta}^S_n + \phi^S_n \beta^S_{n,t-1} + u^S_{n,t}, \quad u^S_{n,t} \sim iidN(0, (\sigma^{u,S}_n)^2) \tag{1.26}$$

where G_t is a global liquidity factor, $R_{n,t}$ is a regional common factor and $S_{n,t}$ is a sectoral common factor. We do not include a constant because $L_{n,t}$ has zero mean by construction. The state equations (1.24) - (1.26) imply that each factor loading at t is a weighted average of the loading at $t-1$ and its long-run average, with the parameter $\phi^{(\cdot)}_n$ regulating the speed of mean reversion, i.e. the memory of the factor loading. The process $\{u_{n,t}\}$ is independent of $\{\varepsilon_{n,t}\}$. Since the variance of the state, $\sigma^{u,(\cdot)}_n$, is constant over time, the model can be written as a linear state-space model and the unknown parameters can be consistently estimated with maximum likelihood estimation via the Kalman Filter after identification of the factors. Mikkelsen et al. (2015) prove that the ML estimator is consistent

[18] An alternative model specification is one where the loadings are static and the factors are dynamic. This setting would still allow to estimate time-varying correlations. However, since the Asset Pricing Theory assumes that stock prices are generated by a set of unpredictable factors, we follow the same rationale in the specification of our model of stock liquidity. We thank Professor Catherine Doz for this suggestion.

if the loadings are stationary, if $T/N^2 \to 0$ and if the maximum variation of the factors is bounded. In our case, factors are observable (cross-sectional averages are essentially equally-weighted portfolios) and we expect the error in measurement in the dependent variable to be smoothed out by the use of region-sector portfolios. In Appendix 1.C we report initial parameters and estimates of the Kalman Filter.

This method allows us to estimate factor loadings, and model-implied covariances, at every time t and with dynamics (memory, variance) that are dictated by the data. So far the literature has estimated time-varying factor loadings using rolling window estimation [see, among others, Bekaert et al. (2009), Karolyi et al. (2012)], with results heavily depending on the length of the window[19]. Inoue et al. (2017) also show that this greatly affects the forecasting performance of a model.

Fig. 1.5 plots the time series of factor loadings for liquidity of two portfolios. Our model allows us to take into account different types of structural instability. For example, the liquidity of Eurozone's Consumer Discretionary stocks has a fairly stable exposure to global liquidity shocks, with an autoregressive parameter of -0.1095, with the top panel of Fig. 1.5 showing that the spikes in the global factor loadings are short lived. Instead, the exposure to regional liquidity shocks is very persistent (AR parameter of 0.8979) and shocks are not easily absorbed. Fig. 1.5 shows another example, where the global factor loading of US Financials has a variance very close to zero and it is not different from the OLS estimate.

[Figure 1.5 about here.]

The dynamic factor model (1.23) - (1.26) nests the static version in Eq. (3.1), where the factor loadings are constant over time. We use a standard likelihood ratio test (LRT) to assess which model fits the data better. The LRT for the full

[19] Gagliardini et al. (2016) prove consistency of the two-pass estimation of a factor model with time-varying parameters that are function of stock specific and macroeconomic variables, as in Shanken (1990).

model can be written as:

$$LRT = -2\log\left(\frac{\prod_{n\in\mathcal{I}_{RS}} \ell_n(\widetilde{\theta}_n)}{\prod_{n\in\mathcal{I}_{RS}} \ell_n(\widehat{\theta}_n)}\right) \tag{1.27}$$

$$= -2\left[\sum_{n\in\mathcal{I}_{RS}} \log \ell_n(\widetilde{\theta}_n) - \sum_{n\in\mathcal{I}_{RS}} \log \ell_n(\widehat{\theta}_n)\right] \tag{1.28}$$

where $\widehat{\theta}_n$ is the vector of estimated parameters in (1.23) - (1.26), $\widetilde{\theta}_n$ is the restricted parameter vector that is related to the static model in (3.1) and $\ell(\cdot)$ is the likelihood function. The Null hypothesis that the static model fits the data better than our dynamic model is rejected at 99% confidence level, with a LRT statistic of 795244.2. In the remainder of the section, we derive our measure of within-region liquidity commonality, TC, and we decompose it into VD and ED.

1.5.2 Total commonality

Let $\boldsymbol{F}_{n,t} = [G_t, R_{n,t}, S_{n,t}]$ be the vector of unknown factors. Assuming that $\mathbb{E}[\boldsymbol{F}_{n,t}] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{F}_{n,t}\varepsilon_{n,t}] = \mathbf{0}$ holds for all n, it follows that the unconditional covariance between any pair of sector-region liquidity indexes is equal to:

$$cov(\mathrm{L}_n, \mathrm{L}_m) = \underbrace{\mathbb{E}[\boldsymbol{\beta}'_{n,t}\boldsymbol{F}'_{n,t}\boldsymbol{F}_{m,t}\boldsymbol{\beta}_{m,t}]}_{\text{Model implied cov}} + \underbrace{\mathbb{E}[\varepsilon_{n,t}\varepsilon_{m,t}]}_{\text{Residual cov}} \tag{1.29}$$

for $n = 1, \ldots, N$, $m = 1, \ldots, N$, $n \neq m$ and $n, m \in \mathcal{I}_{RS}$, where the first and second term are the model-implied and residual covariance, respectively. If the factor model fully describes the comovements between variables, the residual covariance is negligible. Thus, the empirical counterpart of Eq. (1.29), conditional on the estimated factor loadings and covariances, can be written as:

$$c\hat{o}v_t(\mathrm{L}_n, \mathrm{L}_m) = \hat{\boldsymbol{\beta}}'_{n,t}\hat{\Sigma}^F_{n,m,t}\hat{\boldsymbol{\beta}}_{m,t} \tag{1.30}$$

and can be further simplified as:

$$c\hat{o}v_t(\mathrm{L}_n, \mathrm{L}_m) = \hat{\beta}^G_{n,t} v\hat{a}r_t(G_t)\hat{\beta}^G_{m,t} + \hat{\beta}^R_{n,t} c\hat{o}v_t(R_{n,t}, R_{m,t})\hat{\beta}^R_{m,t} + \hat{\beta}^S_{n,t} c\hat{o}v_t(S_{n,t}, S_{m,t})\hat{\beta}^S_{m,t}. \tag{1.31}$$

Eq. (1.31) gives some insights into the covariance structure implied by the model: since stocks n and m are in the same region, the second term simplifies to the variance of the regional factor.

The model-implied covariance has two sources of variation: the factor loadings and $\hat{\Sigma}^F_{n,m,t}$, which can be either constant or time-varying. This matrix is diagonal because of the orthogonalisation procedure and its entries correspond to the variance of the global and regional factors, and the covariances among sectoral factors. We estimate the variance of the global factor G_t using a univariate GARCH(1,1) with normal innovations. For the regional and sector factors, we use the Engle (2002)'s Dynamic Conditional Correlations (DCC) model of order (1,1) with GARCH(1,1) marginal conditional volatility processes with normal innovations. The DCC model is estimated using quasi-maximum likelihood estimation separately for the four regional factors and ten sectoral factors. This is essentially a two-step procedure, where we first identify the orthogonal factors within each group and then estimate the conditional covariance between the factors. This framework is similar to the estimation of a factor GARCH with orthogonal factors, where the dependent variable is assumed to be explained by a small number of orthogonal factors but allowing for Granger causality in the variances. The GARCH-type volatilities displayed by the factors is such that today's volatility of one factors may affect tomorrow's volatility of another factor [see, among others, Alexander (2001) and van der Weide (2002)]. Hafner and Preminger (2009) show that the quasi-maximum likelihood estimator of the factor GARCH parameters in the second step is consistent and asymptotically normal. They assume the factors to be either observable or function of underlying variable, which is similar to our case since we use equally-weighted portfolios to identify the factors.

Following the literature on return comovements [see e.g. Bekaert et al. (2009, 2014)], we estimate liquidity commonality at time t as the MPC implied by the factor model, assuming that residual cross-correlation is negligible. We can check if this assumption holds by estimating the residual covariance in Eq. (1.29) and use it to calculate residual MPC. Examination of the plot of residual MPC indicates that its magnitude is negligible, so that we can assume it is zero. Thus, we estimate TC at time t for region r as:

$$\mathrm{TC}_t^r = \frac{1}{N_S(N_S-1)/2} \sum_{\substack{n \in \mathcal{I}_r}} \sum_{\substack{m \in \mathcal{I}_r \\ m>n}} c\hat{o}rr^F_{n,m,t} \tag{1.32}$$

with

$$c\hat{o}rr^F_{n,m,t} = \frac{1}{\sqrt{v\hat{a}r^F_{n,t}}\sqrt{v\hat{a}r^F_{m,t}}} \times \hat{\boldsymbol{\beta}}'_{n,t} \hat{\Sigma}^F_{n,m,t} \hat{\boldsymbol{\beta}}_{m,t} \tag{1.33}$$

where $v\hat{a}r^F_{n,t} = \hat{\boldsymbol{\beta}}'_{n,t}\Omega^F_{n,t}\hat{\boldsymbol{\beta}}_{n,t}$ and $\Omega^F_{n,t}$ is diagonal, containing the variances of the three mutually uncorrelated factors of sector-region n. $v\hat{a}r^F_{m,t}$ is defined accordingly.

1.5.3 Exposure-driven commonality

Forbes and Rigobon (2002) prove that using Pearson correlation coefficients in comovement analysis is misleading because correlations are biased upwards during periods of market turmoil[20]. Thus, we introduce ED, a variation of the MPC that is not affected by the heteroscedasticity of the factors and it is driven instead by changes in the factor loadings, i.e. the exposure of each sector-region to the global, region or sector common factors. ED at time t for region r is estimated as:

$$\mathrm{ED}_t^r = \frac{1}{N_S(N_S-1)/2} \sum_{\substack{n \in \mathcal{I}_r}} \sum_{\substack{m \in \mathcal{I}_r \\ m>n}} c\hat{o}rr^F_{n,m,t,ED} \tag{1.34}$$

[20]Consider a model that describes the relationship between a single stock returns and a risk factor:

$$r_t = \alpha + \beta F_t + \epsilon_t$$

with assumption $E[\epsilon_t] = E[F_t\epsilon_t] = 0$. Then, split the sample in two parts, one with low factor volatility, σ_f^ℓ and the other with high σ_f^h, and further assume that β is constant and $E[\epsilon_t^2] = c$ for all t. In this case, Forbes and Rigobon (2002) prove that: $\rho^h = \beta\frac{\sigma_f^h}{\sigma_r} > \beta\frac{\sigma_f^\ell}{\sigma_r} = \rho^\ell$. Thus, it follows that $R_h^2 > R_\ell^2$, where R_h^2 and R_ℓ^2 are the coefficient of determination in the high and low volatility sample periods, respectively.

where

$$\hat{corr}^F_{n,m,t,ED} = \frac{1}{\sqrt{\hat{var}^F_n}\sqrt{\hat{var}^F_m}} \times \hat{\boldsymbol{\beta}}'_{n,t}\hat{\Sigma}^F_{n,m}\hat{\boldsymbol{\beta}}_{m,t} \tag{1.35}$$

so that $\hat{\Sigma}^F_{n,m}$ is constant, held fixed to its long-run value. Bekaert et al. (2014) and Dungey and Renault (2017) use changes in loadings to detect contagion across markets.

1.5.4 Volatility-driven commonality

With the same rationale, we introduce VD, a variation of the MPC that is completely driven by the volatility of the factors. VD at time t for region r is estimated as:

$$\text{VD}^r_{t,ED} = \frac{1}{N_S(N_S-1)/2}\sum_{n\in\mathcal{I}_r}\sum_{\substack{m\in\mathcal{I}_r \\ m>n}} \hat{corr}^F_{n,m,t,VD} \tag{1.36}$$

where

$$\hat{corr}^F_{n,m,t,VD} = \frac{1}{\sqrt{\hat{var}^F_{n,t}}\sqrt{\hat{var}^F_{m,t}}} \times \hat{\boldsymbol{\beta}}'_n\hat{\Sigma}^F_{n,m,t}\hat{\boldsymbol{\beta}}_m \tag{1.37}$$

so that $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_m$ are constant, held fixed to their long-run value.

To summarise, Eq. (1.32) - (1.36) allow us to identify three types of commonality in liquidity. The TC in (1.32) is driven by both factor heteroskedasticity (changing variances and covariances) and loadings, the ED in (1.34) changes over time only due to changes in the factor loadings and the VD in (1.36) is driven only by time-variation in factor covariances.

1.6 The determinants of commonality in liquidity

In this section, we turn to the analysis of the economic determinants of liquidity commonality. First, we present estimated commonality. Second, we relate TC^r_t, ED^r_t and VD^r_t to a set of exogenous variables that approximate aggregate demand and supply of liquidity, for each region r.

Fig. 1.6 presents graphs of TC and ED in each of our four currency regions. The graphs suggest that liquidity shocks are positively correlated throughout the entire sample period but the two dependent variables are not always in agreement. During the Great Financial Crisis, ED and VD they have a different behaviour in all countries. In particular, ED experiences some negative shifts during financial crises. However, the patterns observed in Fig. 1.6 could be due to statistical noise and we need further analysis to draw any conclusion. Estimates of VD are not reported to save space.

[Figure 1.6 about here.]

Regional factors can be correlated with one another, which implies that TC_t^r, ED_t^r and VD_t^r can be correlated across regions. To take this into account, following Hameed et al. (2010) and Karolyi et al. (2012), we estimate SUR models where liquidity commonality is explained by a set of variables controlling for capital market conditions and proxies for demand-side and supply-side liquidity factors. To match the frequency of the estimated liquidity commonality with that of the exogenous variables, we resample all variables in the last week of the month. The $N_R \times 1$ vector of commonality in liquidity in month t, $\boldsymbol{Y}_t = [\text{TC}_t^1, \dots, \text{TC}_t^{N_R}]'$, is related to a set of exogenous variables by the following model:

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} + \sum_j \boldsymbol{A}_j \circ \boldsymbol{X}_t^j + \sum_k \boldsymbol{C}_k \circ \boldsymbol{W}_t^k + \boldsymbol{\delta} t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim iidN(\boldsymbol{0}, \boldsymbol{V}) \tag{1.38}$$

where $\boldsymbol{X}_t^j$ denotes supply and demand factors, $\boldsymbol{W}_t^j$ are control variables to capture the general variation in capital market conditions and $\circ$ is the Hadamard product. The coefficients in (1.38) are estimated using an iterative procedure that stops when the estimates converge. To investigate the explanatory power of each exogenous variable, we impose an homogeneity assumption, restricting the parameters in (1.38) to be the same across all regions, which implies that the vectors $\boldsymbol{A}_j$ and $\boldsymbol{C}_k$ become scalars. In Appendix 1.D we report the results of a SUR model with heterogenous parameters. Eq. (1.38), in its homogeneous version, is estimated also with ED and VD as dependent variables.

To control for overall capital market conditions we use market returns, market volatility and average market turnover in each region[21]. These conditions

[21] We do not include market liquidity because its high correlation with market volatility

might influence demand and supply and, subsequently, commonality. As such, we want to control for them before exploring precise demand and supply channels.

Tables 1.4 - 1.6 report the estimation results of Eq. (1.38) for TC_t^r, ED_t^r and VD_t^r as the dependent variable, respectively. To gauge the goodness of fit of the model, we report the average R^2 of individual OLS regressions with the same specification of the relevant SUR model. This procedure has been chosen versus the McElroy-R^2 for SUR models[22] because the parameter homogeneity restrictions substantially reduces the McElroy-R^2. In this section, we comment on the relationship between general capital market conditions and liquidity commonality, while we leave the analysis of demand and supply channels to the next sections.

We find that the R^2-commonality measure used by the extant literature is only able to capture the part of liquidity commonality that is driven by the volatility of the factors. In particular, the results of the SUR model in the VD case - Table 1.5 - replicate closely the "More developed countries" column of Table 5 in Karolyi et al. (2012), which uses a set of countries comparable to ours. This result sheds light on the limits of the procedure used by the literature to estimate liquidity commonality in equity markets. Table 1.5 shows that the coefficient on market turnover is negative and statistically significant, that on market returns is insignificantly different from zero and the market volatility coefficient is positive and significant. We find the same results both in our first model (only control variables) and in our last model, where all variables are included in the model. Differently from Karolyi et al. (2012), in our sample we find a positive and significant trend in commonality that could be interpreted as increasing market integration. The coefficients on large (bigger than one standard deviation) negative market returns are insignificant. This lost significance could be

would induce multicollinearity.

[22] The McElroy-R^2 is calculated as:

$$McR^2 = 1 - \frac{\hat{\epsilon}'(S^{-1} \otimes I)\hat{\epsilon}}{Y'(S^{-1} \otimes I)Y},$$

where $\hat{\epsilon}$ and Y are $TM \times 1$ vectors of stacked residuals and dependent variables respectively, of the M equations in the SUR model. S is the $M \times M$ covariance matrix of residuals and I is a $N \times N$ identity matrix. The Kronecker product between S^{-1} and I forms a $MT \times MT$ matrix which makes the matrix conformable.

due to our use of large caps (liquid stocks), whose liquidity is less affected by extreme market conditions.

Our model allows us to bring the analysis one step forward to identify which part of liquidity commonality is driven by certain market conditions. The main result is that there exists an offsetting mechanism of factor exposures that makes the TC increase less than it would if the loadings were constant. In particular, the control variables are not significant in explaining the combined effect of factor loadings and volatility in the TC case, reported in Table 1.4. Table 1.6 shows that this happens because market volatility and turnover change sign when ED is used as dependent variable. Thus, when volatility increases, the exposure of sector liquidity to common factors move in different directions.

[Table 1.4 about here.]

[Table 1.5 about here.]

[Table 1.6 about here.]

In the remainder of the section, we will explore this mechanism for variables approximating the supply-side and demand-side of liquidity.

1.6.1 The effect of funding liquidity constraints on commonality

Various theoretical and empirical papers show that constraints in funding liquidity influence market liquidity [Coughenour and Saad (2004), Brunnermeier and Pedersen (2009), Hameed et al. (2010)]. Since market makers provide liquidity across many securities [Menkveld (2013), Anand and Venkataraman (2016)], a single shift in the cost of deploying capital could influence the market liquidity of several stocks at the same time, hence increasing commonality. To test this hypothesis we use five indirect measures of funding liquidity constraints that are related to the aggregate supply of liquidity.

Our first measure is the US commercial paper (CP) spread for non-financial corporations. Krishnamurthy (2002) shows that the CP spread reflects the liquidity premium required by money market funds, who are the main investors in CPs. Other authors that use CP spread to approximate liquidity supply include Gatev and Strahan (2006) and Hameed et al. (2010). Our second proxy is the short interest rate for each region. Adrian and Shin (2008) show that short interest rates drive the cost of leverage of broker-dealers and they are inversely related to the size of intermediary balance sheets. We expect an increase in interest rates to be followed by decrease in leverage and funding liquidity[23]. Our third proxy is the stock market performance of the banking sector in each region. A fall in the market valuation of financial intermediaries is likely to signal a weakening of their balance sheets [Hameed et al. (2010)]. Our fourth measure is the VIX Index, which reflects expectations of aggregate market volatility and is a proxy of perceived market risk [Adrian and Shin (2010)]. Higher volatility creates uncertainty in the valuation of assets (collateral) and thus it becomes more difficult to obtain funding [Brunnermeier and Pedersen (2009)]. Our fifth measure is credit risk, which we measure with the TED spread. It captures the risk of default on interbank loans. As an alternative measure we use the St. Louis FED stress indicator, which is a measure of general financial stress that combines interest rates, credit spreads and volatility measures.

Table 1.4 shows that, as funding liquidity becomes constrained, commonality in liquidity increases, consistent with the collateral-based theory of Brunnermeier and Pedersen (2009) and the evidence in Hameed et al. (2010). The coefficients on the CP spread and the TED spread are positive and significant at 99% level. In particular, a 100 basis points increase in the CP spread is connected to a TC that is 5.53% higher, while when the TED spread is 100 basis points higher, TC is expected to increase by 2.78%. Thus, since an increase in the CP or the TED spread is associated with an overall deterioration of aggregate market liquidity, transaction costs increase and become more correlated across many stocks when funding liquidity is low. As a robustness check, we also test the effect of CP spread of financial corporations on commonality and we find a similar effect. Furthermore, decomposing total correlation into volatility- and exposure-driven components sheds light on how commonality is influenced by the liquidity supply factors. Table 1.5 and 1.6 show the two cases. Both coefficients on CP spread and TED spread remain strong (4.61 and 2.12, respectively)

[23] Adrian and Shin (2010) show that leverage is indeed procyclical.

and significant at 99% level when VD is used a dependent variable, while they reverse sign in the ED case. Thus the part of commonality that is driven by volatility is positively affected by funding constraints, while exposure-driven commonality moves in the opposite direction. A possible explanation for this finding for the ED component is that when a market maker has fewer resources, she needs to decide how best to allocate them and might decide to focus on some sectors according to his or her expertise. The negative and significant coefficients of short-term interest rates, the VIX index and the FED stress indicator in Table 1.6 are supportive of our interpretation.

1.6.2 The effect of correlated liquidity demand on commonality

Another factor affecting commonality is correlated liquidity demand. This can arise either from coordinated trading of institutional investors that are hit by a common liquidity shock, such as an outflow of capital [Kamara et al. (2008), Koch et al. (2016)], or from irrational trading due to market sentiment [Baker and Wurgler (2006)]. In both cases, large order imbalances would consume the liquidity available on the order book and, if liquidity providers are not able to restore the depth of the book, price impact would increase. To test this hypothesis we use four measures of investor behaviour that are related to aggregate demand for liquidity.

Our first and second measures are US dollar ETF trading volume and net equity flows, which both capture correlated trading of institutional investors. Koch et al. (2016) show that stocks owned by mutual funds with large turnover exhibit greater commonality than other stocks. ETF volume is a proxy of index-related basket trading [Karolyi et al. (2012)], which has been shown to increase commonality by Gorton and Pennacchi (1993). Net equity flows tells us if US investors are net buyers or sellers of local stocks and Kamara et al. (2008) argue that foreign capital inflows are mainly driven by institutional investors. Our third measure is exchange rate changes (with respect to the SDR, a basket of currencies defined by the IMF). A depreciation of local currency should attract foreign investors and increase commonality [Karolyi et al. (2012)]. Finally, we approximate aggregate investor sentiment by the US investor sentiment index of Baker and Wurgler (2006). Although Huberman and Halka (2001) and Baker and Wurgler (2006) provide evidence in favour of sentiment-based commonality in liquidity, there is no clear theoretical prediction on the direction of the effect on commonality.

Table 1.4 shows that ETF trading volume (expressed as percentage of the stock market capitalisation in each region) and the US sentiment index are positively and negatively related to total commonality, respectively. Their estimated coefficients are significant with 99% confidence. In particular, an increase in local ETF trading volume by 1% of market capitalisation is associated with a 1.9% higher total commonality. Tables 1.5 and 1.6 show that ETF volume is positive and significant also in the ED and VD case, representing the only variable in our study with such a consistency across our three tests. Thus, index-related basket trading explains the variability of all three types of liquidity commonality, strongly supporting the empirical results of Koch et al. (2016). On the other hand, US sentiment is negatively related to TC, implying that liquidity is less associated among the ten sectors in the economy when people are optimistic. It is possible that when all investors are optimistic, individual investors tilt their portfolios towards more risky assets but these tilts are investor-specific and subjective. This is in line with the finding of Baker and Wurgler (2006) that young, high-volatility, non-dividend paying, growth companies are more sensitive to investor sentiment. So when investor sentiment is high, demand for, e.g., IT stocks might be higher than Energy stocks, making demand for liquidity more idiosyncratic. When investors become more pessimistic, their portfolio holdings converge and also their demand for liquidity. The coefficient on US sentiment remains negative and significant in the VD case and it is not significantly different from zero in the ED case.

In general, the relationship between the significant demand-side determinants (ETF volume and US sentiment) and liquidity commonality is consistent across TC, VD and ED, while the supply-side variables have the same impact on TC and VD, while flipping sign in the ED case. This evidence suggests that the dominant regional common shocks that make liquidity co-move are demand driven.

1.7 Conclusions

In this paper we studied the determinants of commonality in liquidity. We measured the contribution of demand and supply-side variables that, amongst other things, provide information on investor flows and on the funding conditions faced by intermediaries. We estimated liquidity commonality using a novel factor model where both factor loadings and factor volatilities vary over time, and a stock's liquidity is exposed to global, regional and sectoral factors. Our framework allowed us to estimate *total commonality* (TC), defined as the conditional mean pair-wise correlation implied by the model, and to identify *volatility-driven commonality* (VD), defined as the component of TC driven by the time-variation in the factor covariance matrix, with factor loadings are assumed constant over time. We also identified *exposure-driven commonality* (ED), defined as the component of TC driven by time-varying factor loadings, while the factor covariance matrix is held constant at its long-run value. Then, we used seemingly unrelated regression to measure the impact of demand and supply economic determinants on TC, VD or ED.

The empirical analysis, using a set of 1909 firms covering 15 countries over the period 11 January 2000 to 20 January 2017, provides some interesting findings. First, the sources of risk that drive the cross-section of liquidity are mainly global and region-specific. Sector factors are less important. Thus the listing location of a company is more important than its business mix in determining liquidity. Second, we found that both demand and supply side factors play a role in explaining commonality. Region-specific index-related trading (proxied by ETF trading volume) is positively related to commonality, supporting the results of Koch et al. (2016). Instead, global market sentiment (proxied by the US sentiment index of Baker and Wurgler (2006)) is negatively related to commonality, implying that, when people become more optimistic, price impact is less correlated among stocks. A possible explanation for this is that when market sentiment is high, a set of investors become more risk averse and tilt their portfolios towards more risky sectors, such as those containing young, non-dividend paying, growth stocks. These have been found to be more sensitive to market sentiment in Baker and Wurgler (2006). On the supply side, in line with Hameed et al. (2010), when funding constraints are binding, liquidity commonality increases. This is driven by the VD component of total commonality and is partially offset by ED. Thus, by controlling for factor volatility, we find that price impact becomes more idiosyncratic when intermediaries find it more costly to raise capital. A possible explanation for this evidence is that market makers focus on sectors where they have more expertise when they are capital constrained, removing liquidity from those where they have little knowledge or experience. Then, if expertise varies across intermediaries, liquidity in different sectors will be provided by different intermediaries and idiosyncratic shocks to their risk-bearing capacity will cause variation in liquidity across sectors. Third, we found that the coefficients of demand-side variables, when significant, are consistent across TC, VD and ED. Instead when supply shocks take place, there is an offsetting mechanism of factor exposures that makes the TC increase less than it would if the loadings were constant. We interpret this result as evidence that the correlated liquidity demand of institutional investors is the strongest regional economic force that makes stock liquidity co-move.

Table 1.1: Summary statistics, trading data (daily)

The table reports the summary statistics of the trading activity and market capitalisation of the companies in our sample. For each region we consider only the companies with at least 60 months of valid observations of last price, volume and shares outstanding, to calculate these statistics. *Mean* and *Med* are the time-series averages of the cross-sectional mean and median, respectively. *Min* and *Max* are the minimum and maximum over time and across all stocks in a region. The remaining statistics are cross-sectional averages of the relevant coefficient: *StDev, Skw* and *Krt* are the average standard deviation, skewness and kurtosis; *AR1* is the OLS estimate of the first autocorrelation coefficient.

	Mean	Med	Min	Max	StDev	Skw	Krt	AR1
Returns(%)								
EZ	0.026	-0.018	-60.811	123.731	2.269	0.353	16.950	0.003
JP	0.026	-0.056	-36.923	50.000	2.416	0.192	8.800	-0.021
UK	0.035	-0.013	-72.487	90.057	2.281	0.039	22.781	0.014
US	0.055	0.018	-89.631	143.905	2.428	0.314	18.802	-0.030
$-Volume(mil)								
EZ	71.610	19.205	30.638	15556.948	65.534	6.429	142.247	0.611
JP	43.888	22.426	65.886	4372.289	38.021	4.197	58.446	0.691
UK	39.364	8.559	26.413	7800.854	36.892	6.526	131.462	0.528
US	169.490	81.680	378.492	29875.677	139.969	5.206	98.438	0.668
$-Market Cap(bil)								
EZ	15.022	6.124	0.019	342.984	6.759	0.621	3.250	0.998
JP	9.509	4.473	0.050	395.957	3.708	0.686	3.659	0.997
UK	10.306	2.419	0.009	299.338	4.072	0.689	3.116	0.999
US	23.840	9.847	0.035	774.691	10.476	0.692	3.367	0.998

Table 1.2: Summary statistics, exogenous variables (monthly)

The table reports time series summary statistics for the exogenous variables used to capture capital market conditions and the level of demand/supply of liquidity. *VIX* is the Chicago Board Options Exchange (CBOE) VIX Index; *TED spread* is the difference between three-month LIBOR and three-month US T-bill; *FED stress* is the St. Louis FED financial stress indicator; *US sentiment* is the index calculated by Baker and Wurgler (2006); *US CP spread* is the difference between the interest rate on a 90-day AA US commercial paper and the three-month US T-bill, for non-financials and financial institutions; *Short term* is the local 3-month T-bill rate; *Banks* is the percentage returns on the Datastream Banks index; *Exchange rate* is the percentage change in the value of a region's currency relative to the Special Drawing Rights basket; *ETF volume* is the dollar trading volume in local iShares Morgan Stanley Capital International (MSCI) ETFs traded on US markets, calculated as a percentage of the total stock market capitalisation; *Net equity flow* is the difference between gross sales of foreign stocks by foreigners to US residents and gross purchases of foreign stocks by foreigners from US residents; *Market turnover* is the average daily turnover, expressed as shares-volume in percentage of shares outstanding; *Market volatility* is the standard deviation of the daily market return.

	Mean	Med	Min	Max	StDev
VIX	20.317	17.840	10.310	68.510	8.543
TED spread	0.433	0.280	0.120	3.150	0.418
FED stress	-0.244	-0.532	-1.651	4.853	1.086
US sentiment	0.160	0.052	-0.866	3.076	0.725
US CP spread non-fin	0.214	0.120	0.020	1.400	0.244
US CP spread fin	0.297	0.160	0.050	2.520	0.354
Short-term rate					
EZ	1.643	1.978	-0.975	5.090	1.704
JP	0.121	0.058	-0.379	0.665	0.203
UK	2.627	3.507	0.202	5.990	2.171
US	1.517	0.780	-0.010	6.180	1.822
Banks					
EZ	0.006	0.451	-33.273	31.035	9.276
JP	-0.005	-0.461	-30.404	25.065	8.057
UK	-0.087	-0.050	-29.544	33.903	7.424
US	0.359	0.740	-34.829	32.477	8.010
Exchange rate					
EZ	-0.031	-0.121	-5.336	7.156	1.773
JP	0.070	0.026	-8.706	6.280	2.097
UK	0.121	-0.009	-4.369	7.840	1.595
US	0.014	0.010	-2.796	3.206	1.129
ETF volume					
EZ	0.191	0.040	1.904e-05	1.360	0.303
JP	1.193	1.309	0.021	6.192	0.880
UK	0.152	0.103	0.005	0.901	0.150
US	0.429	0.474	0.018	2.175	0.349
Net equity flow					
EZ	0.534	0.725	-14.897	9.827	4.209
JP	2.280	1.565	-12.013	23.812	6.705
UK	1.430	1.499	-11.362	17.053	3.370
Market Turnover					
EZ	0.442	0.443	0.211	0.947	0.114
JP	0.451	0.453	0.194	0.902	0.136
UK	0.383	0.335	0.162	0.840	0.162
US	0.758	0.681	0.434	1.692	0.237
Market Volatility					
EZ	1.102	0.976	0.347	4.618	0.592
JP	1.200	1.068	0.348	5.253	0.557
UK	0.995	0.841	0.265	4.867	0.598
US	0.980	0.856	0.201	5.014	0.628

Table 1.3: Static factor model estimates

The table reports the estimates of the static factor model in Eq. (3.1), which features one global, one region and one sector factor and static factor loadings. Factor loadings' estimates are from ordinary least squares. $\bar{R}^2$ is the adjusted R^2 of the model. $\bar{R}^2_G$ is the adjusted R^2 of a model with global factor only, while $\bar{R}^2_{GR}$ is the adjusted R^2 of a model with global and regional factors only. The factors are calculated as cross-sectional averages of standardised illiquidity shocks. By construction they have zero mean but not necessarily unit variance. The dependent variable is illiquidity innovations in each sector-region portfolio. It has a mean of zero but it is not standardised. Estimates are multiplied by 100.

	β^G	t-stats	β^S	t-stats	β^R	t-stats	$\bar{R}^2$	$\bar{R}^2_G$	$\bar{R}^2_{GR}$
Consumer Discretionary, EZ	0.76	25.93	0.59	16.83	0.57	15.78	0.58	0.32	0.44
Consumer Staples, EZ	0.56	25.54	0.55	23.10	0.47	17.56	0.63	0.28	0.41
Energy, EZ	0.25	26.08	0.20	20.37	0.22	17.99	0.62	0.30	0.44
Financials, EZ	0.43	26.52	0.39	20.85	0.40	20.23	0.64	0.29	0.46
Health Care, EZ	0.52	21.95	0.51	19.00	0.50	16.90	0.57	0.24	0.38
Industrials, EZ	0.86	30.83	0.50	14.08	0.69	20.28	0.64	0.39	0.56
Information Technology, EZ	0.75	22.74	0.69	19.20	0.47	11.56	0.54	0.27	0.34
Materials, EZ	1.09	32.15	0.67	17.26	0.81	19.37	0.66	0.40	0.55
Telecommunication Services, EZ	0.27	17.94	0.32	18.99	0.31	16.55	0.52	0.18	0.33
Utilities, EZ	0.52	20.95	0.45	16.53	0.45	14.83	0.52	0.24	0.37
Consumer Discretionary, JP	0.32	18.16	0.29	13.88	0.38	25.19	0.57	0.16	0.48
Consumer Staples, JP	0.97	17.59	1.01	16.77	1.23	25.77	0.59	0.15	0.46
Energy, JP	2.54	10.36	4.58	18.43	3.81	17.88	0.47	0.06	0.26
Financials, JP	0.59	14.89	0.74	16.23	0.71	20.92	0.52	0.12	0.37
Health Care, JP	0.52	13.76	0.74	17.30	0.76	23.13	0.54	0.10	0.38
Industrials, JP	1.19	22.90	0.88	13.42	1.45	32.28	0.67	0.20	0.60
Information Technology, JP	0.59	20.66	0.48	15.62	0.66	26.26	0.61	0.19	0.50
Materials, JP	1.38	17.25	1.09	11.85	1.96	28.35	0.59	0.14	0.52
Telecommunication Services, JP	0.28	14.50	0.34	15.72	0.33	19.43	0.49	0.12	0.35
Utilities, JP	0.95	9.31	1.53	13.55	1.41	15.87	0.38	0.06	0.24
Consumer Discretionary, UK	2.42	25.04	2.18	18.82	2.43	21.95	0.63	0.27	0.48
Consumer Staples, UK	0.58	26.05	0.48	19.48	0.62	23.98	0.65	0.27	0.50
Energy, UK	0.43	25.17	0.32	18.62	0.40	20.54	0.62	0.28	0.47
Financials, UK	0.59	22.89	0.53	17.67	0.57	19.05	0.58	0.25	0.43
Health Care, UK	0.34	16.79	0.37	16.38	0.42	18.18	0.50	0.16	0.35
Industrials, UK	2.53	24.99	1.95	15.16	2.16	18.61	0.58	0.30	0.47
Information Technology, UK	3.93	9.57	7.62	17.25	6.11	12.99	0.39	0.06	0.18
Materials, UK	1.06	24.28	1.17	23.18	0.84	16.80	0.62	0.26	0.38
Telecommunication Services, UK	0.58	18.23	0.57	16.48	0.57	15.60	0.49	0.19	0.34
Utilities, UK	1.05	17.15	1.21	17.98	1.38	19.67	0.54	0.16	0.36
Consumer Discretionary, US	0.34	19.86	0.34	16.14	0.33	17.86	0.53	0.21	0.39
Consumer Staples, US	0.09	22.90	0.09	19.50	0.10	24.44	0.63	0.22	0.47
Energy, US	0.07	18.72	0.09	22.50	0.08	19.65	0.59	0.17	0.35
Financials, US	0.12	22.98	0.10	16.77	0.11	21.13	0.59	0.25	0.46
Health Care, US	0.11	23.09	0.09	17.21	0.09	17.97	0.57	0.26	0.42
Industrials, US	0.10	22.49	0.11	20.55	0.09	19.39	0.60	0.23	0.41
Information Technology, US	0.05	22.68	0.04	19.57	0.04	18.35	0.59	0.24	0.41
Materials, US	0.35	13.63	0.55	18.41	0.41	15.01	0.46	0.11	0.25
Telecommunication Services, US	0.11	14.22	0.16	18.03	0.16	19.36	0.51	0.11	0.33
Utilities, US	0.54	14.49	0.64	15.76	0.58	14.85	0.44	0.14	0.28

Table 1.4: The determinants of total liquidity commonality

The table reports the estimation results of time-series regressions of monthly liquidity comovement in four regions - US, UK, Japan and Eurozone - against various demand and supply factors over the period 07/2000 - 01/2017. Liquidity commonality in one region is defined as the mean pair-wise correlation (MPC) among the ten GICS sectors' liquidity. The coefficients are taken from seemingly unrelated regression (SUR) models, estimated through an iterative algorithm and constraining the parameters to be the same for all regions. The first set of variables control for general market conditions, *US CP spread* is the spread between 90-day commercial paper rate and a 3-month T-bill, *Local bank returns* is the percentage return on the Datastream Banks Index, *Net equity flow* is the difference between "sales of foreign stocks by foreigners to US residents" and "purchases of foreign stocks by foreigners from US residents". The dependent variable in the SUR models is the total liquidity commonality implied by our factor model.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Capital market conditions														
Market return	0.0228		0.0260	0.0364	0.0434	0.0256	0.0312	0.0180	0.0408	0.0275	0.00503	0.0373	-0.0122	-0.0312
Market volatility	-0.444		-0.393	-0.740	-0.875	-0.572		-0.670	-0.868	0.233	-0.491	-0.697	-0.674	-0.407
Market turnover	1.037		0.639	0.737	0.768	1.174	0.714	5.635^a	0.773	3.961^c	1.026	-0.460	0.118	-2.513^c
Time trend	0.0443^a	0.0444^a	0.0468^a	0.0493^a	0.0463^a	0.0443^a	0.0432^a	0.0416^a	0.0455^a	0.0340^a	0.0449^a	0.0290^a	0.0392^a	0.00327
Large/small up/down returns														
$R_m^{Down,Large}$		0.0114												
R_m^{Small}		0.0717												
$R_m^{Up,Large}$		0.0431												
Supply-side factors														
Short-term interest rate			0.244											0.157
US Non-financials CP spread				5.531^a										8.966^b
US Financials CP spread					3.491^a									-3.001
Local bank returns						-0.0412								-0.0194
VIX							-0.0446							0.00775
FED stress indicator								-0.158						-2.095^b
TED spread									2.781^a					2.869
Demand-side factors														
Net equity flow										0.0499				0.0624
Exchange rate (vs SDR)											0.116			0.126
ETF volume												1.940^a		1.968^a
US sentiment index													-2.955^a	-3.097^a
TN	796	796	796	780	796	796	796	796	796	576	792	792	732	692
Average R^2	0.186	0.122	0.195	0.214	0.200	0.197	0.190	0.196	0.200	0.178	0.192	0.188	0.269	0.278

Table 1.5: The determinants of volatility-driven liquidity commonality

This table reports the estimation results of time-series regressions of monthly liquidity comovement in four regions - US, UK, Japan and Eurozone - against various demand and supply factors over the period 07/2000 - 01/2017. Liquidity commonality in one region is defined as the mean pair-wise correlation (MPC) among the ten GICS sectors' liquidity. The coefficients are taken from seemingly unrelated regression (SUR) models, estimated through an iterative algorithm and constraining the parameters to be the same for all regions. The first set of variables control for general market conditions, *US CP spread* is the spread between 90-day commercial paper rate and a 3-month T-bill, *Local bank returns* is the percentage return on the Datastream Banks Index, *Net equity flow* is the difference between "sales of foreign stocks by foreigners to US residents" and "purchases of foreign stocks by foreigners from US residents". The dependent variable in the SUR models is the volatility-driven liquidity commonality implied by our factor model, holding the factor loadings constant at their long-run average.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Capital market conditions														
Market return	0.0540		0.0707	0.0592	0.0639	0.0540	0.0349	0.0221	0.0634	0.0144	0.0541	0.0547	0.0491	-0.00655
Market volatility	0.789^{c}		1.042^{b}	0.629	0.599	0.782^{c}		0.227	0.592	-0.344	0.777^{c}	0.748^{c}	0.685	1.661^{a}
Market turnover	-9.853^{a}		-10.75^{a}	-9.841^{a}	-9.959^{a}	-9.845^{a}	-9.462^{a}	-1.151	-9.947^{a}	2.778	-9.835^{a}	-10.75^{a}	-10.47^{a}	-13.43^{a}
Time trend	0.0655^{a}	0.0629^{a}	0.0681^{a}	0.0660^{a}	0.0640^{a}	0.0655^{a}	0.0664^{a}	0.0594^{a}	0.0636^{a}	0.0478^{a}	0.0660^{a}	0.0586^{a}	0.0678^{a}	0.0326^{a}
Large/small up/down returns														
$R_m^{Down,Large}$		0.0311												
R_m^{Small}		0.0163												
$R_m^{Up,Large}$		0.0262												
Supply-side factors														
Short-term interest rate			0.680^{a}											0.731^{a}
US Non-financials CP spread				4.607^{a}										12.10^{a}
US Financials CP spread					2.508^{b}									-6.152^{c}
Local bank returns						-0.00191								0.0227
VIX							-0.0140							0.0109
FED stress indicator								-0.405						-2.726^{a}
TED spread									2.125^{b}					4.266^{c}
Demand-side factors														
Net equity flow										0.0784^{b}				0.0769^{b}
Exchange rate (vs SDR)											0.000653			0.0209
ETF volume												0.649^{c}		0.592
US sentiment index													-2.328^{a}	-3.553^{a}
TN	796	796	796	780	796	796	796	796	796	576	792	792	732	692
Average R^2	0.197	0.153	0.216	0.246	0.229	0.207	0.214	0.213	0.229	0.221	0.201	0.202	0.260	0.331

Table 1.6: The determinants of exposure-driven liquidity commonality

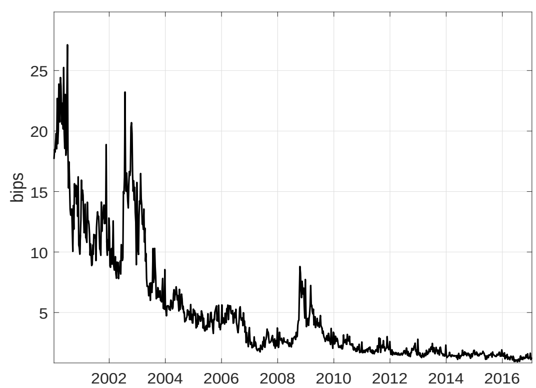
The table reports the estimation results of time-series regressions of monthly liquidity comovement in four regions - US, UK, Japan and Eurozone - against various demand and supply factors over the period 07/2000 - 01/2017. Liquidity commonality in one region is defined as the mean pair-wise correlation (MPC) among the ten GICS sectors' liquidity. The coefficients are taken from seemingly unrelated regression (SUR) models, estimated through an iterative algorithm and constraining the parameters to be the same for all regions. The first set of variables control for general market conditions, *US CP spread* is the spread between 90-day commercial paper rate and a 3-month T-bill, *Local bank returns* is the percentage return on the Datastream Banks Index, *Net equity flow* is the difference between "sales of foreign stocks by foreigners to US residents" and "purchases of foreign stocks by foreigners from US residents". The dependent variable in the SUR models is the exposure-driven liquidity commonality implied by our factor model, holding the factors' covariance matrix at its long-run average.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Capital market conditions														
Market return	-0.00932		-0.0207	-0.0111	-0.0135	-0.00962	0.0293^{c}	-0.00536	-0.0122	0.0217	-0.0117	-0.00854	-0.0120	0.0125
Market volatility	-0.814^{a}		-0.999^{a}	-0.676^{a}	-0.654^{a}	-0.785^{a}		-0.356	-0.633^{a}	0.547^{c}	-0.838^{a}	-0.866^{a}	-0.936^{a}	-0.528^{b}
Market turnover	2.145^{a}		3.220^{a}	2.901^{a}	3.077^{a}	2.128^{a}	1.601^{a}	2.725^{a}	3.150^{a}	-2.190^{b}	2.117^{a}	1.854^{a}	2.889^{a}	4.934^{a}
Time trend	0.00160	0.00199	-0.00542^{b}	0.00102	0.00162	0.00161	0.000333	-0.0000533	0.00166	-0.00916^{a}	0.00187	-0.00246	0.0000418	-0.0200^{a}
Large/small up/down returns														
$R_m^{Down,Large}$		0.0561^{c}												
R_m^{Small}		0.00729												
$R_m^{Up,Large}$		-0.00553												
Supply-side factors														
Short-term interest rate			-0.374^{a}											-0.580^{a}
US Non-financials CP spread				-1.390^{a}										-0.217
US Financials CP spread					-0.990^{a}									-1.107
Local bank returns						0.00618								-0.00201
VIX							-0.0523^{a}							-0.0505^{b}
FED stress indicator								-0.288^{b}						-0.117
TED spread									-0.825^{a}					0.679
Demand-side factors														
Net equity flow										-0.0277				-0.0329
Exchange rate (vs SDR)											0.0346			0.00428
ETF volume												0.773^{a}		1.315^{a}
US sentiment index													0.0851	0.541^{a}
TN	796	796	796	780	796	796	796	796	796	576	792	792	732	692
Average R^2	0.0842	0.0364	0.145	0.119	0.128	0.0867	0.116	0.107	0.118	0.105	0.0891	0.118	0.138	0.268

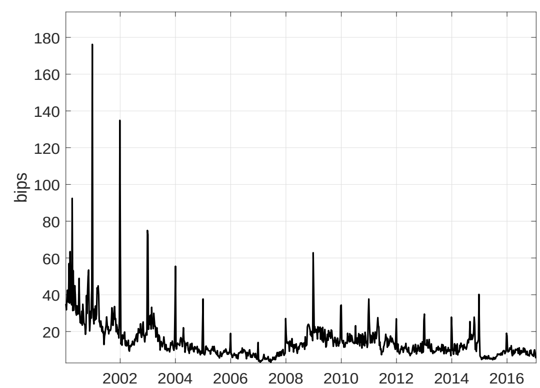
Figure 1.1: Deseasonalisation effect on the illiquidity indexes

The figure plots the illiquidity index of the Utilities sector in the US and the UK. Panels (a) and (b) refer to the raw time-series while Panels (c) and (d) report the illiquidity index calculated with the adjusted trading volume.

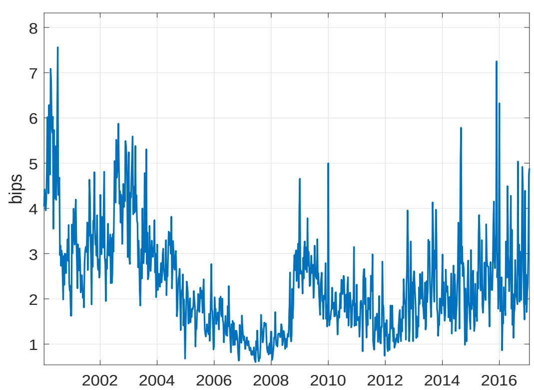
(a) US, Utilities



(b) UK, Utilities



(c) UK, Utilities



(d) UK, Utilities

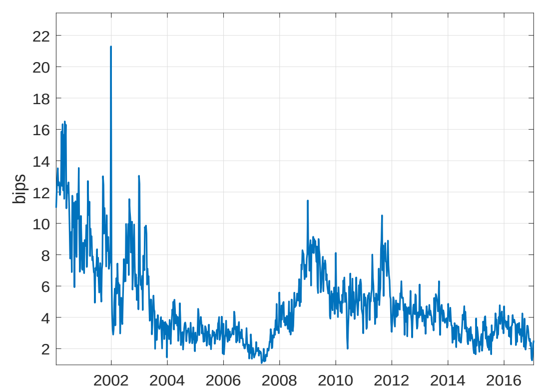


Figure 1.2: Stock illiquidity and liquidity shocks

The figure reports in Panel (a) the plots of the stock illiquidity indexes of the fourty sector-region portfolios (ten GICS sectors across US, Japan, UK and Euro zone) and in in Panel (b) the correspoding liquidity shocks. The illiquidity index is calculated using Eq. (1.8), which represents the portfolio counterpart of the Amihud (2002)'s measure, adjusted for deterministic components of trading volumes. The liquidity shocks are calculated as the inverse of the residual from an ARMA(1,1) model as defined in Eq. (1.10). All measures are standardised.

(a) Stock illiquidity **(b)** Liquidity shocks

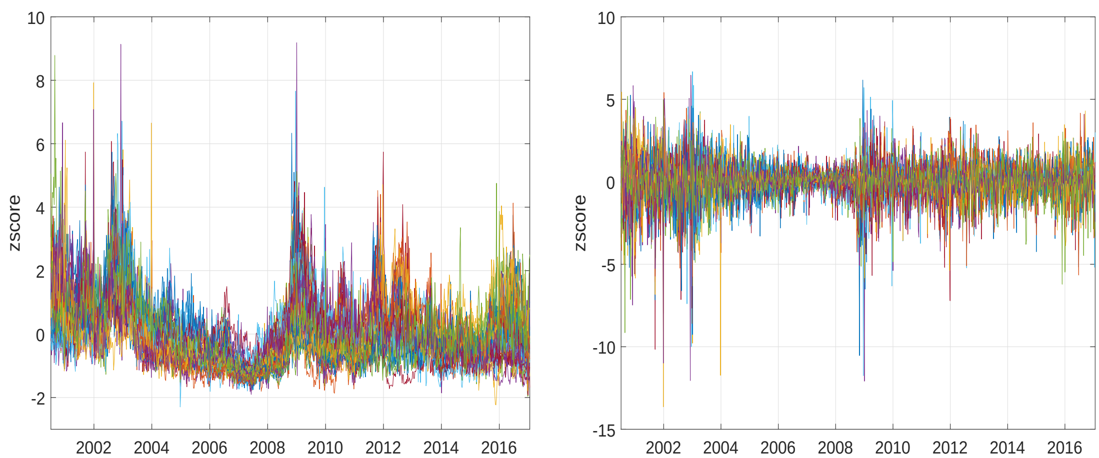


Figure 1.3: Sort-dependent liquidity correlation matrices

The figure reports, in the upper panels the correlation matrices of the dependent variable (the liquidity shocks), while the bottom panels report the same correlation matrix after taking into account the global factor. The left panels show the correlation matrices when the portfolios are ranked by sectors, whereas the right panels show the same correlations when the portfolios are ranked by region.

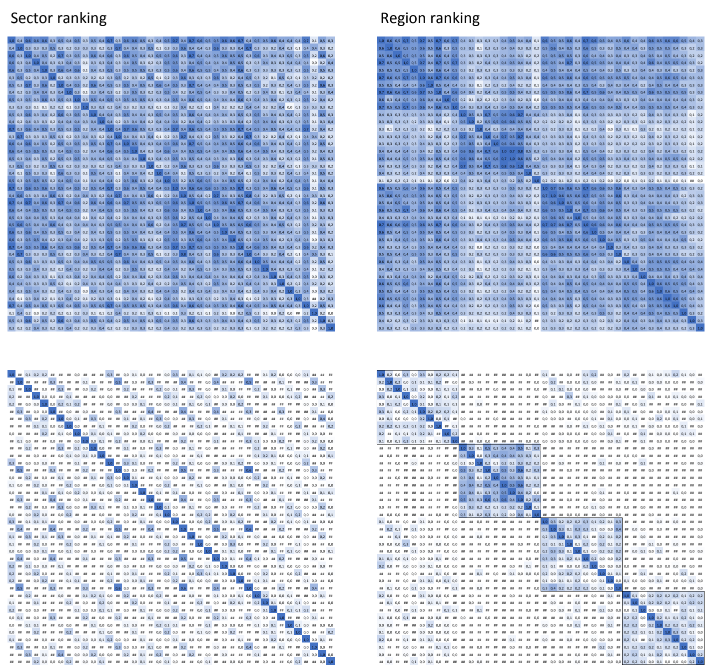
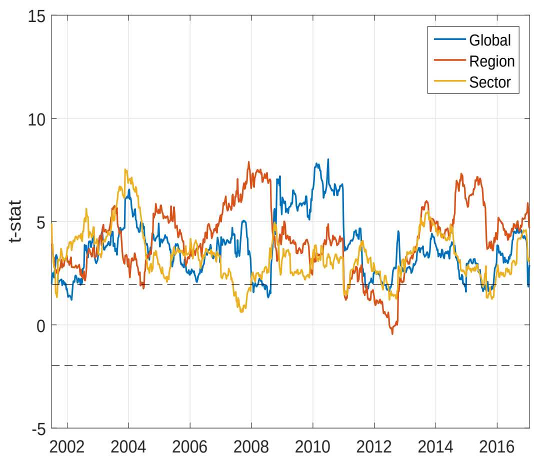


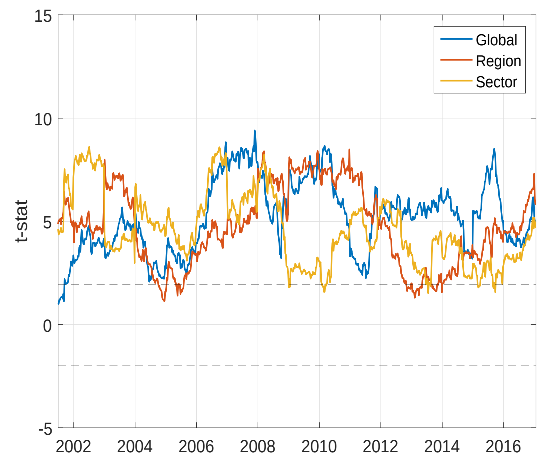
Figure 1.4: Time-varying factor significance

The figure plots the t-statistics of the factor loadings estimated with OLS from Eq. (3.1) with a one-year rolling window for Utililities and Financial sectors in the US and UK. The factors are cross-sectional averages of liquidity shocks and they have been orthogonalised.

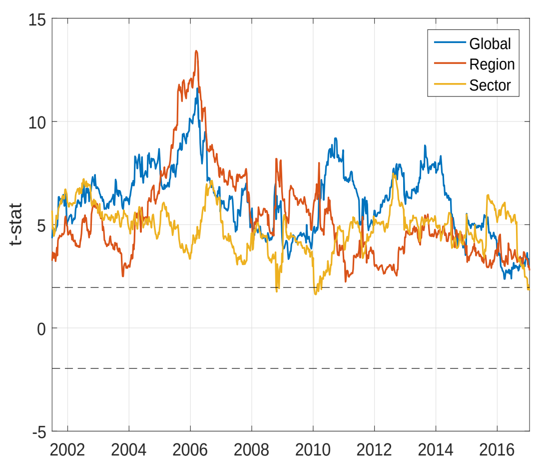
(a) Utilities, US



(b) Utilities, UK



(c) Financials, US



(d) Financials, UK

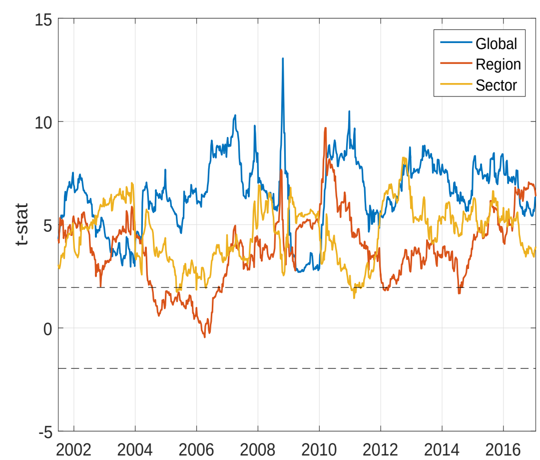
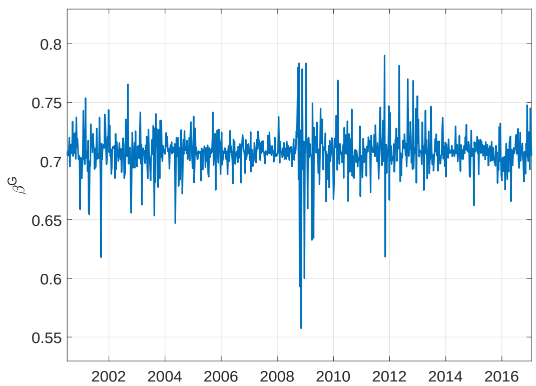


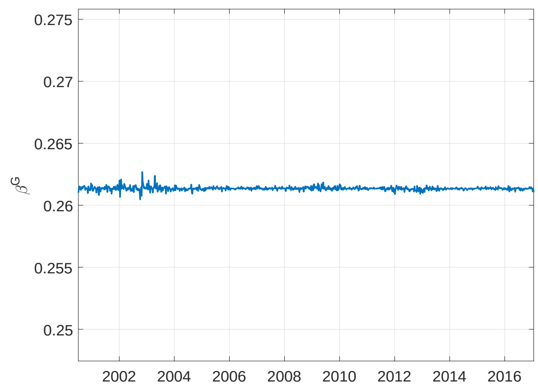
Figure 1.5: An example of factor loadings estimates

The figure reports the plots of the time-varying factor loadings for the Consumer Discretionary and Financials sectors in Euro zone and US.

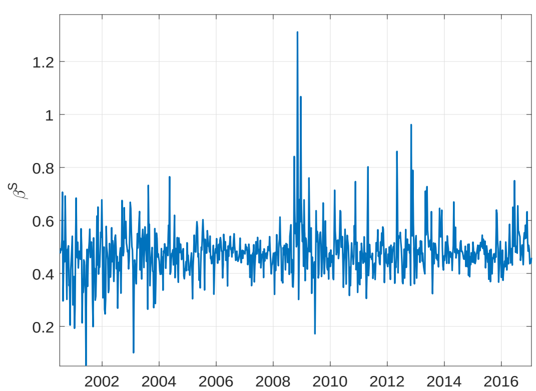
(a) Consumer Discretionary, EZ



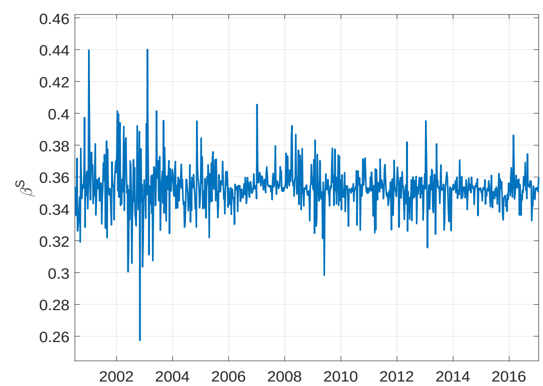
(b) Financials, US



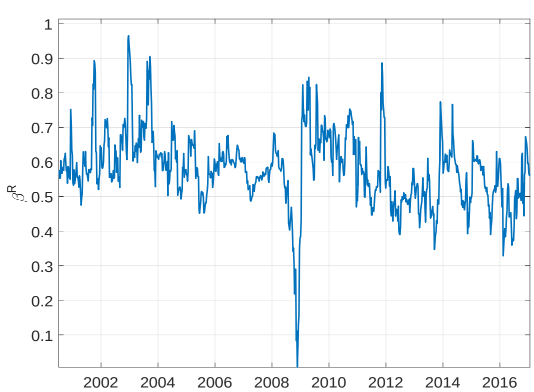
(c) Consumer Discretionary, EZ



(d) Financials, US



(e) Consumer Discretionary, EZ



(f) Financials, US

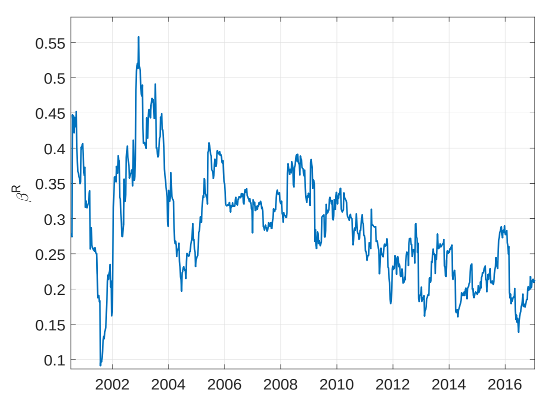
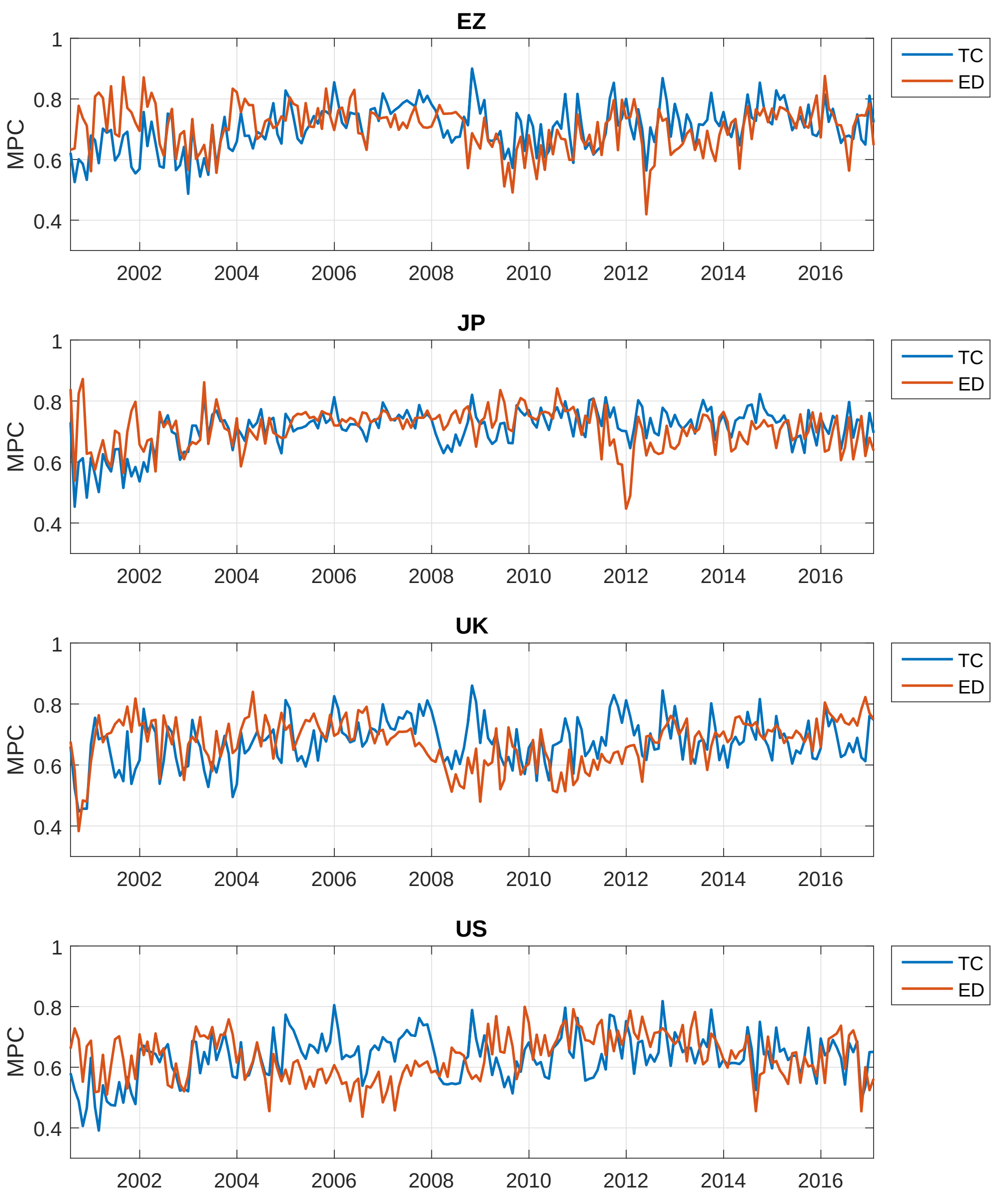


Figure 1.6: Commonality in liquidity within region

The figure reports the plots of the liquidity commonality within regions, calculated as the mean pair-wise correlation MPC among the ten GICS sectors in each region. We report the plot of both the total commonality (TC) and exposure-driven (ED) measures. For presentation purposes, ED is transformed to have the same mean and variance of TC.



Appendix 1.A Data description

Table 1.A.1: Variables description

Variable	Description	Source
Trading data		
Last price, intraday low and high	PX_LAST corresponds to the price of the last transaction reported to the exchange, adjusted for subsequent splits but not for dividends. The lowest and highest intraday prices are PX_LOW and PX_HIGH, respectively.	Bloomberg
Trading volume	Composite volume in number of shares multiplied by Last Price. This assumes that all the transactions where executed at the last price	Bloomberg and own computation
Market capitalisation	Shares outstanding multiplied by Last Price	Bloomberg and own computation
Capital market conditions		
Market return and turnover	Value-weighted average of, respectively, the return (in % per month), and the turnover (average of daily turnover in %) of all individual stocks in each region in a given month	own computation
Market volatility	Standard deviation (in %) of the daily market return of a region in a given month. Daily market returns are computed as value-weighted average of the returns of all individual stocks in each region.	own computation
Large/small up/down market returns	Large positive (negative) returns are defined as market returns that are more than one standard deviation above (below) the unconditional mean market return for each region, and zero otherwise. Small returns are defined as market returns with one standard deviation from the mean market return.	own computation
Supply-side factors		
Short-term interest rate	For US, Japan and UK we use the 3-month T-bill rate (% per annum). For the Eurozone we use a composite 3-month rate of the countries in the Eurozone, constructed by Datastream.	Datastream
US nonfinancial and financial commercial paper spread	Difference between the rate (% per annum) on the 90-day AA commercial paper and the three-month T-bill (% per annum). Both for financial and nonfinancial institutions	Federal Reverse's website
Local bank returns	Percentage return on Datastream Banks index for each region	Datastream
CBOE VIX	Measure of expected volatility. It is the square root of the annualized forward price of the 30-day variance of the S&P 500 return. This forward price is based on the replication of total variance by a portfolio of options delta-hedged with stock index futures.	Datastream
TED spread	Indicator of general credit risk. It is the difference between the rates on interbank loans (three-month LIBOR) and short-term US government debt (three-month T-bill)	Federal Reverse's website
St. Louis FED financial stress indicator	Indicator of the degree of financial stress in the markets. It is constructed from 18 weekly data series: seven interest rate series, six yield spreads and five other indicators. Average is zero. Value of zero represents normal financial market conditions. Values above zero suggest above-average financial market stress.	Federal Reverse's website
Demand-side factors		
Net equity flow	For each region, we calculate the difference of "Gross sales of foreign stocks by foreigners to US residents" and "Gross purchases of foreign stocks by foreigners from US residents" and scale it by the sum of of gross sales and purchases of foreign stocks by foreigners to/from US residents. A positive net (%) equity flow signals that US residents are net buyers of foreign stocks	Treasury International Capital (TIC)
Exchange rate	Monthly percentage change in the value of each region's currency relative to the Special Drawing Rights (SDR) basket of currencies defined by the IMF. A positive exchange rate return indicates a depreciation of the currency relative to the SDR	IMF's International Financial Statistics
ETF volume	Dollar trading volume in iShares Morgan Stanley Capital International (MSCI) ETFs traded on US markets. This is calculated as a percentage of the total stock market capitalisation of the region	Datastream
US sentiment index	The sentiment index calculated by Baker and Wurgler (2006). The index is based on the first principal component of six standardised sentiment proxies, where each of the proxies has first been orthogonalised with respect to a set of macroeconomic conditions. Positive values indicate optimistic sentiment	Jeff Wurgler's website

Appendix 1.B Deseasonalisation procedure

The adjustment is performed to $v_{s,r,t} \equiv \log(V_{s,r,t})$, estimating the following model in two stages:

$$v_{s,r,t} = b_{s,r}D_t + u_{s,r,t} \quad \text{(mean equation)} \tag{1.B.1}$$

$$\log(u^2_{s,r,t}) = \gamma_{s,r}D_t + \epsilon_{s,r,t} \quad \text{(variance equation)} \tag{1.B.2}$$

where D_t contains a constant, a linear trend and dummy variables for all the weeks of December and January to control for holiday effects. This structure implies a known form of heteroscedasticity with seasonalities. D_t can accommodate, for instance, a quadratic trend and week-of-the-year dummies. However, we prefer a simpler structure because the adjustment with a richer specification is economically equal. Then, the estimated variance from (1.B.2), $\exp(\hat{\gamma}_{s,r}D_t)$, is used to standardise the residuals from the mean equation as follows:

$$\hat{z}_{s,r,t} = \hat{u}_{s,r,t} \exp(-\frac{1}{2}\hat{\gamma}_{s,r}D_t). \tag{1.B.3}$$

In line with Gallant et al. (1992), we use only the variables whose coefficients are statistically different from zero to construct the residuals. Next, we build a linear combination so that the sample mean and variance of the adjusted log-volume, $av_{s,r,t}$, match the ones of $v_{s,r,t}$:

$$av_{s,r,t} = E(v_{s,r_t}) + \frac{\sqrt{var(v_{s,r,t})}}{\sqrt{var(\hat{z}_{s,r,t})}}\hat{z}_{s,r,t}. \tag{1.B.4}$$

Thus, the adjusted volume index to use in formulation (1.8) can be calculated as $AV_{s,r,t} = \exp(av_{s,r,t})$.

Appendix 1.C Kalman Filter

Table 1.C.1 reports the values that we use to initialise the iterative procedure of the Kalman Filter. The results are economically independent of the initial values. We use OLS estimates to improve the speed of the algorithm.

Table 1.C.1: Initialisation Kalman Filter

The table reports the initial conditions of the parameters that define the linear-gaussian Kalman Filter that we employ to estimate the parameters of our model via Maximum Likelihood. The model for the liquidity of sector-region n at time t, L_{nt} is:

$$L_{n,t} = \beta^G_{n,t} G_t + \beta^R_{n,t} R_{n,t} + \beta^S_{n,t} S_{n,t} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, (\sigma^\varepsilon_n)^2)$$

$$\begin{aligned}
\beta^G_{n,t} &= (1-\phi^G_n)\bar{\beta}^G_n + \phi^G_n \beta^G_{n,t-1} + u^G_{n,t}, \quad u^G_{n,t} \sim iidN(0, (\sigma^{u,G}_n)^2) \\
\beta^R_{n,t} &= (1-\phi^R_n)\bar{\beta}^R_n + \phi^R_n \beta^R_{n,t-1} + u^R_{n,t}, \quad u^R_{n,t} \sim iidN(0, (\sigma^{u,R}_n)^2) \\
\beta^S_{n,t} &= (1-\phi^S_n)\bar{\beta}^S_n + \phi^S_n \beta^S_{n,t-1} + u^S_{n,t}, \quad u^S_{n,t} \sim iidN(0, (\sigma^{u,S}_n)^2)
\end{aligned}$$

In the table, $\widehat{\beta}_{ols}$ indicates the OLS estimate of the loading on the relevant factor, s^2 is the OLS estimate of the regression variance. We maximise the likelihood function using a constrained optimisation and we report the lower and upper bounds.

	$t=0$	Lower bound	Upper bound
$\bar{\beta}^G$	$\widehat{\beta}^G_{ols}$	-5	5
$\bar{\beta}^S$	$\widehat{\beta}^S_{ols}$	-5	5
$\bar{\beta}^R$	$\widehat{\beta}^R_{ols}$	-5	5
ϕ^G	0.5	-0.99	0.99
ϕ^S	0.5	-0.99	0.99
ϕ^R	0.5	-0.99	0.99
$\log(var(\beta^G))$	$\log(s^2)$	$100\log(s^2)$	0
$\log(var(\beta^S))$	$\log(s^2)$	$100\log(s^2)$	0
$\log(var(\beta^R))$	$\log(s^2)$	$100\log(s^2)$	0
$\log((\sigma^\varepsilon)^2)$	$\log(s^2)$	$100\log(s^2)$	0

Table 1.C.2 reports the estimation results of the unknown parameters of Eqs. (1.23) - (1.26), estimated with maximum likelihood through the Kalman Filter. The estimates that we report are filtered, i.e. in-sample predicted values without smoothing.

Table 1.C.2: Parameters estimates

The table reports the parameters estimated with Maximum Likelihood through the Kalman Filter. Estimates are filtered, i.e. in-sample predicted values (no smoothing). The factors are calculated as cross-sectional averages of standardised illiquidity shocks. By construction they have zero mean but not necessarily unit variance. The dependent variable is illiquidity innovations in each sector-region portfolio. It has a mean of zero but it is not standardised. The long-run betas are multiplied by 100. The model for the liquidity of sector-region n at time t, L_{nt} is:

$$L_{n,t} = \beta^G_{n,t} G_t + \beta^R_{n,t} R_{n,t} + \beta^S_{n,t} S_{n,t} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, (\sigma^{\varepsilon}_n)^2)$$

$$\begin{aligned}
\beta^G_{n,t} &= (1-\phi^G_n)\bar{\beta}^G_n + \phi^G_n \beta^G_{n,t-1} + u^G_{n,t}, \quad u^G_{n,t} \sim iidN(0, (\sigma^{u,G}_n)^2) \\
\beta^R_{n,t} &= (1-\phi^R_n)\bar{\beta}^R_n + \phi^R_n \beta^R_{n,t-1} + u^R_{n,t}, \quad u^R_{n,t} \sim iidN(0, (\sigma^{u,R}_n)^2) \\
\beta^S_{n,t} &= (1-\phi^S_n)\bar{\beta}^S_n + \phi^S_n \beta^S_{n,t-1} + u^S_{n,t}, \quad u^S_{n,t} \sim iidN(0, (\sigma^{u,S}_n)^2)
\end{aligned}$$

	$\bar{\beta}^G$	ϕ^G	$\bar{\beta}^S$	ϕ^S	$\bar{\beta}^R$	ϕ^R
Consumer Discretionary, EZ	0.7078	-0.1095	0.4780	0.2860	0.5760	0.8979
Consumer Staples, EZ	0.2883	0.2976	0.2978	-0.0337	0.3735	-0.6191
Energy, EZ	2.1738	0.9841	1.7868	0.9412	1.9500	0.0487
Financials, EZ	0.3079	0.3132	0.3676	0.1600	0.3218	0.1653
Health Care, EZ	0.5495	0.3366	0.5364	0.6903	0.4406	0.8860
Industrials, EZ	0.8895	-0.2048	1.0263	-0.1187	1.2484	0.7823
Information Technology, EZ	0.5586	0.1696	0.4612	0.0659	0.5295	0.2126
Materials, EZ	0.0918	0.9706	0.0828	-0.3185	0.1001	0.4078
Telecommunication Services, EZ	0.2569	-0.2237	0.2233	0.9488	0.2092	0.4212
Utilities, EZ	2.1268	0.2541	3.8735	0.6525	2.7116	0.1091
Consumer Discretionary, JP	0.3996	0.2796	0.3178	0.2634	0.3839	0.1839
Consumer Staples, JP	0.0701	0.6593	0.0824	0.3975	0.0794	0.1946
Energy, JP	0.4251	-0.3006	0.3876	0.1003	0.3992	0.5061
Financials, JP	0.5014	0.7412	0.6312	0.3481	0.6190	-0.2442
Health Care, JP	0.5882	0.0855	0.4676	0.3933	0.5455	0.0722
Industrials, JP	0.1171	0.2780	0.1180	0.8897	0.1098	0.3017
Information Technology, JP	0.5142	0.9900	0.5212	0.1568	0.5248	-0.0612
Materials, JP	0.5081	0.0092	0.7341	0.3324	0.7818	0.3007
Telecommunication Services, JP	0.3389	-0.2097	0.3113	0.9233	0.3949	0.3983
Utilities, JP	0.1047	0.1982	0.0898	0.4153	0.0966	-0.1734
Consumer Discretionary, UK	0.8054	0.2681	0.5031	0.1935	0.7031	-0.0150
Consumer Staples, UK	1.1753	0.9900	0.9453	0.0767	1.4104	0.6518
Energy, UK	2.3072	0.2289	1.8975	0.2942	2.0224	0.1134
Financials, UK	0.0998	0.0331	0.1155	0.5944	0.0890	0.5436
Health Care, UK	0.6599	0.1682	0.6044	0.7842	0.4293	0.2319
Industrials, UK	0.6084	0.1087	0.5088	0.7826	0.6613	0.6218
Information Technology, UK	4.2074	-0.8512	5.2794	0.2538	4.6002	0.4150
Materials, UK	0.0488	0.7727	0.0568	0.9235	0.0388	-0.5106
Telecommunication Services, UK	1.0943	-0.0859	0.6624	0.1878	0.7792	-0.0678
Utilities, UK	1.4515	0.7125	1.4049	0.1232	1.8344	0.6058
Consumer Discretionary, US	1.0206	0.1579	0.9869	0.2381	0.7772	0.9900
Consumer Staples, US	0.3332	-0.3713	0.4262	0.4059	0.3477	0.1127
Energy, US	0.2822	0.1672	0.2711	0.4786	0.2902	0.3852
Financials, US	0.2614	0.0011	0.3534	0.0880	0.2784	0.9767
Health Care, US	0.5304	0.6104	0.5266	0.0151	0.5171	0.0013
Industrials, US	0.1052	-0.9037	0.1531	0.3420	0.1426	0.5734
Information Technology, US	0.5059	0.6278	0.4005	-0.0086	0.4339	-0.1787
Materials, US	0.8354	0.2507	1.2266	0.6061	1.1787	-0.1081
Telecommunication Services, US	0.9218	0.4239	1.1338	0.2372	1.2102	0.0936
Utilities, US	0.5370	0.7222	0.6167	0.5121	0.4862	0.8206

Appendix 1.D Robustness checks

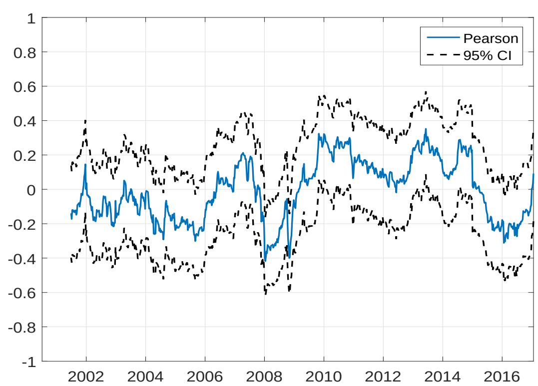
1.D.1 Conditional orthogonalisation

In Section 1.3, we performed a sequential orthogonalisation of the factors to ensure that, for each liquidity portfolio n, the triplet $[\hat{G}_t, \hat{R}_{n,t}, \hat{S}_{n,t}]$ is a set of mutually uncorrelated factors. As robustness check, we estimate the 52-week rolling Pearson correlation between $\hat{G}_t$ and $\hat{R}_{r,t}$, for $r = 1, \ldots, N_R$. Fig. 1.D.1 plots the resulting conditional correlation together with its 95% confidence interval, showing that for most of the sample, the conditional Pearson coefficient is not significantly different from zero.

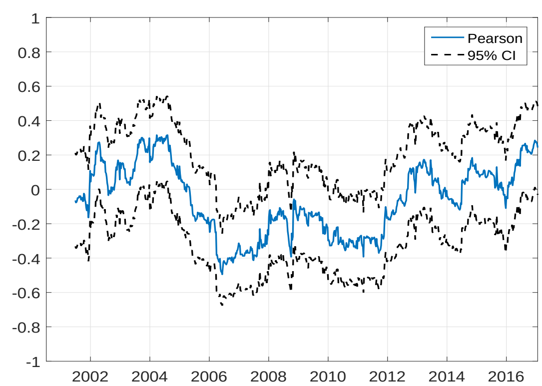
Figure 1.D.1: 52-week rolling correlation

The figure reports the plots of the time-varying correlations between the global factor and the four region factors estimated using 52-week rolling Pearson correlation.

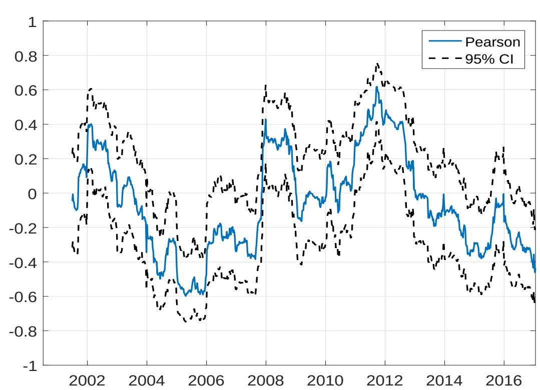
(a) Global factor and EZ



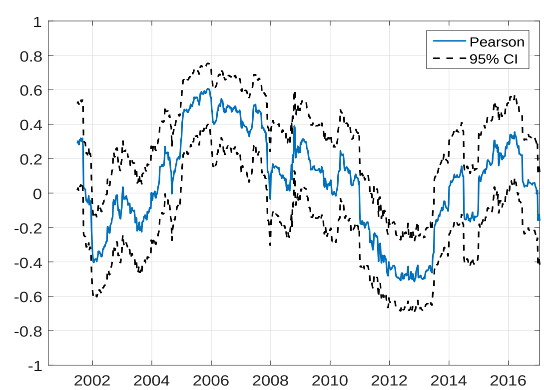
(b) Global factor and JP



(c) Global factor and UK



(d) Global factor and US



1.D.2 Heterogeneity

In this section we analyse if the relationship between economic variables and liquidity commonality holds in all regions uniformly. Tables 1.D.1 reports the estimation results of the unrestricted SUR model, i.e. without imposing the homogeneity restrictions on the parameters.

The two main demand-side drivers of liquidity commonality are ETF volume and US sentiment. ETF volume is positively related to liquidity commonality, but with a significant coefficient for Eurozone and Japan only. Thus, when there is an increase in index-related trading, the price impact of all Eurozone's and Japan's stocks increases. On the other hand, the US sentiment index of Baker and Wurgler (2006) is positively related to liquidity commonality in the Eurozone and UK, while negatively related to liquidity commonality in US and Japan. In the US and Japan, when people become more optimistic, liquidity commonality is less correlated across sectors, supporting our interpretation above.

The results for the supply-side factors are more ambiguous but in line with the evidence reported in Tables 1.4 - 1.6. Variables that approximate funding constraints and credit conditions - local short-term interest rate, US commercial paper spread, TED spread - are positively related to liquidity commonality. This is in line with our results in the restricted SUR model and with Karolyi et al. (2012) and Brunnermeier and Pedersen (2009). Japan is the only region where the supply-side variables do not play a role.

Table 1.D.1: Unrestricted SUR, total commonality

The table reports the estimation results of time-series regressions of monthly liquidity comovement in four regions - US, UK, Japan and Eurozone - against various demand and supply factors over the period 07/2000 - 01/2017. Liquidity commonality in one region is defined as the mean pair-wise correlation (MPC) among the 10 GICS sectors' liquidity. The coefficients are taken from seemingly unrelated regression (SUR) models, estimated through an iterative algorithm. The dependent variables in the SUR models are the **total** MPC, estimated with our factor model using the innovations of liquidity, calculated as the residuals from an ARMA model estimated in-sample.

EZ									
Market return	0.0160	0.00701	0.0661	0.0207	0.0257	0.00346	-0.00805	0.00185	0.0603
Market volatility	-0.787	-0.994	-0.381	-1.073	-1.640^b	-1.180	-1.143	-1.347^c	-0.711
Market turnover	2.560	7.586^b	-0.312	1.159	1.545	8.845^b	7.136^b	10.93^a	10.87^a
Time trend		0.0496^a	0.0941^a	0.0541^a	0.0475^a	0.0452^a	0.0489^a	0.0317^a	0.0702^a
Short-term interest rate			1.966^a						
US Non-financials CP spread				10.55^a					
TED spread					5.539^a				
Net equity flow						-0.000897			
Exchange rate (vs SDR)							0.257		
ETF volume								5.142^a	
US sentiment index									1.608^b
JP									
Market return	-0.0487	-0.0502	-0.0715	-0.0624	-0.0636	-0.0595	-0.0341	-0.0258	-0.0996
Market volatility	-2.702^a	-2.299^a	-2.353^a	-2.056^a	-2.042^a	-2.232^a	-2.543^a	-2.208^a	-2.067^a
Market turnover	16.96^a	13.75^a	16.15^a	14.46^a	15.83^a	14.03^a	13.87^a	10.23^a	8.380^b
Time trend		0.0325^a	0.0276^a	0.0325^a	0.0299^a	0.0296^a	0.0329^a	0.0250^a	0.0271^a
Short-term interest rate			-3.180						
US Non-financials CP spread				0.0701					
TED spread					-0.563				
Net equity flow						0.0811			
Exchange rate (vs SDR)							-0.205		
ETF volume								1.156^b	
US sentiment index									-3.519^a
UK									
Market return	0.00729	-0.00362	0.0276	0.00619	0.0141	-0.0334	-0.0365	-0.0228	0.0529
Market volatility	0.597	0.743	1.289	0.358	-0.227	0.713	0.756	0.461	1.214
Market turnover	1.089	1.081	-3.812	-1.441	-1.292	0.0129	0.753	3.499	3.992
Time trend		0.0332^a	0.0515^a	0.0360^a	0.0304^a	0.0198^c	0.0321^a	0.0271^b	0.0553^a
Short-term interest rate			0.834^b						
US Non-financials CP spread				8.562^a					
TED spread					4.898^a				
Net equity flow						0.0142			
Exchange rate (vs SDR)							-0.0172		
ETF volume								4.737	
US sentiment index									1.515^c
US									
Market return	0.0142	0.00475	-0.0127	0.0110	0.0224		-0.0339	-0.00583	-0.0692
Market volatility	-1.046	-1.164	-1.117	-0.872	-1.452		-1.152	-1.043	-0.778
Market turnover	0.753	1.483	0.747	1.006	1.232		1.420	1.255	-1.596
Time trend		0.0320^a	0.0246^b	0.0377^a	0.0328^a		0.0336^a	0.0355^a	0.0316^a
Short-term interest rate			-0.397						
US Non-financials CP spread				4.343^c					
TED spread					2.781^c				
Exchange rate (vs SDR)							0.347		
ETF volume								-0.746	
US sentiment index									-3.004^a

Chapter 2

High-frequency quoting and liquidity commonality★

2.1 Introduction

The last years have seen a large increase in high-frequency quoting (HFQ): An activity carried out by high-frequency traders (HFT), firms that use advanced technology and proprietary capital to act as market makers on multiple securities [Menkveld (2013)]. Compared to traditional market makers, such as banks, the HFT's business model is less diversified and more exposed to funding constraints. In the limit, with only one HFT supplying liquidity, a reduction in its capital, driven by a loss in one stock, might make the HFT less willing to put capital at risk and thus less willing to supply liquidity for all stocks. Therefore, we would observe strong interdependence between liquidity across stocks. This has various implications for market quality. First, concentration of liquidity provision increases systemic risk [Linton et al. (2012)] and consequently regulators' concerns[1]. Second, it increases the probability of illiquidity contagion across the

★Preliminary versions of this chapter have been presented at the "2017 PhD Research Days" (Cass Business School, 7^{th} June 2017) and the "2017 CFE-CFStatistics Conference" (Birkbeck, University of London, 17^{th} December 2017)

[1]The Senior Supervisors Group in April 2015 stated that "the extent to which algorithmic trading activity, including HFT, is adequately captured in banks' risk management frameworks, and whether standard risk management tools are effective for monitoring the risks associated

stocks covered by the HFT [Cespa and Foucault (2014)]. Third, execution costs of baskets of stocks increase if all stocks become expensive to trade at the same time, unless high liquidity periods perfectly offset low liquidity ones. This paper investigates whether and to what extent these new liquidity providers affect the interconnectedness of equity markets, captured by high-frequency liquidity commonality.

There is a strong debate on the impact of HFT on markets' quality, both in the academia and financial industry[2]. While HFT supporters find an increase in liquidity, lower volatility, and a positive contribution to price discovery [Hendershott et al. (2011), Hasbrouck and Saar (2013), Brogaard et al. (2014)], other authors document negative externalities for human trades and unchanged execution costs of large orders [Muravyev and Pearson (2013), Kim and Murphy (2013), Stiglitz (2014)][3]. However, the effect of HFT on liquidity commonality is still unexplored. A few papers have proposed explanations for high-frequency liquidity commonality that can be indirectly related to HFT activity. Coughenour and Saad (2004) argue that liquidity co-moves because each NYSE specialist firm provides liquidity for more than one common stock, sharing capital and information. Domowitz et al. (2005) show that liquidity commonality is due to co-movements in supply and demand induced by cross-sectional correlation in order types (market and limit orders).

This paper proposes high-frequency quoting as a new supply-side explanation for liquidity commonality. We examine two research questions. First, we ask whether a long-run increase in HFQ is related to higher liquidity commonality, where stock liquidity is measured by bid-offer (BO) spread, effective spread and order book depth. Second, we investigate if HFQ has an impact on liquidity commonality also during the trading day and we connect their relationship to the level of stock volatility, liquidity, order flow and order imbalance. To test our hypotheses, we use tick-by-tick data for the FTSE100 stocks listed on the London Stock Exchange (LSE), from September 2010 to July 2011. We estimate stock-level liquidity commonality as the explanatory power of common liquid-

with this activity, are areas of inquiry that all supervisors need to explore". Source: https://www.newyorkfed.org/newsevents/news/banking/2015/an150430.html.

[2]High-frequency trading refers to all strategies that are implemented using low-latency infrastructure. High-frequency quoting is the most used.

[3]See Menkveld (2016) for an excellent survey.

ity factors, estimated by principal component analysis (PCA) as in Hasbrouck and Seppi (2001), on individual stock liquidity.

Our analysis yields three main contributions. Our first contribution is to shed light on the impact of microstructure noise and long memory on the structure of a factor model for stock liquidity. We introduce signature plots for eigenvalues and number of factors estimated using the Bai and Ng (2002) estimator. So far, the literature have used signature plots to decide at which frequency to sample prices before calculating realised volatility [Andersen et al. (2003)]. For this part of the analysis, we approximate liquidity by the depth of the order book at the best prices.

We find that a three factor model can be used for data sampled at frequencies from eight to 15 minutes. At the highest frequencies, when the signal to noise ratio is very low, we need a higher number of factors to explain a fixed amount of variability and this can be partly explained by the large autocorrelation of liquidity at very high frequencies. To strike a balance between signal and noise, we use a three-factor model with data sampled at 10-minute frequency.

Using these results, we can answer our first research question, which forms our second contribution. We test if an increase in high-frequency quoting leads to higher commonality in liquidity, representing an increase in market interconnectedness, in the UK equity markets. Thus, we propose a supply-side driver of high-frequency liquidity commonality. To test this hypothesis, we identify an exogenous variation in HFQ using the LSE technological upgrade of 14 February 2011, when the Millennium Exchange trading system was introduced. The computer upgrade exogenously increased HFT participation without directly increasing commonality. Similarly, Brogaard et al. (2014) use the LSE technological upgrade of TradElect and Hendershott et al. (2011) use the introduction of Autoquote by the NYSE in 2003 to identify HFT. First, we test the difference in eigenvalues five months before (September 2010 - January 2011) and after (March 2011 - June 2011) the shock. Second, we perform a panel data analysis to formally test an increase in the liquidity commonality, controlling for market conditions.

We find that liquidity commonality increased after the introduction of the Millennium Exchange and this result is robust to different measures of liquidity. For depth, the variability explained by the first three principal components increased from 13% to 15%. Bid-offer spread commonality increased from 21.8% to 28.24%, while commonality in effective spreads increased from 12.26% to 17.26%. Furthermore, the results on the panel data analysis shows that, even after controlling for stock volatility, the long-run average of liquidity commonality increased from 8.28% to 15.15% after the Millennium Exchange, suggesting that HFQ increases liquidity commonality. This result strongly supports our hypothesis that high-frequency traders coordinates their quoting activity across multiple securities. There are two possible explanations of why HFTs act in this way. The first is that HFTs acquire information on one security from other securities. They could look at correlated equities or to order books of the same stock in other exchanges. This argument is consistent with the model of Cespa and Foucault (2014), where liquidity providers in one asset class learn information from other asset prices and this leads to liquidity spillovers. The second explanation is funding liquidity and lack of capital, as in the models by Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2009). These mechanisms, the cross-asset learning and the funding constraints channel, are not mutually exclusive. However, we think that at high frequency the commonality in liquidity can be explained by the information channel since it is difficult to think of a situation when funding costs would suddenly rise, either intraday or overnight.

Our third contribution is the analysis of the intraday relationship between HFT's liquidity provision activity, liquidity commonality, returns volatility, liquidity, order flow and order imbalance. We use simple graphical analysis to understand the behaviour of HFTs in relation to the other microstructure variables. We both analyse overall results and by groups of large and small stocks, ranked by market capitalisation.

First, we find that, at the market opening, there is a systematic illiquidity that affects all stocks, as traders are processing overnight market-wide information. On average, HFQ is high in the morning and it increases until 12pm, which is justified both by the uncertainty of prices that require fast quote rebalancing and by the large profit available for the HFT, in terms of bid-offer spread. Thus, HFTs are active when all stocks are illiquid and market makers are needed the most. Second, liquidity commonality is always higher for large stocks, suggesting the presence of stronger underlying common factors. When measured using the order-to-trade ratio, to control for the size of trades, also HFQ is the highest for large stocks. Third, we find that HTFs focus on larger stocks in the last hours of trading, to take advantage of the large trading volumes. Finally, we find evidence that suggests that liquidity providers absorb demand shocks more efficiently at the end of the day, using marketable orders. These need less rebalancing than limit orders and allow to quickly build inventory at the end of the day. Thus, liquidity is provided in a different way throughout the day.

The remainder of the paper is organised as follows. Section 3.3 describes the institutional setting and the data. Section 2.3 presents our measure of liquidity commonality and the signature plots. Section 2.4 tests whether HFQ increases liquidity commonality. Section 2.5 analyses the relationship between HFQ and other microstructure variables. Section 2.6 concludes.

2.2 Institutional setting and data

European equity markets are fragmented and trading can take place on multiple venues, with or without pre-trade transparency. Transactions can be executed on "Lit" venues - traditional exchanges and Multilateral Trading Facilities (MTFs) - that combine visible and hidden liquidity, OTC, in Broker Crossing Network, in dark pools and by systematic internalisers (SIs).

We focus on the period between January 2010 and December 2011, when lit trading was distributed across LSE, Chi-X, BATS and Turquoise. Our data base comprises all trades and best quotes updates (i.e. any update of the best bid and ask quotes) executed or posted on the LSE, for the stocks that entered the FTSE100 Index. For our analysis we select stocks with at least 100 trades per day[4] and that did not trade with multiple classes (e.g. RBSa and RBSb). In total, we have 91 stocks. See Appendix 2.A for details on data cleaning procedures. During the sample period analysed, the LSE has acted as lone primary exchange[5], capturing around 50% of trading volume, and in this paper we focus on the activity occurring on this market. Refer to Riordan et al. (2011) for more

[4] Riordan et al. (2011) exclude stocks with any one day with less than ten trades, but after aggregating the transactions recorded in the same millisecond, while we do it before aggregating them.

[5] Chi-X acquired BATS Europe on December 2011, and the new entity BATS Chi-X Europe became a Recognised Investment Exchange in May 2013.

details on European markets fragmentation. The order book is reconstructed by Thompson Reuters Tick History (TRTH)[6]. Events are timestamped when they are received by Thompson Reuters data centre, hence they do not match with the exchange timestamps. Since both Thompson Reuters and LSE are in London, and order book data is provided at the microsecond quality, this is not an issue for our analysis as we expect timestamping errors to be sub-millisecond. Transactions executed by systematic internalisers (SI) and OTC are not included in our data base. Finally, market capitalisation data come from Bloomberg.

The LSE operates various order books (i.e. SETS, SETSQX, SEAQ, ETR, EQS), each dedicated to different types of securities. The shares of FTSE100 companies trade on the Stock exchange Electronic Trading Service (SETS), which is a fully electronic order-driven book with additional liquidity provided by registered market makers. Thus, an investor's order is electronically transmitted to the SETS by a broker and automatically matched to an existing order. For illiquid stocks or periods, the other side might be provided by a market maker. This combination of multiple liquidity sources reduces the number of shares needed to be held by the market maker, hence their market risk. In addition to visible order types, hidden and iceberg limit orders (for orders that meet the Large-in-scale MiFID requirements) are available on LSE since December 2009, thus offering pre-trade transparency on certain large transactions.

In terms of exchange fees, the LSE has to compete with attractive maker-taker models employed by entrant markets like Chi-X. These schemes are appealing to fast traders, who can provide liquidity where they receive a rebate, while consuming liquidity (to build inventory) where it is cheaper to do so. The LSE adopted, since May 4 2010, a pricing schedule similar to a maker-taker model. By default, both passive and aggressive orders are charged between 0.20 bps and 0.45 bps. However, high-volume traders that meet certain requirements can apply for a more convenient schedule: a flat rate of 0.29 bps on aggressive rates and a full rate waiver for executed passive orders[7]. In our sample period

[6]We use the *Time & Sales* function and for each stock we analyse the files with extension ".L". Additionally, TRTH provides for each stock a Consolidated European Trade Tape with all trades reported in MiFID zone comprising OTC and systematic internalisers (".xt") and a European Consolidated Best Bid&Offer from regulated markets and MTFs (".xbo").

[7]The LSE abandoned the maker-taker model in September 1 2009 but, after losing a substantial market share, the Exchange restored the maker-taker model in May 3 2010.

the price schedule do not change.

Various players interact on SETS. “Member firms” are corporations or sole practitioners admitted to trading. “Agents” are member firms who act on behalf of a customer. Member firms can register as “market makers” in one or more securities if they are able to provide and maintain certain levels of liquidity. They offer executable two-side quotes and deal either on the order book, off the book or both. Because SETS is fully electronic, market makers on SETS are mostly HFTs who quickly post and adjust quotes to make demand and supply meet in the market. As of May 2017, there are 21 registered market makers on the LSE (e.g. KCG Holdings, Optiver, Virtu, Morgan Stanley, UBS). The market makers on SETS have three main quoting obligations. First, they have to maintain an executable quote in each security for at least 90% of regular trading day. The quotes must be maintained until the end of the closing auction too. Second, their quotes have a minimum size requirement of one Exchange Market Size (EMS), which for SETS securities is 1% of Average Daily Turnover (ADT), calculated over a period of *one year* and subject to an upper cap of around £25,000 and a lower cap of £2500[8]. Third, market makers' quotes have to respect the maximum spread regime, which depends on the ADT of the security. For instance, for the most liquid stocks (ADT $\geq$ €50m) the maximum spread is 100 bps and for the next liquidity group (€5m $\leq$ ADT $<$ €50m) it is 250 bps. In practice, this implies that market makers' executable quotes are automatically rejected if they do not follow these rules. Note that the LSE can occasionally relax the spread regime when a stock experiences wide price movements[9]. An important consideration is that, in our sample period, a member firm can effectively engage in high-frequency market making strategies, and be eligible for a fee waiver, without registering as market maker and escaping the obligations mentioned above.

Finally, other players that add liquidity to European equity markets are internalisers and Systematic Internalisers (SI). MiFID I/II defines SIs as “investment

[8] Note that there is no more upper cap after MiFID II. See the MIT & TE Parameters spreadsheet for more details.

[9] For more details see: https://www.londonstockexchange.com/traders-and-brokers/rules-regulations/rules-lse.pdf and the changes and updates document: http://www.londonstockexchange.com/traders-and-brokers/rules-regulations/change-and-updates/stock-exchange-notices/2010/n1810_attach1.pdf.

firms which, on an organised, frequent, systematic and substantial basis, execute client orders against proprietary capital outside a regulated market, MTF or OTF without operating a multilateral system", but only MiFID II introduced a set of quantitative rules to identify SIs. SIs are traditional sell-side banks with sufficient size and technology to fill orders internally against their own books. To classify as SI, the firm needs to undertake risk-taking transactions using the firm's own account, i.e. a simple agency cross does not represent risk-taking activity. Furthermore, SIs are obliged to meet pre-trade transparency requirements, in particular to provide public quotes in liquid instruments traded in regulated markets. The quotes are transmitted by services like TRADEcho, from the SI to the LSE market data channels. Quotes are also available on the SI's website and on Approved Publication Arrangements (APAs). SI transactions are reported separately from the regulated markets.

In our sample period there was only a MiFID I guidance on the definition of SI and most firms did not voluntarily apply to become SI. Consequently, our data contains public SI quotes plus quotes of internalisers (firms that were acting as SI without being registered) and internalised transactions, i.e. executed off-book but reported on the LSE.

2.2.1 The technological change: Millennium Exchange

The technological upgrade we use to identify HFQ is the update, on 14 February 2011, of LSE's electronic trading platform, from TradElect-5 to the Linux-based system Millennium Exchange. This update provides the identification of positive variation of HFT activity.

During our sample, the LSE competes with Chi-X, BATS and Turquoise for order flow. In addition to the attractive pricing schemes, the MTFs offer low trading infrastructure that allowed, for instance, BATS to reach an order latency of 200 microseconds in May 2010 and Turquoise to get to 126 microseconds in October 2010 [Riordan et al. (2011)]. Latency (and especially spread in latencies between venues) is a crucial factor for HFTs' profits. The LSE before February 2011 was unable to beat the competition in terms of speed because the TradElect-5 system guaranteed an order latency of only 3 milliseconds. With Millennium Exchange the latency decreased to 113 microseconds. Since the change in order latency is only a few milliseconds, the event has a direct impact on computer-based trading and HFTs, but only an indirect impact on human traders.

Given that we are interested in testing if an increase in the participation of HFTs have increased liquidity commonality, we split our sample around the Millennium Exchange upgrade, which exogenously increases HFQ without directly increasing liquidity commonality. We use two five-month periods:

1. Sep 2010 - Jan 2011 (Pre-shock)
2. Mar 2011 - Jul 2011 (Post-shock)

So far, the literature has dealt with the endogeneity issue caused by the reverse causation between HFT and the level of liquidity, introducing various instruments to make proper inference (e.g. introduction of NYSE Autoquote in 2003 by Hendershott et al. (2011)). We follow a similar approach even though the causality from liquidity commonality to HFT is not very clear. We would expect liquidity commonality to have a negative impact on HFT because some HFT strategies are based on cross-market arbitrage opportunities that are profitable as long as the markets for two instruments are not perfectly correlated.

2.2.2 Order book summary statistics

In this section we present summary statistics of variables that capture the state of the order book in the two observation periods.

We measure stock liquidity by half quoted spread, effective spread and half quoted depth. The quoted spread captures the tightness of the order book and in normal functioning markets it is always positive, representing a cost for liquidity takers and a profit for liquidity providers. The quoted half spread (QS) measures the execution cost of a single transaction (i.e. without assuming that the spread will be the same when closing the position). Let $A_{i,t}$ and $B_{i,t}$ be the ask and bid price for stock i at time t, and $M_{i,t}$ the midquote, then QS in basis points is:

$$QS_{i,t} = \frac{A_{i,t} - B_{i,t}}{2M_{i,t}} \, 10,000 \tag{2.1}$$

where t is the timestamp of a quote update (i.e. tick-by-tick) for stock i. Then, we take the Time-Weighted Average (TWA) of $QS_{i,t}$ for each stock-day to give more weight to the QS that is quoted for longer on the book. Since QS does not take into consideration transaction prices, it is a good approximation of the execution cost of small trades only.

The Effective Spread (ES) is the difference between the price paid by the trader and the benchmark value at which the trader wished to trade when he submitted the order, which is assumed to be the prevailing midquote. Since it is based on transaction price, ES also captures trades executed within the spread. In December 2009 the LSE introduced different types of hidden orders: hidden limit order and hidden orders pegged to either mid, bid or ask. Let $P_{i,q}$ be the execution price of transaction q for stock i, then ES is calculated as:

$$ES_{i,q} = I_{i,q} \, \frac{P_{i,q} - M_{i,q}}{M_{i,q}} \, 10,000 \tag{2.2}$$

where $I_{i,q}$ is the trade indicator for buyer (+1) or seller (-1) initiated trade. We classify transactions using a modification of the Lee and Ready (1991). We classify a trade as buyer (seller) initiated if the trade price is above (below) the prevailing mid-quote at the time of the trade. Then, trades that execute at mid-point are classified using the tick test [see Hautsch (2012) for more details]. Before computing the ES, consecutive transaction with the same timestamp and direction are aggregated, resulting in a block trade with price equal to the VWAP of the children transactions. Single transactions executed outside the spread are not considered in the ES calculation. We calculated ES both with simple and volume-weighted average, winsorising stock-day observations on the right-tail at 99.9% level.

Depth measures the liquidity available at the best prices, which captures how much the order book can absorb if a large trade arrives. A deep order book is a good news for traders that demand immediacy, since large market orders can be executed at the best prices. However, depth might be large regardless of the tightness of quote prices. One advantage of depth is that it varies at very high-frequency, due to trade executions but also cancellations and amendments of orders. Thus, it allows us to capture a good degree of activity of the players interacting on the order book. Half-depth in GBP for stock i at time t is calculated as:

$$Depth_{i,t} = (V^a_{i,t} A_{i,t} + V^b_{i,t} B_{i,t})/2 \tag{2.3}$$

where $V_{i,t}^a$ and $V_{i,t}^b$ are the volume of best ask and bid quotes. GBP-Depth can vary due to any quote revision, so it is the variable with the highest frequency of variation. We aggregate Depth with TWA.

Table 2.1 reports daily summary statistics of trading activity and liquidity measures. Panel A and Panel B report the results for the period September 2010 - January 2011 (pre-shock) and March 2011 - July 2011 (post-shock), respectively. Each panel is split between three size groups: All firms, the biggest ten and the smallest ten. All numbers are cross-sectional averages of time-series mean or median.

[Table 2.1 about here.]

On average, the two periods are comparable in terms of realised volatility and turnover, so that we do not expect market makers to mechanically widen quoted spreads after the introduction of the Millennium Exchange. The ten largest and smallest stocks have an ADT around £100m and just below £5m, respectively, which means they classify in the first and third liquidity group for the determination of maximum spread obligations of market makers. Half quoted spread and effective spread decrease across all size groups after the technological shock, suggesting an increase in liquidity, in line with Hendershott et al. (2011) and Brogaard et al. (2014). On average, the half quoted spread decreased by 0.29 bps. The increase in liquidity is unlikely to be demand-driven because the smallest ten stocks experienced a large drop in both trading volume (ADT) and trade size after the Millennium Exchange. Average trade duration went from 45.63 to 41.62 seconds, while for the largest ten stocks it went from 14.33 to 12.85 seconds. The increase can be partly due to the lower latency available.

Only about 5% of transactions are executed within the spread but average effective spreads are lower than quoted half spread in all periods and size groups. This suggests that trades are executed in short periods of time when quoted spreads are narrow. We also report summary statistics for the size of transactions executed against hidden liquidity and those executed off the book. The latter are individual transactions executed outside the spread (see Appendix 2.A for more details) that are not part of the ES calculation. On average, hidden trades are about 20% and 50% larger than normal trades before and after the Millennium Exchange, respectively, while they are always about 100% bigger for small stocks. The average trade size of off-book trades in the Mar 2011 - July 2011 period is around £155,000 compared to £10,000 of normal transactions.

During the observation period we find both different reporting rules for trades/quotes and some reporting delays after the Millennium Exchange. In Table 2.1 we report that the average reporting delay after the Millennium Exchange is 1.77 seconds. We believe this was due to an asynchronism between quote and trade reporting. The reporting delay is solve on 6 June 2011. See Appendix 2.A for more details.

2.2.3 HFQ summary statistics

In addition to using the introduction of Millennium Exchange to get exogenous HFT variation, we are interested in identifying the intraday variation in HFT activity to explain the behaviour of market makers. We use variations of order-to-trade ratio (OTR) to measure HFT activity during the day[10].

Measures based on message traffic are useful to assess risks linked to trading systems overload and they are directly related to technology improvement. OTR identifies mostly passive strategies like market making and does not correlate with statistical arbitrage (which requires low latency but it does not require the submission of a large number of orders). This feature makes it a convenient measure for our purpose. Similarly to Hendershott et al. (2011) and Boehmer et al. (2013), we use a metric based on the hourly message traffic received by the exchange. We call it High-Frequency Quoting (HFQ) and is calculated as:

$$HFQ_{i,h} = \frac{Messages_{h,i}}{Volume_{h,i}}, \tag{2.4}$$

where $Messages_{h,i}$ is the number of quote updates in hour h for stock i and $Volume_h$ represents the trading volume, either in GBP10,000 or in number of trades. The rationale for this measure is that high-frequency market making is based on almost continuous update of bid and ask quotes. A large message traf-

[10]Other indirect measures of HFT activity include the lifetime of orders [Hasbrouck and Saar (2009)] and pure message traffic [Hendershott et al. (2011)].

fic may also identify misconduct practices such as quote spoofing. This measure has some drawbacks. First, illiquid stocks (where HFTs are less active) might genuinely experience a high OTR because few trades are executed despite the large number of orders sent. Second, the OTR does not take into account the speed at which orders are sent. Third, using the number of transactions at the denominator does not take into account that they can be very small (split large order).

Table 2.2 reports the summary statistics of HFQ measured either as messages per GBP10000 trading volume or messages per transaction. Panel A and Panel B report the results for the period September 2010 - January 2011 (pre-shock) and March 2011 - July 2011 (post-shock), respectively. Each panel is split between three size groups: All firms, the biggest ten and the smallest ten. All numbers are cross-sectional averages of time-series mean or median.

[Table 2.2 about here.]

We find that high-frequency quoting activity increases after the technological change, across all size groups and both using turnover (HFT-volume) and number of transactions (HFT-trades) at the denominator. HFT-volume for Big10 stocks is smaller than Small10 because the scale of trading volume is increasing in the firm size. Using HFT-trades we neutralise the scale of volume trading and we find instead that HFTs are most active in the ten largest stocks, with 14.3 messages per transaction after the introduction of the Millennium Exchange, compared to 10.5 for the ten smallest stocks. Brogaard et al. (2014) also find that HFTs are most active in the ten largest FTSE250 stocks, using a different data source and calculation method.

Note that this result is bias downwards by the large number of transactions occurring in the ten largest stocks: about 5000 per day versus 1000 per day on the ten smallest stocks.

2.3 Estimating liquidity commonality with high-frequency data

In this paper we estimate stock-level liquidity commonality as the R^2 of a regression of individual stock liquidity against the common liquidity factors. We obtain daily and hour-specific liquidity commonality using intraday data. Thus, we face three main identification issues. First, we need to define liquidity. Second, we need to identify the number of common factors and estimate them. Third, we need to choose a sampling frequency that is a trade-off between signal and noise.

For the empirical tests of this section, we approximate liquidity by the depth of the order book at the best prices. To avoid giving too much weight to levels of depth that are not meaningful (for instance, available liquidity at bad prices that moves to the top of the book for a few milliseconds after a trade execution) we follow Mykland and Zhang (2016) and pre-average depth at one-minute frequency using a time-weighted average (TWA) scheme.

Depth, and liquidity in general, is a stationary variable with a strong serial correlation. Since we are interested in studying its the relationship across stocks, we follow Hasbrouck and Seppi (2001) and Corwin and Lipson (2011) and remove intraday seasonalities standardising individual depth by time of the day. While these papers use standardised variables in levels, others such as Korajczyk and Sadka (2008) propose to filter liquidity variables with some autoregressive process to remove the long memory of this variable. Thus, we formulate a standard factor model of liquidity and we are interested in estimating the number of factors as a function of the sampling frequency for both liquidity in levels and filtered. Following the notation in Bai and Ng (2002), we assume that the time variation in the liquidity of N stocks can be explained by r common factors in the following model:

$$\underset{(T\times N)}{L} = \underset{(T\times r)}{F}\ \underset{(r\times N)}{\Lambda'} + \underset{(T\times N)}{e} \tag{2.1}$$

where F is a matrix of r unknown common factors and Λ are the respective factor loadings. When N and T are large, a simple method to estimate F is by principal component analysis (PCA). Assuming a number of k factors, the factors and loadings are estimated solving the optimisation problem

$$V(k) = \min_{\Lambda, F^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (L_{it} - \lambda_i^k F_t^k)^2 \tag{2.2}$$

subject to a normalisation to disentangle F^k and Λ^k. If $T > N$, as in our case, it is convenient to concentrate out F^k and use the normalisation $\Lambda^{k'}\Lambda^k/N$. This implies that the estimated loadings $\widetilde{\Lambda}^k$ are equal to $\sqrt{N}$ multiplied by the eigenvectors corresponding to the k-th largest eigenvalues of the covariance matrix $L'L$. Then, the factors can be estimated by the least squares formula simplified as $\widetilde{F}^k = L\widetilde{\Lambda}^k/N$. Our main concern is to choose the right number of factors for our liquidity factor model. Bai and Ng (2002) propose various modifications of information criteria for model selection with an additional penalty that is a function of both N and T. We employ the following two:

$$PC_1(k) = V(k, \widetilde{F}^k) + k\hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln\left(\frac{N+T}{NT}\right) \tag{2.3}$$

$$IC_1(k) = \ln(V(k, \widetilde{F}^k)) + k \left(\frac{N+T}{NT}\right) \ln\left(\frac{N+T}{NT}\right) \tag{2.4}$$

where $V(k, \widetilde{F}^k)$ is the average residual variance of a factor model with k factors. We follow the suggestion in Bai and Ng (2002) and set $\hat{\sigma}^2 = V(kmax, \widetilde{F}^{kmax})$, where $kmax$ is an arbitrary set integer such that $r < kmax$. We choose $kmax = 20$. The main difference between $PC_1(k)$ and $IC_1(k)$ is the dependence of $PC_1(k)$ on $\hat{\sigma}^2$, which essentially is penalising a model where $kmax$ factors are not able to explain a lot of variability. Essentially, this is a penalty for signal-to-noise ratio (SNR) which is useful in our high-frequency setting to analyse the number of significant factors as a function of sampling frequency. For this reason we prefer PC_1, which can be used for estimating the number of factors as a standard information criteria:

$$\hat{k} = \underset{0 \le k \le kmax}{\arg\min}\ PC_1(k) \tag{2.5}$$

Bai and Ng (2002) prove that this estimator of the number of factors is consistent also for some autocorrelation and cross-correlation in the residuals. In simulations, they show that the method performs well.

2.3.1 Factor model structure and microstructure noise

We want to analyse the explanatory power of the standard principal component estimator when faced with microstructure noise and strong autocorrelation[11] to choose the best sampling frequency for our empirical analysis. Because microstructure noise is higher at high sampling frequency, we build various signature plots to unveil how sampling frequency influences, first, the estimator of the number of factors proposed by Bai and Ng (2002) and, second, the strength of liquidity commonality. Signature plots have been used by Andersen et al. (2003) and Mykland and Zhang (2016), among others, to show that the realised volatility highly overstate true volatility when sampling frequency is too high. Signature plots are used in practice to choose a frequency that strikes a balance between information and noise.

In the multivariate analysis of transaction data, and in particular for the estimation of the $N \times N$ covariance matrix of returns, the main concern is the Epps effect, i.e. correlations tend towards zero as the sampling frequency increases, due to the asynchronicity of transactions. Analysing quotes poses a different challenge. In theory, in markets with designated market makers, there are always quotes available for buying and selling. However, these quotes are updated at different speed for different stocks, leading to some quotes being more stale then others. This also causes an Epps effect. For instance, if the depth of one stock does not change within one minute, its variance is zero and correlation is consequently zero.

The impact of long memory on PCA is an issue that has gone almost unexplored in the literature. Vanhatalo and Kulahci (2016), using Montecarlo simulations, show that autocorrelation per se does not affect the results of PCA. However, when the variables have very different autocorrelation coefficients (es-

[11]Starting with Barndorff-Nielsen and Shephard (2004), the literature has developed various methods to estimate the quadratic covariation of returns in the presence of noise. Lately, Aït-Sahalia and Xiu (2017) propose a PCA estimator for high-frequency data. All these methods focus on log-returns and they are not designed for long memory processes like stock liquidity. A comparison of these methods is beyond the scope of this paper.

pecially of different sign), a higher number of principal components are needed to explain a fixed amount of variability, even if the number of true factors is unchanged. Thus, even if autocorrelation is higher at high frequency, we need to be concerned about the dispersion of the coefficient, rather than its magnitude. We investigate this effect by building a signature plot of the cross-sectional dispersion of the first order autocorrelation coefficient.

Fig. 2.1 plots the number of factors estimated with the PC_1 criteria of Bai and Ng (2002) in Eq. (3.5), as a function of the sampling frequency, from 1 minute to 30 minutes, using one year of data in 2010. We set $kmax$ (maximum number of factors) to 20. At each re-sampling step, we take the TWA of individual stock liquidity. We plot the optimal number of factors both for the data in levels (the log of depth) and for the filtered data, which is obtained by taking the residuals of an AR(2) model that is re-estimated at each re-sampling step[12]. Both transformations are standardised by time of the day following Hasbrouck and Seppi (2001). Alongside PC_1 we report the signature plot of SNR, calculated using a model estimated with $kmax$ factors:

$$SNR = \frac{N^{-1}\sum_{i=1}^{N}\text{var}(\widetilde{\lambda}_i^{kmax}\widetilde{F}_t^{kmax})}{V(kmax, \widetilde{F}^{kmax})} \tag{2.6}$$

This is the ratio of the average variance of the estimated common component divided by the average variance of the residuals.

At one minute frequency, the PC_1 criteria estimates that stock liquidity is driven by two common component. With a decrease of sampling frequency, the number of factors increases until it reaches five factors at 30-minute frequency. The SNR, in Panel (b) of Fig. 2.1 shows that this is due to a lower residual variance as we decrease the sampling frequency. Intuitively, at lower frequencies there is less noise and a model with $kmax$ factors has lower residual variance than the same model with higher frequency data. The PC_1 criteria, differently from IC_1, penalises models with high SNR and it is suitable for our applica-

[12]The choice of AR(2) follow from Korajczyk and Sadka (2008) and it might not capture all the serial correlation of some stocks. However, it is enough to model most of the autocorrelation of individual stocks.

tion[13]. We also report the signature plot of PC_1 for the filtered data and notice that it is not a function of sampling frequency. One factor is always enough to explain the commonality in filtered data.

[Figure 2.1 about here.]

To understand if the strength of commonality depends on the sampling frequency, we report in Fig. 2.2 the signature plot of the first two eigenvalues of the $N \times N$ liquidity covariance matrix. Since the variables are standardised to have unit variance, the total variance of L is equal to N and, under the assumption of no-commonality, each eigenvalue - hence the variance of each principal component - is equal to one. Fig. 2.2 plots the eigenvalues, together with their 95% confidence interval calculated assuming multivariate normality[14]. Thus, the standard error of the first eigenvalue, $\lambda^{(1)}$, is equal to $\lambda^{(1)}\sqrt{2/T}$. Since T decreases with sampling frequency, the standard errors will be higher and higher.

We find that liquidity commonality is less strong at higher sampling frequencies, reflecting the SNR pattern that we found in Fig. 2.1. At one-minute frequency, the first PC explains roughly $6/91 = 6.6\%$ of the variability of stock liquidity, while at 30-minute frequency it explains about $12/91 = 13.2\%$. Thus, the two factors suggested by the PC_1 criteria of Bai and Ng (2002) at one-minute frequency explains less variability than at lower frequencies. At the highest frequency, when the noise is the highest, we need more PCs to explain the same amount of variability. However, due to the noise level at this frequency, the PC_1 criteria suggests not to consider them. The filtered data shows very mild commonality, which becomes substantial at lower frequencies. Thus, although at lower frequencies it might be sensible to use filtered variables, we choose to use stock liquidity in levels to estimate commonality with high-frequency data.

[Figure 2.2 about here.]

Finally, we want to explore if the patterns of Fig. 2.2 is due to the heterogeneity of autocorrelation structure of stock liquidity. Fig. 2.3 plots the cross-

[13] In unreported results, we find that IC_1 estimates that the optimal number of factors is equal to two at any frequency.

[14] Although this assumption does not hold, both Hasbrouck and Seppi (2001) and Corwin and Lipson (2011) use it to have some guidance on the significance of the eigenvalues.

sectional dispersion of the N estimated AR1 coefficients as a functions of sampling frequency. As expected, the median autocorrelation is always positive and the highest at one-minute frequency. It decreases with the data frequency and stabilises around 0.6 from five-minute frequency onwards. However, its cross-sectional dispersion remains quite stable across the signature plot, from a minimum of about 0.45 to a maximum of 0.9. As a robustness check, we also report the signature plot for the filtered data, which is not expected to display significant autocorrelation at the first lag. In unreported results, we produce this plot calculating the average of the first three autocorrelation coefficients, but the results are qualitatively equal.

Because the dispersion of autocorrelation coefficients is stable across sampling frequencies, we exclude it as a possible factor driving the patterns in Figs. 2.1 - 2.2. Instead, the steep decrease in the median autocorrelation from one- to five- minute is related to the steep increase in the first and second eigenvalues, suggesting that very high levels of memory increases the number of PCs needed to explain a certain amount of variability.

[Figure 2.3 about here.]

The results of this section shed light on the impact of long memory and sampling frequency on the structure of a factor model for stock liquidity. First, we find empirical support for Hasbrouck and Seppi (2001)'s decision of using a three factor model when the data are sampled at 15-minute frequency. The PC_1 criteria of Bai and Ng (2002) suggests that a three factor model can be used for frequencies from eight to 15 minutes. Second, we find that at the highest frequency, when the noise is the highest (low SNR), we need a higher number of PCs to explain a fixed amount of variability. This can be partly explained by the large autocorrelation of liquidity at very high frequencies. However, the PC_1 criteria with a penalty related to the signal to noise ratio does not suggest to use more than two factors at one-minute frequency.

To strike a balance between signal and noise, we choose to use a three-factor model with data sampled at 10-minute frequency.

2.4 Testing if HFQ increases liquidity commonality

In this section we present our methodology to test if the technological change of 14 February 2011, when the LSE introduced the Millennium Exchange trading system, led to an increase in liquidity commonality. First, we compare the eigenvalues in two five-month period before and after the shock, extending the analysis of Hasbrouck and Seppi (2001). Second, we perform a panel data analysis to formally test an increase in the liquidity commonality, controlling for market conditions.

2.4.1 Test 1: Eigenvalues analysis

Table 2.3 reports the strength of liquidity commonality before and after the introduction of the Millennium Exchange on 14 February 2011. We exclude February 2011 from the analysis due to the large number of data errors and the computer glitch of 25 February 2011. Panel A and Panel B report the results for the period September 2010 to January 2011 and March 2011 to July 2011, respectively. All variables have been resampled at 10-minute frequency using different sampling schemes: Depth and bid-offer (BO) spread are aggregated using a time-weighted average; Effective spread is aggregated with simple average; Trading volume is the sum of volume within 10-minute time bins. Following Hasbrouck and Seppi (2001), all variables have been standardised by time-of-the-day. We report the first three eigenvalues of the $N \times N$ covariance matrix of stock liquidity variables and also the results for returns and trading volume. The last column of Table 2.3 reports the strength of commonality, which is calculated as the sum of the eigenvalues divided by $N = 91$. Since the variables are standardised, under the null of no commonality all eigenvalues would be equal to one. Table 2.3 shows that all liquidity variables - depth, bid-offer spread and effective spread - are driven by common factors.

[Table 2.3 about here.]

We find that liquidity commonality increased after the introduction of the Millennium Exchange and this result is robust to different measures of liquidity. For depth, the variability explained by the first three principal components increased from 13% to 15%. BO spread commonality increased from 21.8% to 28.24%, while commonality in effective spreads increased from 12.26% to 17.26%. The results are reported in Panel B of Table 2.3. The first eigenvalue of depth, bid-offer spread and effective spread is statistically higher than before the technology shock, at 99% confidence level. The difference in eigenvalues has been tested assuming multivariate normality of the underlying data, as explained in Section 2.3. Although this might be a strong assumption, various authors have followed this approach as it gives an approximate estimate of the standard errors [Hasbrouck and Seppi (2001), Corwin and Lipson (2011)]. Furthermore, we find that commonality in trading volume is not statistically different after the Millennium Exchange, implying that the increase cannot be explained by demand-side factors.

The limitation of this test is that we cannot control for market conditions or stock-specific characteristics to be sure that liquidity commonality increased regardless of, for instance, stock return volatility. For this reasons, we perform a panel data analysis, which we present in the next section.

2.4.2 Test 2: Panel data analysis

In this section we present a panel data analysis to test if liquidity commonality increased after the introduction of Millennium Exchange. Our dependent variable is monthly commonality in quoted half depth sampled at 10-minute frequency. Depth has the advantage of varying at very high frequency compared to other liquidity variables. Even at one-minute frequency it is rarely stale for FTSE100 stocks.

To test if liquidity commonality for stock i in month t, $\text{LIQ\_COM}_{i,t}$, increased after the technological update, we formulate the following fixed-effect (FE) model that takes into account both firm effect and time effect:

$$\text{LIQ\_COM}_{i,t} = \alpha_0 + \alpha_1 D_t + \beta_0 RV_{i,t} + \beta_1 RV_{i,t} D_t + \eta_i + \gamma_t + e_{i,t}, \quad e_{i,t} \sim iidN(0, \sigma^2) \tag{2.1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. D_t is a dummy variable that takes the value of 1 from March 2011 until the end of the sample, $RV_{i,t}$ is the monthly realised volatility, η_i is the firm fixed effect and γ_t is a flexible time trend that influences all stocks at time t by the same amount. In our sample, we expect the time effect to be particularly strong because of the calendar effect in trading volumes of December and January, so the γ_t variables play a crucial role in the model. Eq. (2.1) is a two-way FE model and it can be estimated in the parametrised version presented above or also by demeaning the variables. It can be estimated via OLS if η_i and γ_t are set to equal to zero and by random-effect estimator if η_i and γ_t are estimated together with $e_{i,t}$, assuming they are random variables. The variance σ^2 is homogeneous across time and stocks.

Table 2.4 reports the results of five panel data models. Model (1) and (2) are the OLS and firm FE model, respectively. Model (3) corresponds to the estimation results for the two-way FE model in Eq. (2.1). This model is able to capture 5.8% of variability in liquidity commonality and yields interesting findings. First, the panel analysis confirms that liquidity commonality increases after the shock, as we find a statistically significant increase in the post-shock period. The long-run average of $\text{LIQ\_COM}_{i,t}$ increases from 8.28% to 15.15% after the Millennium Exchange, suggesting that HFT activity increases liquidity commonality. This result strongly supports our hypothesis that high-frequency traders coordinates their quoting activity across multiple securities, increasing liquidity commonality. This could be due to both informational reasons, such as in the cross-asset learning model of Cespa and Foucault (2014), and funding constraints, such as in the models by Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2009). However, we think that, although these mechanisms are not mutually exclusive, at higher frequencies the informational channel prevails also because funding costs are not changing at high frequencies.

Second, returns volatility enters the equation with a positive sign, implying that when returns are more volatile stock liquidity co-moves more strongly with market-wide liquidity. An increase of 10% in stock volatility leads to 6.43 higher liquidity commonality (i.e. 6.43% higher explanatory power of common factors). We interpret this result as follows: when the fundamental value of a stock is difficult to assess, liquidity becomes less idiosyncratic and tends to follow the market liquidity. This result supports the evidence by Hameed et al. (2010), who find, at the aggregate level, that when market volatility is high liquidity commonality is higher. Model (4) and (5) of Table 2.4 report the random-effect version of the models and results are qualitatively the same.

[Table 2.4 about here.]

2.4.3 Robustness checks

We carry out several robustness checks. First, we estimate the model before and after the shock. Second, we use bid-offer instead of realised volatility as control variable. Finally, in unreported results, we apply a logistic transformation to our dependent variable and we also estimate all models using OLS plus standard errors clustering to correct for firm and time effect, as proposed by Petersen (2009). However, the clustering of standard error is a robust correction only when the number of clusters is large, which is not the case for our time cluster (ten months)[15].

Table 2.5 reports the estimation results for two separate periods, before Millennium Exchange (September 2010 to January 2011) in Panel A and after Millennium Exchange (March 2011 to July 2011) in Panel B. Overall, the results confirm the findings of our previous section. Controlling for stock volatility, the average liquidity commonality in the first period is always lower than in the second period. It is interesting to notice that stock volatility is positive and significantly different from zero only before the Millennium Exchange.

[Table 2.5 about here.]

Table 2.6 repeats the analysis using the bid-offer spread instead of the volatility as exogenous variable. Also here the results are similar.

[Table 2.6 about here.]

2.5 Intraday liquidity commonality and HFQ

In this section, we analyse the intraday relationship between HFQ, liquidity commonality, returns volatility, liquidity, order flow and order imbalance. We use the five-month period after the introduction of the Millennium Exchange

[15] The same reasoning holds for the Fama-MacBeth estimator, which was developed to correct strong time effect in financial data.

(i.e. from March 2011 to July 2011) to analyse these relationships in the new institutional setting[16].

Fig. 2.4 plots, for every hour of the trading day, the average HFQ, liquidity commonality and the bid-offer spread. Each plot consists of three lines: one for all stocks, one representing the largest ten and one the smallest ten firms ranked by market capitalisation at the beginning of the sample. Overall, we find that HFTs increase their participation from market opening until 12:00PM, when stocks are most illiquid and market makers are needed the most. This can be explained by two forces. First, some HFTs are designated market makers that are obliged to post quotes for 90% of the trading day. These players build their inventory in the first part of the day and try to avoid holding large positions overnight. Second, market-makers have the opportunity to earn larger spreads. This is in line with Menkveld (2013), who shows that HFTs act as an accelerated version of traditional market makers, i.e. they earn the bid-offer spread while managing inventory risk[17].

[Figure 2.4 about here.]

Fig. 2.4c shows that the average bid-offer spread at market opening is about 6 bps, with the 10 smallest stocks trading at 9 bps and the 10 largest ones trading below 2.5 bps. The shape of the intraday average BO spread is similar across the three stock categories, with a steep decrease from 8:30AM to 10AM, and a further decrease after 2:00PM. Furthermore, Fig. 2.4b suggests that this lack of liquidity is common across stocks at market opening, because the price has not been discovered and the market is still gathering stock-specific information. So, we can think of the market opening as a situation of systematic illiquidity, where the market has digested only the market-wide overnight information. This affects all stocks and commonality is high. HFQ is high because prices are uncertain and quotes need a lot of rebalancing, and at the same time there is

[16] Furthermore, we do not find a change in the intraday relationship between variables before and after the shock, apart from the parallel shifts in liquidity and its commonality.

[17] The common strategy of an high-frequency trading firm (e.g. KCG Holdings, Virtu Financial) consists in quoting on two market venues. In one incumbent venue (e.g. the London Stock Exchange) they have a small participation rate and in the other, a high-growth entrant market venue (e.g. BATS Chi-X Europe), they have a high (60-80%) participation rate. These players aim at closing the trading day with an almost zero inventory.

little trading interest. As the time goes by, stock-specific information arise and liquidity commonality decreases. We find that, at any time during the day, liquidity commonality is always higher for larger stocks, which suggests that there is a stronger common factor among large stocks.

Fig. 2.4a plots HFQ throughout the day. The average across all stocks has an inverted-U shape, suggesting that after 12:00PM HFTs start to be less active, with the number of messages per GBP10,000 of trading going from 60 down to about 20 at 4:00PM. This suggests that, on average, there is an inverted relationship between intraday HFQ and liquidity commonality. However, we find a different pattern for the largest ten stocks, where HFQ keeps increasing during the day, reaching a level of about 11 messages per GBP10,000 trading volume at the end of the day. Thus, at the end of the day HFTs focus on large stocks. This evidence is particular striking because the large trading volume at the end of the day should bias downwards the estimate of HFQ. A possible explanation for this result is that at the end of the day, HFTs concentrate on the larger stocks because trading volumes are very large. The high liquidity commonality (particularly large for big stocks) is likely to be demand-driven, due to a lot of uninformed basket trading benchmarked against VWAP or end-of-day price.

2.5.1 Liquidity provision throughout the day

We further explore the behaviour of HFTs throughout the day, using a set of variables that are related to the demand and supply of liquidity. For each stock, we calculate hourly realised volatility, the size (standard deviation) of its order flow and the size (standard deviation) of average order imbalance. The order flow, which is calculated as the sum of signed transactions within one hour, is generally considered as containing information, which can be either market-wide or idiosyncratic [Hasbrouck and Seppi (2001)]. Order imbalance, which is based on limit orders, is the visible demand/offer imbalance and it is a high-frequency predictor of miquote change [Cont et al. (2014)]. Order imbalance is calculated at tick-by-tick frequency and then average at one minute frequency using a TWA sampling scheme. The tick-by-tick order imbalance is calculated as:

$$OI_{i,t} = \frac{V^a_{i,t} A_{i,t} - V^b_{i,t} B_{i,t}}{Depth_{i,t}} \tag{2.1}$$

where $V^a_{i,t}$ and $V^b_{i,t}$ are the volume of best ask and bid quotes.

Fig. 2.5 plots, for every hour of the trading day, realised volatility and the standard deviation of both the order flow and the order imbalance. Each plot consists of three lines: one for all stocks, one representing the largest ten and one the smallest ten firms.

[Figure 2.5 about here.]

In Fig. 2.5, we find that intraday RV has a U-shaped pattern, with smaller stocks experiencing higher volatility at any time of the day. The high volatility at the end of the day together with the evidence that all stocks are liquid is puzzling because in a linear equilibrium the variance of returns is increasing in illiquidity, i.e. $\Delta p_t = Illiq_t OF_t$, where OF_t is the order flow at time t. However, this formula can be reconciled if the variance of the order flow is very high. Fig. 2.5b shows that this is indeed the case and the pattern is particularly convex for the smallest 10 stocks in our sample. We find a U-shaped pattern in the size of the order flow. In the morning, the information is systematic and we can argue that some trading happens on the back of overnight news. At the end of the day a lot of trading takes place and information is more idiosyncratic. Fig. 2.5c plots the standard deviation of the order imbalance and we find that it is the lowest at the end of the day for all stock categories, which suggests that liquidity providers are more efficient at absorbing demand shocks at the end of the day. This finding, together with the lower overall HFQ, suggests that at the end of the day HFTs are able to predict temporary demand shocks and provide liquidity using marketable orders (which need less rebalancing of limit orders). This behaviour would also help them to quickly build and unwind inventory before the end of the trading day.

Overall, these results points to the evidence that liquidity is provided in a different way throughout the day. In the morning, when bid-offer spread are large across many stocks (high liquidity commonality) market makers are active in the market because they are incentivised by the large potential profits. They enter passive two-side limit orders without information, as classic market makers, earning the spread to compensate for inventory risk and adverse selection.

In the afternoon, some liquidity provider might be able to predict order flow and want to absorb temporary demand shocks quickly with market orders[18].

2.6 Conclusions

This paper investigated whether and to what extent high-frequency quoting (HFQ), a type of activity carried out by high-frequency traders (HFT), affect the interconnectedness of equity markets, captured by high-frequency liquidity commonality. First, we asked whether an increase in HFQ is related to higher liquidity commonality. Second, we investigated the intraday relationship between HFQ, liquidity commonality and the level of stock volatility, liquidity, order flow and order imbalance. To test our hypotheses, we used tick-by-tick data for the FTSE100 stocks listed on the London Stock Exchange, from September 2010 to July 2011. We estimated stock-level liquidity commonality as the explanatory power of common liquidity factors, estimated by principal component analysis (PCA), on individual stock liquidity.

We find that liquidity commonality increased after the introduction of the Millennium Exchange and our results are robust to different liquidity measures and controlling for both volatility and bid-offer spread. This result strongly supports our hypothesis that high-frequency traders coordinates their quoting activity across multiple securities. Furthermore, we find that HFTs increase their participation from market opening until 12pm, when stocks are most illiquid and market makers are needed the most. At the end of day, they focus more on larger stocks where trading is higher, to compensate for the smaller bid-offer spreads that depress their per-transaction profit. Finally, we find evidence that suggests that liquidity providers use marketable orders to absorb demand shocks more efficiently at the end of the day.

[18] Also, quoted bid-offer spread might seem very narrow because the comprise systematic internalisers quotes.

Table 2.1: Summary statistics (order book)

The table presents daily trading activity and liquidity measures for 91 stocks that entered the FTSE100 index between January 2010 and December 2011, with at least 100 transactions for each day of the sample. We focus on the order book of the SETS platform of the London Stock Exchange. Panel A and Panel B report the statistics for the period September 2010 to January 2011 (pre-shock) and March 2011 to July 2011 (post-shock), respectively. All numbers are cross-sectional averages of time-series *Mean* or *Median*. Half depth and Half BO spread are time-weighted average; Effective spread is reported both as simple and volume-weighted average; hidden trades are defined as those executed within the spread; off-book trades are defined as single transactions, outside the spread, that do not respect the minimum tick restrictions. Sweep trades are aggregated in a single transaction.

Panel A: Sep 2010 - Jan 2011

	Big 10		All		Small 10	
	Mean	Median	Mean	Median	Mean	Median
Market cap.	97.62		24.32		2.72	
Number of trades	5510.80	5377.30	2117.32	2017.01	1076.36	987
Total volume ('000 GBP)	94131.89	88379.24	27543.82	25266.39	5502.93	4732.25
Trade size (GBP)	17054.77	15728.39	10668.98	9686.77	5237.58	4603.17
Trade size, hidden	13724.64	10884.76	12709.11	8002.10	9739.27	5429.98
Trade size, off-book	246887.18	137201.18	155504.27	75769.31	67865.40	36046.50
Trade duration (sec)	14.33	13.62	45.63	43.25	62.39	58.65
Half depth (GBP)	57636.72	55856.75	45020.63	43806.06	13218.49	12472.90
Half BO spread (bp)	2.19	2.19	4.75	4.71	6.66	6.57
Eff. spread (bp)	1.75	1.74	3.60	3.59	4.85	4.83
VW Eff. spread (bp)	1.81	1.80	3.76	3.71	4.97	4.91
Realised volatility (bp)	129.36	126.50	144.82	141.21	170.87	163.97
Perc. hidden trades	5.27		5.37		5.66	
Perc. off-book trades	1.43		0.85		0.83	
Reporting delay (sec)	0		0		0	

Panel B: Mar 2011 - Jul 2011

	Big 10		All		Small 10	
	Mean	Median	Mean	Median	Mean	Median
Market cap.	100.97		25.57		3.07	
Number of trades	5711.19	5473.60	2221.32	2078.88	1093.08	998.40
Total volume ('000 GBP)	105584.43	96292.48	27996.20	25154.69	4511.47	3915.74
Trade size (GBP)	17497.84	16257.78	10038.15	9231.12	4217.63	3879.70
Trade size, hidden	18009.24	15428.20	15580.43	10573.91	8172.04	5343.51
Trade size, off-book	264171.80	160215.57	185773.38	86636.72	77389.84	28271.64
Trade duration (sec)	12.85	12.69	41.62	40.44	61.33	59.17
Half depth (GBP)	67198.95	66171.97	46435.36	44646.93	11806.76	11389.01
Half BO spread (bp)	1.96	1.92	4.46	4.39	6.24	6.15
Eff. spread (bp)	1.56	1.53	3.44	3.39	4.54	4.49
VW Eff. spread (bp)	1.61	1.58	3.51	3.45	4.66	4.57
Realised volatility (bp)	122.46	118.57	145.02	139.87	172.08	165.31
Perc. hidden trades	4.31		4.15		3.79	
Perc. off-book trades	0.97		0.64		0.40	
Reporting delay (sec)	1.38		1.77		2.33	

Table 2.2: Summary statistics (HFT)

The table presents daily trading activity and liquidity measures for 91 FTSE100 stocks that entered the index between January 2010 and December 2011, with at least 100 transactions each any day of the sample. Panel A and Panel B report the statistics for the period September 2010 to January 2011 (pre-shock) and March 2011 to July 2011 (post-shock), respectively. All numbers are cross-sectional averages of time-series *Mean* or *Median*.

Panel A: Sep 2010 - Jan 2011

	Big 10		All		Small 10	
	Mean	Median	Mean	Median	Mean	Median
HFT-volume (GBP10000)	9.20	8.85	15.63	14.74	21.29	19.92
HFT-trades	12.86	12.65	12.80	12.37	8.83	8.41

Panel B: Mar 2011 - Jul 2011

	Big 10		All		Small 10	
	Mean	Median	Mean	Median	Mean	Median
HFT-volume (GBP10000)	9.35	9.08	16.51	15.56	29.60	26.90
HFT-trades	14.30	13.92	13.18	12.76	10.50	10.05

Table 2.3: Liquidity commonality

The table reports the strength of liquidity commonality before and after the introduction of the Millennium Exchange on 14 February 2011. Panel A and Panel B report the results for the period September 2010 to January 2011 and March 2011 to July 2011, respectively. All variables have been resampled at 10-minute frequency. Depth and bid-offer (BO) spread are aggregated using a time-weighted average. Effective spread is aggregated with simple average. Trading volume is the sum of volume within 10-minute time bins. The last column of the table reports the strength of commonality, which is calculated as the sum of the eigenvalues divided by N. The number of stock is $N = 91$. We test the null hypothesis of equal eigenvalues, assuming multivariate normality. a,b and c indicate 99%, 95% and 90% confidence levels.

Panel A: Sep 2010 - Jan 2011

	Eigenvalues			
	First	Second	Third	Cumulative Explained Variance (%)
Return	23.61	2.78	1.76	30.94
Depth	6.14	4.24	2.21	13.84
BO spread (%)	12.56	4.51	2.77	21.80
BO spread (ticks)	10.34	5.61	2.85	20.66
Effective spread	5.90	3.34	1.92	12.26
Trading Volume	8.51	1.77	1.47	12.90

Panel B: Mar 2011 - Jul 2011

	Eigenvalues			
	First	Second	Third	Cumulative Explained Variance (%)
Return	25.26^{b}	2.78	2.39^{a}	33.44
Depth	7.02^{a}	4.58^{a}	2.32^{c}	15.30
BO spread (%)	18.36^{a}	4.49	2.86	28.24
BO spread (ticks)	18.04^{a}	3.84^{a}	2.48^{a}	26.77
Effective spread	10.56^{a}	3.30	1.84	17.26
Trading Volume	8.49	1.81	1.53	13.01

Table 2.4: Panel estimation

The table reports estimation results for five panel data models. The dependent variable is the stock-specific monthly liquidity commonality. *Volatility* is the monthly realised volatility calculated with pre-averaged one-minute data. *Post* is a dummy variable taking value 1 from March 2011. *Post*RV* is an interaction variable. Model (1) is estimated with Pooled OLS.

	(1)	(2)	(3)	(4)	(5)
Volatility	0.173	0.281	0.643	0.222	0.463
	(0.84)	(1.10)	(2.39)	(1.00)	(2.02)
Post	1.691	1.992	6.866	1.849	6.596
	(0.95)	(1.16)	(3.35)	(1.10)	(3.25)
Post*RV	-0.258	-0.309	-0.461	-0.285	-0.406
	(-0.94)	(-1.17)	(-1.70)	(-1.10)	(-1.53)
Constant	12.409	11.753	8.285	12.112	9.373
	(9.32)	(7.35)	(4.54)	(8.37)	(5.67)
Firm fixed effect		Yes	Yes		
Firm random effect				Yes	Yes
Time dummies			Yes		Yes
Observations	910	910	910	910	910
R-squared	0.001	0.002	0.058	0.001	0.042

Table 2.5: Panel estimation (sample split)

The table reports estimation results for two separate periods, before Millennium Exchange (September 2010 to January 2011) in Panel A and after Millennium Exchange (March 2011 to July 2011) in Panel B. Model (1) is estimated with Pooled OLS.

Panel A: Sep 2010 - Jan 2011

	(1)	(2)	(3)	(4)	(5)
Volatility	0.173	0.570	1.304	0.256	0.525
	(0.81)	(1.59)	(3.30)	(1.06)	(2.08)
Constant	12.409	9.984	4.286	11.902	9.001
	(8.99)	(4.48)	(1.69)	(7.59)	(5.03)
Firm fixed effect		Yes	Yes		
Firm random effect				Yes	Yes
Time dummies			Yes		Yes
Observations	455	455	455	455	455
R-squared	0.001	0.007	0.089	0.001	0.046

Panel B: Mar 2011 - Jul 2011

	(1)	(2)	(3)	(4)	(5)
Volatility	-0.085	-0.211	-0.105	-0.121	-0.049
	(-0.49)	(-0.91)	(-0.41)	(-0.66)	(-0.25)
Constant	14.100	14.889	12.100	14.322	11.682
	(12.32)	(9.96)	(5.91)	(11.52)	(6.99)
Firm fixed effect		Yes	Yes		
Firm random effect				Yes	Yes
Time dummies			Yes		Yes
Observations	455	455	455	455	455
R-squared	0.001	0.002	0.055	0.001	0.037

Table 2.6: Panel estimation (bid-offer)

The table reports estimation results for five panel data models. The dependent variable is the stock-specific monthly liquidity commonality (transformed to be defined on the real line). *Bid-offer spread* is the time-weighted monthly average bid-offer spread. *Post* is a dummy variable taking value 1 from March 2011. *Post*BO* is an interaction variable. Model (1) is estimated with Pooled OLS.

	(1)	(2)	(3)	(4)	(5)
Bid-offer spread	0.087	-1.159	-0.702	-0.078	0.034
	(0.37)	(-2.39)	(-1.40)	(-0.28)	(0.12)
Post	1.047	1.580	5.288	1.458	5.233
	(0.63)	(1.01)	(2.84)	(0.94)	(2.82)
Post*BO	-0.210	-0.413	-0.350	-0.314	-0.245
	(-0.61)	(-1.28)	(-1.09)	(-0.98)	(-0.77)
Constant	13.061	18.896	15.546	13.833	12.013
	(11.04)	(8.21)	(6.11)	(9.80)	(7.44)
Firm fixed effect		Yes	Yes		
Firm random effect				Yes	Yes
Time dummies			Yes		Yes
Observations	910	910	910	910	910
R-squared	0.000	0.012	0.056	0.000	0.040

Figure 2.1: Signature plot: number of factors

This figure plots the number of factors estimated with the PC_1 criteria of Bai and Ng (2002), reported in Eq. (3.5), as a function of the sampling frequency, from 1 minute to 30 minutes, using one year of data in 2010. At each re-sampling step, we take the time-weighted average of individual stock liquidity. We plot the optimal number of factors both for the data in levels (the log of depth) and for the filtered data, which is obtained by taking the residuals of an AR(2) model that is re-estimated at each re-sampling step. Both transformations are standardised by time of the day. SNR is the ratio of variance of estimated common component over the variance of the residuals.

(a) PC_1 criteria, $kmax = 20$ **(b)** Signal-to-noise ratio at $kmax$

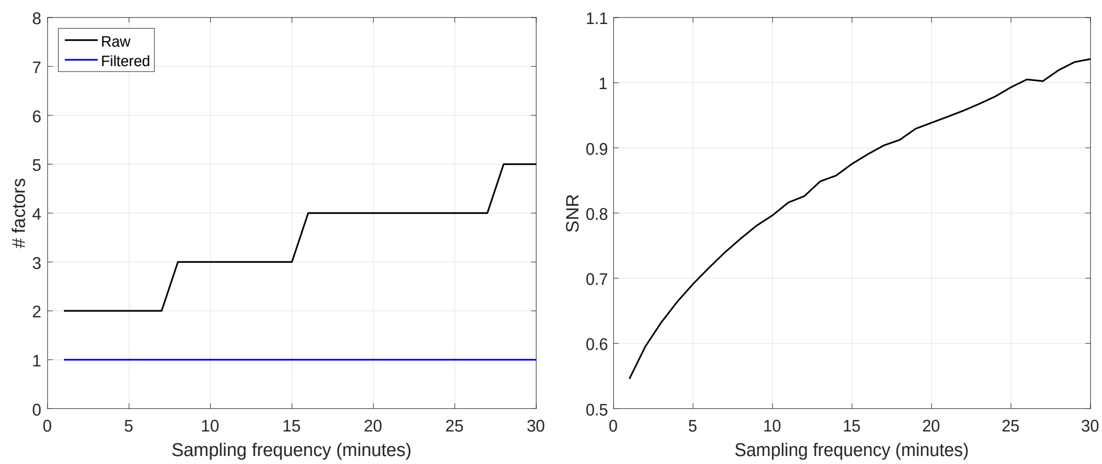


Figure 2.2: Signature plot: eigenvalues

This figure plots the first and second eigenvalue of the $N \times N$ liquidity covariance matrix, together with their 95% confidence interval (calculated assuming multivariate normality), as a function of the sampling frequency, from 1 minute to 30 minutes, using one year of data in 2010. At each re-sampling step, we take the time-weighted average of individual stock liquidity. We plot the eigenvalues both for the data in levels (the log of depth) and for the filtered data, which is obtained by taking the residuals of an AR(2) model that is re-estimated at each re-sampling step. Both transformations are standardised by time of the day.

(a) First eigenvalue

(b) Second eigenvalue

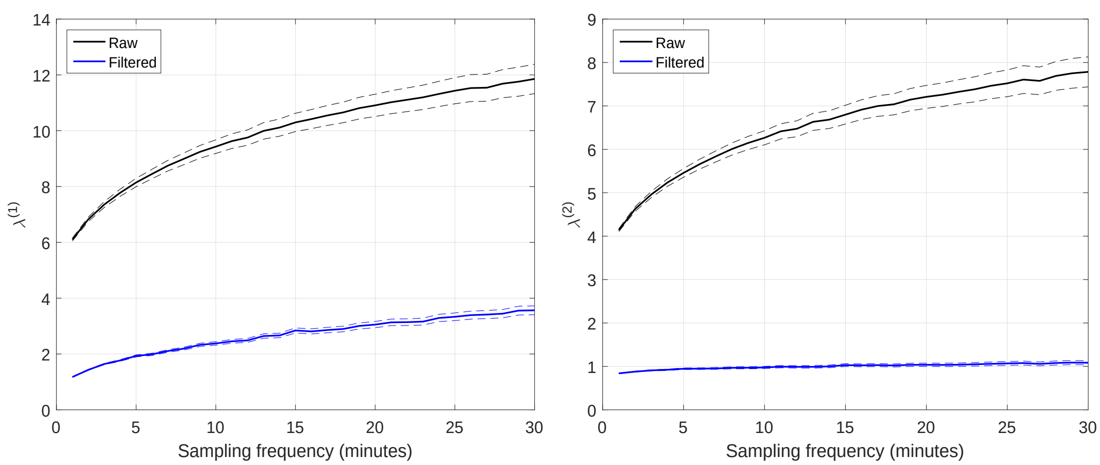
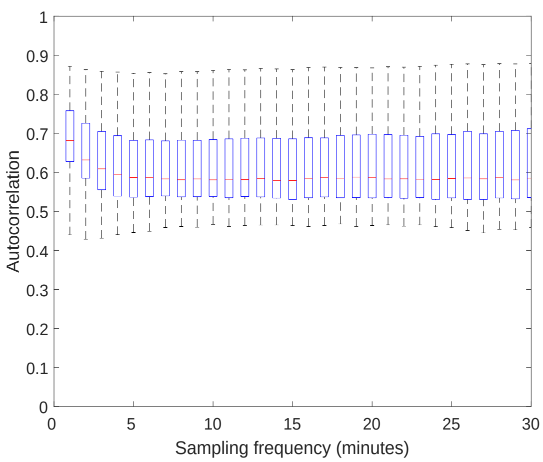


Figure 2.3: Signature plot: autocorrelation dispersion

This figure plots the cross-sectional dispersion of the N estimated AR1 coefficients as a functions of sampling frequency, from 1 minute to 30 minutes, using one year of data in 2010. At each re-sampling step, we take the time-weighted average of individual stock liquidity. We plot the eigenvalues both for the data in levels (the log of depth) and for the filtered data, which is obtained by taking the residuals of an AR(2) model that is re-estimated at each re-sampling step. Both transformations are standardised by time of the day.

(a) Depth



(b) AR(2)-filtered depth

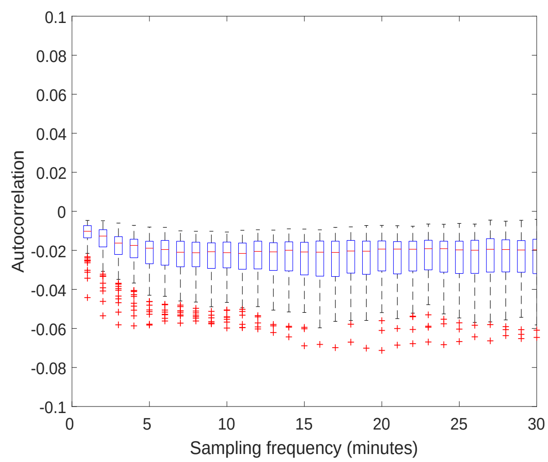


Figure 2.4: Intraday patterns in HFQ, commonality and BO

The figure presents, for every hour of the trading day, the average HFQ, liquidity commonality and bid-offer spread. Each plot presents the statistics for all stocks, plus the largest 10 and the smallest 10 firms.

(a) High-frequency quoting **(b)** Liquidity commonality **(c)** Bid-offer spread

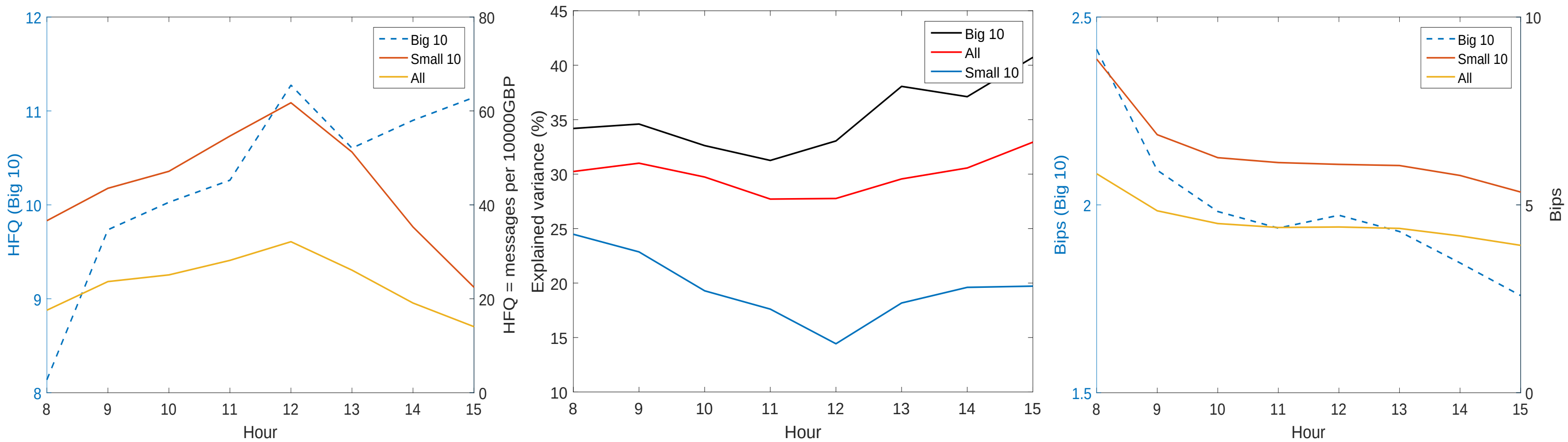
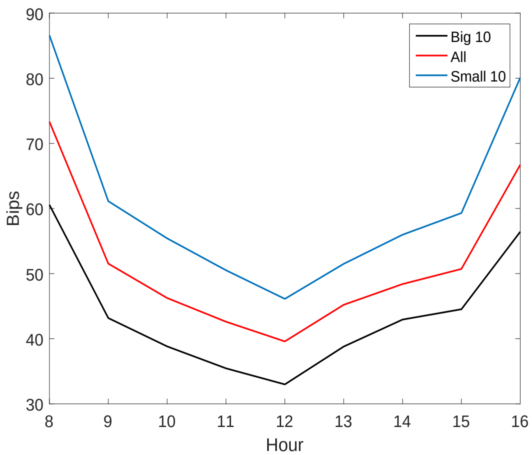


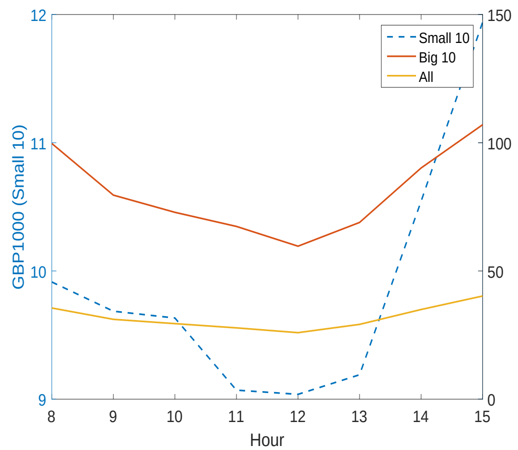
Figure 2.5: Intraday patterns in volatility and order flow

The figure presents, for every hour of the trading day, the average realised volatility, the standard deviation of the order flow and the order imbalance. Each plot presents the statistics for all stocks, plus the largest 10 and the smallest 10 firms.

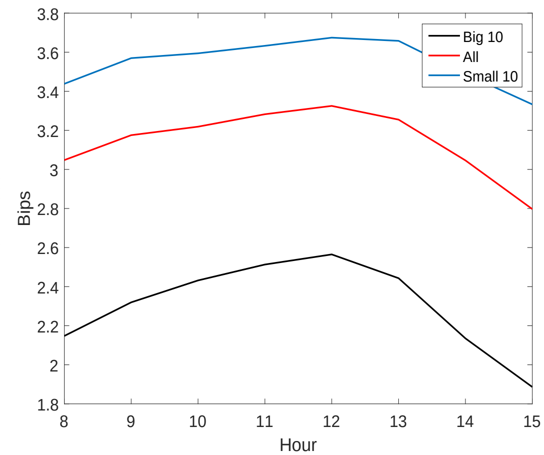
(a) Realised volatility



(b) Size order flow



(c) Size order imbalance



Appendices

Appendix 2.A Data details

2.A.1 Trade/quote reporting rules and delays

During the observation period we find both different reporting rules for trades/quotes and some reporting delays after Millennium Exchange. In this section we present them and we explain our procedure for quotes-trades matching.

In the *Time & Sales* data base provided by Thompson Reuters Tick History (TRTH), upon arrival of a parent trade, all resting orders which the trade is matched against are reported on the book. Before the Millennium Exchange, the children trades are reported in bulk, i.e. after the prevailing quotes we find a list of children transaction that automatically executed to satisfy the trading demand in full or partially. The forthcoming best quotes are the remaining demand/offer interests on the book after the parent trade has been executed. Thus, if a trade is sweeping the book, our aggregated trade has a VWAP away from the previous prevailing midquote and the new quoted spread would be large unless liquidity is quickly replenished.

However, from 14 February 2011, also the quote updates are reported. i.e. after every children transaction we find a new data entry containing the available liquidity remained on the book. Furthermore, we find a trade reporting delay from 14 February 2011, which is then corrected on 6 June 2011, probably due to the new trading system transmitting orders quicker than trades were reported.

How we matched trades and quotes To match the transactions with the relevant quote (after the Millennium Exchange only) we used an algorithm similar to the one proposed by Hautsch (2012). First, we search for a perfect match using the quotes up to 30 seconds before the trade timestamp. Second, the unmatched transactions are matched to the quote prevailing at the median delay time before. The algorithm proceeds as follows. For each stock-day, for each transaction we seek, in the previous 30 seconds:

1. A quote with the same price (either bid or ask) and size equal to the trade volume. Or,
2. A quote preceding the quotation with the same price (either bid or ask) and with change in size equal to the transaction volume,
3. Calculate median daily delay time,
4. Match the transactions that are not perfectly matched ($< 1\%$) against the quotes prevailing at the median delay time before the trade.

The timestamps of the matched quotes above are used to calculate midquotes, flag trades and calculate effective spreads. Trades executed at the prevailing midquote are flagged using the tick rule as usual.

This has implications on the HFQ measure To compare the number of messages (quote updates) before and after Millennium, we add the number of children trades to the number of quote updates before Millennium.

2.A.2 Data cleaning

The rationale behind our data cleaning method is to use, when possible, the rules of SETS (e.g. circuit breakers) to eliminate the same data errors that other authors proposed to eliminate using statistical procedures [Brownlees and Gallo (2006) and Hautsch (2012)].

- We exclude February 2011 because we find several data errors and because of a computer glitch on 25 February 2011, when SETS was halted until 12:15.
- We remove days with an Exchange Delivery Settlement Price (EDSP) intraday auction. This auction occurs on the expiry day of FTSE100 Index Futures and Options (the third Friday of every month), when trading on the underlying stocks is halted, from 10:10 AM to between 10:15 and 10:29 AM, to determine the price at which the derivative is settled. At the end of the auction, the share price for each stock is the price at which most volume can be executed. In our data, the EDSP auction is flagged by large negative bid-offer spreads.
- We remove two half-day trading days per year, i.e. the last trading day before Christmas and the last trading day before New Year's Day.

- We remove the first 15 minutes and the last minute of each trading day (we keep the last minutes of the day because we are interested in the activity towards the end of day)
- We delete quotes with prices that breach the circuit breakers static threshold of SETS, which is set to 10% deviation from the opening price (revised down to 8% on 19 January 2015). As opening price we use the median midquote in the first hour of trading[19].
- We delete all data entries with non-positive price or volume (of either a quote or trade), trades with missing price or volume, quotes with negative bid-offer spread
- Single trades whose price is outside the spread are excluded but we keep track of them. These are likely to be large transactions executed off the book and reported with a delay.
- We cap the half quoted spread at 100 bps.
- We winsorise the right tail of liquidity measures distribution at 99.9%

2.A.3 Special trades

Also, we are able to identify the following special transactions:

1. "Sweep trades", i.e. trades that walk the book, are reported as multiple transactions, with the first executed at the best quotes and the successive ones at increasingly worse prices. So they are normally aggregated into one trade.
2. Off-book trades: single transactions executed outside the spread. Note three features of these trades. First, some are executed at prices that would have breached the static circuit breaker threshold, but they are in line with closing price of previous days (these could be also data errors). Second, they are very large. Third, often they do not conform with the minimum tick size, i.e. the price seems to be a percentage off the midquote (pegged orders). These trades can not be flagged as buy or sell. Because of their characteristics they could be either data errors or transactions negotiated off the book and reported with a delay. We exclude these trades from Effective spread and volume calculations.

[19]We prefer this rule to the method of Brownlees and Gallo (2006) based on local estimators.

3. SIs trades are subject to pre-trade transparency. The SI trades are reported in a separate book, but the point here is that a broker does not have incentive to be a SI, so if an order is internalised not on systematic basis the broker is not a SI and the trade is reported on the SETS tape
4. Trades within the spread. Only a small part of trades are executed within the spread. LSE allows hidden liquidity on LIT since December 2009, so these are likely to be orders matched against fully or partially hidden liquidity.

2.A.4 Variables

1. Last midquote
2. Last price (VWAP for aggregated trades) excluding off-exchange transactions.
3. Time-weighted average (TWA) of $Q_{i,t}$ and $D_{i,t}$. The weights are calculated as the time that the relevant liquidity variable is available on the order book. TWA is good both to remove time-asyncronicity and to give more importance to liquidity available/valid for longer.
4. $V_{i,t}$ = Total trading volume is the sum of trade volumes in GBP
5. Trading volume of off-exchange trades (subtract from $V_{i,t}$ to obtain clean volume)
6. Trading volume of within-spread trades
7. Average transaction size of all trades, off-exchange and within-spread
8. Number of trades (after aggregation) of all three types
9. Trades duration: average time between trades
10. $TI_{i,t}$ = Trade imbalance, defined as buy volume minus sell volume (then this can be standardised according to the total volume at the aggregation frequency). It might be equal to zero in some minutes.
11. Simple average of $OI_{i,t}$. We do not take the TWA to give more weight to extreme observations that might indicate some temporary demand imbalance.
12. Simple average of $ES_{i,t}$ excluding off-exchange transactions.
13. Number of messages (any update of any quote)

Appendix 2.B Additional results

Table 2.B.1 presents daily trading activity and liquidity measures for 91 FTSE100 stocks that entered the index between January 2010 and December 2011, with at least 100 transactions each any day of the sample. Period Mar 2010 - Jul 2010

Table 2.B.1: Descriptive statistics (Mar 2010 - Jul 2010)

The table presents daily trading activity and liquidity measures for 91 FTSE100 stocks that entered the index between January 2010 and December 2011, with at least 100 transactions each any day of the sample. Period. All numbers are cross-sectional averages of time-series *Mean* or *Median*. Half depth and Half BO spread are time-weighted average; Effective spread is reported both as simple and volume-weighted average; hidden trades are defined as those executed within the spread; off-book trades are defined as single transactions, outside the spread, that do not respect the minimum tick restrictions. Sweep trades are aggregated in a single transaction.

	Big 10		All		Small 10	
	Mean	Median	Mean	Median	Mean	Median
Market cap.	87.45		21.47		2.51	
Number of trades	7765.21	6946.80	2543.36	2315.60	1146.19	1060.70
Total volume ('000 GBP)	138282.42	120009.17	34839.99	30395.75	6940.07	6048.41
Trade size (GBP)	17797.34	16768.72	11070.15	10217.83	6347.93	5685.51
Trade size, hidden	13079.26	11080.24	13639.61	8540.98	12369.59	6706.49
Trade size, off-book	160674.59	109374.51	133078.60	77845.28	75314.25	42952.48
Trade duration (sec)	11.16	10.97	39.36	37.91	55.32	53.08
Half depth (GBP)	59091.80	57056.79	46669.62	44751.80	15212.44	14813.63
Half BO spread (bp)	2.51	2.51	4.79	4.75	7.07	6.98
Eff. spread (bp)	2.00	1.99	3.65	3.63	5.15	5.08
VW Eff. spread (bp)	2.07	2.06	3.82	3.78	5.37	5.25
Realised volatility (bp)	173.23	166.96	164.70	155.84	200.29	191.51
Perc. hidden trades	5.15		4.50		4.40	
Perc. off-book trades	2.02		0.93		0.85	
Reporting delay (sec)	0		0		0	

Table 2.B.2 reports the hourly average of liquidity commonality together with a set of variables measuring high-frequency quoting, market conditions, trading behaviour and informational efficiency. We report the averages within each period.

Table 2.B.2: Hourly patterns in microstructure variables and liquidity commonality

This table reports Mar 2011 - Jul 2011

	8	9	10	11	12	13	14	15	16
BO spread									
Mean (bips)	5.83	4.84	4.50	4.40	4.41	4.37	4.17	3.92	3.54
Commonality (R^2)	30.48	30.71	28.68	26.46	26.37	28.51	29.71	30.61	24.08
Effective spread									
Mean (bips)	4.35	3.69	3.46	3.38	3.38	3.40	3.28	3.14	2.96
Commonality (R^2)	19.94	20.65	18.98	18.02	17.93	19.85	19.47	21.44	19.32
Half depth									
Mean (1000GBP)	67.43	81.00	86.78	88.96	87.90	92.08	100.24	110.63	138.58
Commonality (R^2)	16.17	15.38	15.39	15.93	14.37	17.14	19.41	22.13	21.81
High-frequency quoting									
Messages per GBP10000	17.57	23.66	25.09	28.18	32.15	26.10	19.10	14.06	7.93
Price impact									
Amihud (% per GBP1mil)	0.77	0.48	0.44	0.46	0.47	0.43	0.32	0.23	0.23
λ(order flow)	5.72	4.43	3.86	3.57	3.36	3.77	4.15	4.38	4.34
Trading variables									
Total volume (GBP1mil)	2.51	2.75	2.48	2.20	1.99	2.42	3.41	4.43	2.79
RV midquote (bips)	42.57	38.94	34.43	31.53	28.92	33.58	36.26	37.89	24.21
RV price (bips)	55.01	51.53	46.25	42.61	39.59	45.21	48.39	50.72	33.35
Order flow									
Size (std.)	0.80	0.81	0.82	0.83	0.84	0.82	0.79	0.75	0.72
Autocorrelation	0.03	0.05	0.05	0.05	0.04	0.03	0.03	0.02	0
Order Imbalance									
Size (std.)	0.30	0.32	0.32	0.33	0.33	0.33	0.30	0.28	0.26
Autocorrelation	0.41	0.46	0.47	0.49	0.50	0.48	0.44	0.40	0.35
Informational efficiency									
Returns autocorr. (abs.)	0.17	0.17	0.18	0.19	0.20	0.19	0.17	0.17	0.23
Variance ratio	0.30	0.26	0.26	0.26	0.26	0.27	0.25	0.25	0.33

Appendix 2.C Some comments on PCA

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a dataset consisting of a large number of interrelated variables, transforming them into a new set of variables, the principal components (PCs). The number of PCs should be small, though retaining much of the variation of the original data, and the PCs should be uncorrelated so that they are not redundant.

Consider $\boldsymbol{X} = (X_1, \ldots, X_n)'$, a vector of n random variables (RVs), and suppose that we are interested in studying the relationship between these variables. Unless n is small, it would not be very useful to look at all the $\frac{1}{2}n(n-1)$ covariance and n variances. Instead, it would be ideal to transform the variables such that the new variables retain much of the information contained in the covariances. The first step is to look for a linear combination, for example $\boldsymbol{\alpha}_1'\boldsymbol{X}$, of the elements of $\boldsymbol{X}$, having the maximum variance. The vector $\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_n)'$ is a weighting vector of finite coefficients that maximises:

$$Var(\boldsymbol{\alpha}_1'X) = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1, \tag{2.C.1}$$

where $\boldsymbol{\Sigma} \equiv Cov(\boldsymbol{X})$. Since it is not possible, for finite $\boldsymbol{\alpha}_1$, to find a solution to the quadratic form in (2.C.1), we need to impose a normalisation constraint on $\boldsymbol{\alpha}_1$. Different constraints can be imposed, but a classic one is $\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 1$, which means that $\boldsymbol{\alpha}_1$ has unit length. We can now express the problem as a constrained optimisation:

$$\begin{aligned} \max \quad & \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 \\ \text{s.t.} \quad & \boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 1 \end{aligned} \tag{2.C.2}$$

Assuming that $\boldsymbol{\Sigma}$ is known, the problem in (2.C.2) can be solved setting up the standard Lagrange multipliers and solving the first order condition (FOC):

$$\mathcal{L} = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda(\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 - 1) \tag{2.C.3}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}_1} = 0 \tag{2.C.4}$$

$$\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda\boldsymbol{\alpha}_1 = 0 \tag{2.C.5}$$

$$(\boldsymbol{\Sigma} - \lambda\mathbf{I}_n)\boldsymbol{\alpha}_1 = 0. \tag{2.C.6}$$

Thus, one can notice that, to satisfy the last two equations, λ must be an eigenvalue of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}_1$ is its corresponding eigenvector. To decide which of the n eigenvalues of $\boldsymbol{\Sigma}$ we should choose, notice that the quantity to maximise is $\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1'\lambda\boldsymbol{\alpha}_1 = \lambda$, by equation (2.C.5). Then, it is possible to conclude that

$\lambda = \lambda_1$ is the largest eigenvalue of $\boldsymbol{\Sigma}$ and it also corresponds to the variance of the first PC.

The second PC of $\boldsymbol{X}$ is the linear combination $\boldsymbol{\alpha}_2'\boldsymbol{X}$ that maximises $\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2$, with an additional constraint: we do not want PC2 to contain information that is already contained in PC1. This means PC2 must be orthogonal to PC1, i.e. $Cov(\boldsymbol{\alpha}_1'\boldsymbol{X}, \boldsymbol{\alpha}_2'\boldsymbol{X}) = 0$. But:

$$\begin{aligned} Cov(\boldsymbol{\alpha}_1'\boldsymbol{X}, \boldsymbol{\alpha}_2'\boldsymbol{X}) &= \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 && \text{(2.C.7)} \\ &= \boldsymbol{\alpha}_2'\lambda_1\boldsymbol{\alpha}_1 \quad \text{because of (2.C.5)} && \text{(2.C.8)} \\ &= \lambda_1\boldsymbol{\alpha}_2'\boldsymbol{\alpha}_1 && \text{(2.C.9)} \\ &= \lambda_1\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_2, && \text{(2.C.10)} \end{aligned}$$

so the orthogonality condition can be expressed using either of the formulas above. For simplicity, the last one is chosen and we write down the problem as the following constrained optimisation:

$$\begin{aligned} \max \quad & \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 && \text{(2.C.11)} \\ \text{s.t.} \quad & \boldsymbol{\alpha}_2'\boldsymbol{\alpha}_2 = 1 && \text{(2.C.12)} \\ & \boldsymbol{\alpha}_2'\boldsymbol{\alpha}_1 = 0, && \text{(2.C.13)} \end{aligned}$$

which is solved as above. Hence:

$$\begin{aligned} \mathcal{L} &= \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda(\boldsymbol{\alpha}_2'\boldsymbol{\alpha}_2 - 1) - \phi\boldsymbol{\alpha}_2'\boldsymbol{\alpha}_1 && \text{(2.C.14)} \\ \frac{\partial\mathcal{L}}{\partial\boldsymbol{\alpha}_2} &= 0 && \text{(2.C.15)} \\ 2\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - 2\lambda\boldsymbol{\alpha}_2 - \phi\boldsymbol{\alpha}_1 &= 0. && \text{(2.C.16)} \end{aligned}$$

Now we need to find the values of the Lagrange multipliers, λ and ϕ, that satisfy equation (2.C.16). Multiplying (2.C.16) by $\frac{1}{2}\boldsymbol{\alpha}_1'$ from the left gives:

$$\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_2 - \frac{1}{2}\phi\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 0, \tag{2.C.17}$$

which gives the solution $\phi = 0$, given constraints (2.C.12) - (2.C.13). Therefore, we are left with:

$$\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda\boldsymbol{\alpha}_2 = 0 \tag{2.C.18}$$

$$(\boldsymbol{\Sigma} - \lambda\mathbf{I}_n)\boldsymbol{\alpha}_2 = 0. \tag{2.C.19}$$

Thus, λ is again one of the n eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}_2$ is its corresponding eigenvector. Since maximising $Var(\boldsymbol{\alpha}_2'\boldsymbol{X})$ is the same as maximising λ, we are going to choose the largest eigenvalue that is not equal to λ_1, assuming that $\boldsymbol{\Sigma}$ does not have repeated eigenvalues. So, $\lambda = \lambda_2$, the second largest eigenvalue. The procedure can be generalised and the r-th PC of $\boldsymbol{X}$ is the linear combination $\boldsymbol{\alpha}_r'\boldsymbol{X}$, with $Var(\boldsymbol{\alpha}_r'\boldsymbol{X}) = \lambda_r$ and $\boldsymbol{\alpha}_r$ the corresponding eigenvector. Note that there exists other methods to derive the principal components, for example through geometric optimisation. See Jolliffe (2002) for more details.

In the classic derivation of PCA, it is assumed that the covariance matrix of $\boldsymbol{X}$, $\boldsymbol{\Sigma}$, is known and that it does not have repeated eigenvalues.

The equality of eigenvalues (and hence of variances of PCs) is more a theoretical than a practical problem, because it would be very unlikely that two or more stocks have exactly the same variance.

When $\boldsymbol{\Sigma}$ is unknown it is replaced with an unbiased covariance estimator, which makes some distributional assumptions (e.g. multivariate normality of $\boldsymbol{X}$). Furthermore, the main inferential results on eigenvalues rely on the independence (or weak dependence) of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ as well as multivariate normality. These asymptotic results show that the eigenvalues of the sample covariance matrix are consistent and asymptotically normal estimators of the population eigenvalues, at least when the data follow a multivariate normal distribution.

However, the derivation of PCA does not require the normality assumption. So, if the aim is not to make inference, PCA works well also with non-Gaussian data. Indeed, PCA is extremely efficient at representing multivariate normally distributed data, where uncorrelatedness implies independence. Thus, when data is non-gaussian, higher order statistics might need to be taken into account.

Chapter 3

The dynamics of systematic risk in a large equity portfolio▼

3.1 Introduction

The analysis of comovements between stock returns is at the heart of empirical asset pricing. Portfolio managers invest internationally [Heston and Rouwenhorst (1995), Bekaert et al. (2009)] and this requires the knowledge of relevant common factors and their importance for stock returns, both for risk management and for asset allocation purposes.

Searching for the most influential factors, various researchers have concluded that global, country- and region-specific factors are more important than industry ones in explaining the cross-section of expected returns [e.g., Heston and Rouwenhorst (1995), Griffin (2002), Bekaert et al. (2009), Fama and French (2017), Ando and Bai (2017)][1]. Nevertheless, factor loadings have been shown to vary

▼A joint research paper with Prof. Eric Hillebrand, Jakob Mikkelsen and Prof. Giovanni Urga entitled "Global comovements of stock returns using a multi-level factor model with time-varying parameters" is based on the results of this chapter. The paper has been presented at the "19th OxMetrics User Conference" (Paris, 11th September 2017), the "40th International Panel Data Conference" (Thessaloniki, 7th July 2017), the "18th OxMetrics User Conference" (London, 12th September 2016), and the "2016 PhD Research Days" (London, 9th June 2016).

[1]There is a small number of papers that argue that the relative influence of industry and country factors depends on the sample period. However, Bekaert et al. (2009) shows that the relevance of the industry factors was a short-lived phenomenon.

over time, making it difficult for the investor to estimate the riskiness of a firm and its cost of equity. For instance, Fama and French (1997) remark that "...there is strong variation through time in the CAPM and the three-factor risk loadings of industries..." and the problems are even stronger for individual firms and investment projects. Consequently, also the relative importance of global versus group specific factors is time-varying and difficult to estimate. However, so far the literature has assessed the time-varying relevance of global and regional factors using rolling window estimation [see, among others, Bekaert et al. (2009), Hirata and Otrok (2013)], whose results dependent on a nuisance parameter - the length of the window - and whose consistency has not been proven[2].

In the field of asset pricing, several researchers have estimated factor models with time-varying loadings but the results are limited to models with observable factors (CAPM and Fama-French extensions). At first, various authors have assumed that the factor loading is a function of economic state variables [Robichek and Cohn (1974), Shanken (1990), Rosenberg and Guy (1995), Ferson et al. (2002), Santos and Veronesi (2004)], leading to the derivation of the conditional CAPM. Intuitively, a firm in distress is more likely to report low earnings when the economy is in a bad state. However, Ghysels (1998) and Lewellen and Nagel (2006), among others, point out that the loading estimates are highly dependent on the assumed information set and, in case of misspecification, the unconditional CAPM works better. To avoid this drawback, a recent strand of literature makes use of non-parametric estimation to retrieve factor loadings from high-frequency data [e.g. Bollerslev and Zhang (2003), Patton and Verardo (2012) and references therein]. Unfortunately, non-parametric estimation disregards the parameters driving the dynamics of the loadings, which have been shown to contain important information for asset prices. For instance, Armstrong et al. (2013) develop a model where uncertainty about a firm's loading is negatively related to expected returns. The main limitation of this literature is that the analysis of factor loadings highly depends on the prior identification of the factors. In particular, because the factors in the asset pricing theory are unknown, the authors focus solely on the loading of the market factor.

[2]Recently, Gagliardini et al. (2016) prove the consistency of the two-pass estimation of a factor model with time-varying parameters that are function of stock specific and macroeconomic variables, as in Shanken (1990).

Thus, the main aim of this paper is twofold. First, we evaluate the contribution of unobserved global and regional factors to the overall variance in the portfolio. Second, we test whether the dynamics (uncertainty and persistence) of time-varying loadings are related to asset prices and investors.

To answer these questions, we use a panel of 2000 stock returns from six worlds regions from January 2006 to January 2016. We formulate a factor model where stock returns are assumed to be function of two types of factors: global (one observed financial factor and one latent non-financial factor) and region-specific (one latent factor per region). This modelling choice is motivated by Boivin and Ng (2006), who show that increasing N does not always help the estimation of the common factors when there is large cross-sectional correlation in groups of variables. Goyal et al. (2008) use the clustering of Nasdaq- and NYSE-listed stocks to identify common factors. Ando and Bai (2017) let the group membership of a stock be an unknown parameter to be estimated. The global financial factor in our model is the S&P500 Financials Index while the global and region latent factors are estimated with principal components analysis (PCA). Furthermore, in our model the loadings of each return on the factors vary over time, thus the model is an extension of Breitung and Eickmeier (2015). The principal component is a consistent estimator of the unknown factors in presence of time-varying loadings for large panels with $N, T \to \infty$: Bates et al. (2013) prove average consistency in t, while Mikkelsen et al. (2015) prove uniform consistency in t if $\frac{T}{N^2} \to 0$ is satisfied.

There is a great deal of empirical evidence in favour of modelling loadings as time-varying parameters. Stock and Watson (2009) find significant improvement in forecasting macro variables when coefficients are allowed to change after a structural break. Del Negro and Otrok (2008) and Eickmeier et al. (2015) estimate factor models with loadings modelled as genuine random walks using large panels of macro data. On the theoretical side, Bates et al. (2013) show that the principal component estimator remains consistent if the loadings are stationary, experience a structural break, or they are random walks of the type $\lambda_{it} = \lambda_{i,t-1} + T^{-3/4}\eta_{it}$. Recently, Mikkelsen et al. (2015) prove that the maximum likelihood estimator (MLE) of the parameters of autoregressive loadings processes is consistent. We apply this theory to specify and estimate our dynamic factor model. Andersen et al. (2006) find that stock beta is best approximated by a stationary $\mathrm{I}(0)$ process, due to a cancellation process in the ratio of covariance and market variance. Stationary fluctuations of factor loadings reconcile with theories of corporate finance. First, systematic risk is a decreasing function of investment because the level of investment increases with the availability of low risk projects [Berk et al. (1999), Cooper and Priestley (2011)], in alignment with pro-cyclical macroeconomic factors. Predictions from real options models [e.g. Carlson et al. (2006)] agree with this mechanism because undertaking a real investment can be considered as exercising a risky option, which makes systematic risk falling. Second, the beta of a firm increases around earning announcement [Savor and Wilson (2016)] and before Seasoned Equity Offerings [Carlson et al. (2010)], while gradually decreasing afterwards. Finally, entertaining takeover bids temporarily modifies the beta of a firm according to the difference with the beta of the target Hackbarth and Morellec (2008), because the new entity will have an average beta of the two firms.

Factor models can be either estimated on portfolios, as proposed by Fama and French (1997), or on individual stocks. Estimating the model stock by stock has two main advantages. First, it avoids the loss of information caused by grouping the stocks into portfolios when testing for the pricing of market anomalies, such as size or value [Ang et al. (2017), Gagliardini et al. (2016)]. Second, it allows to test if the dynamics of factor loadings (uncertainty and persistence) are related to expected returns. Lately, Armstrong et al. (2013) provide evidence of a negative relationship between factor loading uncertainty and future stock returns in a CAPM setting. We extend this to the case of global and region latent factors.

The empirical analysis suggests interesting findings. First, using canonical correlation analysis, we find that our estimated factors are linear combinations of Fama and French's market, value and size factors. Thus, our estimated factors correctly capture the risk at which firms are exposed. Second, we find that the relative importance of unobserved regional and global factors is time-varying, with the global one becoming more relevant when firms are exposed to global shocks. For instance, Energy stocks started to be more exposed to global shocks, both during the Great Financial Crisis (GFC) and from the beginning of 2015, due to the shocks to oil price. Third, the dynamics (persistence and variance) of the factor loading are related to the profile of a company. We find that bigger firms have larger exposure to financial and regional common factors, while there is no clear difference across global factor loading quantiles, supporting the result of Fama and French (2017). I.e. a global version of the factor model is not able to price the cross-section of stock returns. Expected returns are higher when the variance of financial and global factor loadings is large, while they are lower when the variance of the regional factor loading is large. This decreasing relationship is in line with the finding of Armstrong et al. (2013) for US stocks. However, our model suggests that there is a premium for holding stocks whose global systematic risk is very volatile. Finally, expected returns are decreasing in the persistence of financial and global factor loadings, implying that there is no premium for holding firms with highly persistent factor exposures.

The remainder of the paper is organised as follows. Section 3.2 presents the model and the estimation procedure. Section 3.3 describe our data base and the identification of the factors. Section 3.4 presents the estimation results. Section 3.5 compares our time-varying loading factor model with one with constant loadings. Section 3.6 uses our model to analyses the comovements between stock returns. Section 3.7 connects the loadings persistence and variance to the profile of the firm. Section 3.8 concludes.

3.2 Model

Factor models are useful to analyse comovements in equity markets [Bekaert et al. (2009), Bekaert et al. (2014)], to test for market integration [Flood and Rose (2005)], and for contagion across countries and asset classes [Dungey and Martin (2007), Belvisi et al. (2016)]. In this section, we introduce the two-level factor model with time varying loadings, and the estimation procedures for factor extraction and loading estimation via maximum likelihood estimation (MLE).

3.2.1 A two-level factor model with time-varying loadings

We have N stocks in total. We divide them into regions $R_1, R_2, \ldots, R_K$. Each region has n_k stocks, thus $\sum_{k=1}^{K} n_k = N$. The log-return $r_{i,t}$ on stock i in week t

is modelled as:

$$r_{i,t} = a_{i,t}O_t + b_{i,t}G_t + \sum_{k=1}^{K} c_{i,t}F_{k,t}\mathbb{1}_{\{i\in R_k\}} + e_{i,t}, \quad e_{i,t} \sim N(0,\psi_i) \tag{3.1}$$

for $i = 1,\dots,N$ and $t = 1,\dots,T$. O_t is an observable global factor, G_t is an unobservable (latent) global factor and $F_{k,t}$ is an unobservable factor specific to stocks in region R_k, for $k = 1,\dots,K$. The factor loadings vary over time according to the following specification:

$$\begin{aligned}
a_{i,t} &= (1-\phi_i^O)\bar{a}_i + \phi_n^O a_{i,t-1} + \eta_{i,t}^O, \quad \eta_{i,t}^O \sim iidN(0,q_i^O)\\
b_{i,t} &= (1-\phi_i^G)\bar{b}_i + \phi_n^G b_{i,t-1} + \eta_{i,t}^G, \quad \eta_{i,t}^G \sim iidN(0,q_i^G)\\
c_{i,t} &= (1-\phi_i^j)\bar{c}_i + \phi_n^j c_{i,t-1} + \eta_{i,t}^j, \quad \eta_{i,t}^j \sim iidN(0,q_i^j)
\end{aligned}$$

Note that O_t can be thought of as a known factor or it can be estimated using PCA. The model could easily accommodate multiple regional factors, but to keep the structure interpretable we assume that there is one factor for each region. After a sign identification, this allows us to have a multi-factor model with interpretable statistical factors (each region factor is the main regional driver of returns). The total number of factors in the model is denoted by $m = K + 2$.

The main innovation of the model is to let the loadings vary over time. We specify the dynamics of the factor loadings by stacking $a_{i,t}, b_{i,t}, c_{i,t}$ in the loading vector $\lambda_{i,t}$. The formulation in Eq. (3.1) implies a sparsity condition in the loading matrix, so that $\lambda_{i,t}$ is an m-dimensional vector that contains the same number of non-zero elements for all i. For instance $\lambda_{1t} = (a_{1t}, b_{1t}, c_{1t}, 0,\dots,0)'$ and $\lambda_{2t} = (a_{2t}, b_{2t}, 0, c_{2t}, 0,\dots)'$. The non-zero elements of $\lambda_{i,t}$ evolve as the following vector autoregression:

$$\lambda_{it} = (\mathrm{I} - \Phi_i)\bar{\lambda}_i + \Phi_i\lambda_{i,t-1} + \eta_{it} \tag{3.2}$$

where $\bar{\lambda}_i = E(\lambda_{it}) = (\bar{a}_i, \bar{b}_i, \bar{c}_i)'$ is the unconditional mean vector, $\Phi_i = diag(\phi_i^O, \phi_i^G, \phi_i^R)$ is the persistence parameter matrix and the characteristic roots of Eq. (3.2) lie outside the unit circle. $Q_i \equiv E(\eta_{it}\eta_{it}') = diag(q_i^O, q_i^G, q_i^R)$ is the covariance matrix of the innovations η_{it}, which is a Gaussian white noise process. Thus, the loadings of stock i on the three factors evolve as independent autoregressive (AR) processes of order one, around their respective unconditional means, $\bar{a}_i, \bar{b}_i$ and $\bar{c}_i$, with AR coefficient ϕ_i^f, $f \in \{O, G, R\}$, and condition $|\phi_i^f| < 1$ satisfied for all f. The higher ϕ_i^f the higher the weight of the factor loading at $t-1$ in determining the loading today and the lower the weight on its unconditional mean. Stationarity of the loadings on market factors has been demonstrated by, among others, Andersen et al. (2006) and Patton and Verardo (2012) and we extend this to the case of global and regional factors.

Furthermore, to simplify the exposition we cluster the model by region:

$$\begin{bmatrix} r_{1,\cdot t} \\ r_{2,\cdot t} \\ \vdots \\ r_{K,\cdot t} \end{bmatrix} = \begin{bmatrix} \mathcal{A}_{1t} \\ \mathcal{A}_{2t} \\ \vdots \\ \mathcal{A}_{Kt} \end{bmatrix} O_t + \begin{bmatrix} \mathcal{B}_{1t} & \mathcal{C}_{1t} & 0 & \cdots & 0 \\ \mathcal{B}_{2t} & 0 & \mathcal{C}_{2t} & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \mathcal{B}_{Kt} & 0 & \cdots & \cdots & \mathcal{C}_{Kt} \end{bmatrix} \begin{bmatrix} G_t \\ F_{1t} \\ F_{2t} \\ \vdots \\ F_{Kt} \end{bmatrix} + \begin{bmatrix} e_{1,\cdot t} \\ e_{2,\cdot t} \\ \vdots \\ e_{K,\cdot t} \end{bmatrix} \tag{3.3}$$

where $r_{k,\cdot t}$ is the vector of returns on the n_k stocks in region R_k and $k = 1, \dots, K$. $\mathcal{A}_{k,t}$, $\mathcal{B}_{k,t}$ and $\mathcal{C}_{k,t}$ are the $n_k \times 1$ vectors of loadings on the financial, global and regional factor, respectively. The peculiarity of the two-level structure is that stocks in region R_k are not influenced by shocks that are specific to other regions, which render the estimation of the factors more challenging than in the case where all stocks load on all factors. Breitung and Eickmeier (2015) and Wang (2010) show that regional and global factors can be disentangled by adding the sparsity as in Eq. (3.3). Finally, the model for the N stocks can be written in a more compact form as:

$$r_t = \mathcal{A}_t O_t + \mathcal{B}_t^* F_t + e_t, \tag{3.4}$$

or

$$r_t = \Lambda_t F_t^* + e_t, \tag{3.5}$$

where $r_t = (r'_{1t}, \dots, r'_{Kt})'$, $\Lambda_t = (\mathcal{A}_t, \mathcal{B}_t^*)$ and $F_t^* = (O_t, F_t)'$. The covariance matrix of idiosyncratic errors e_t is $\Psi_0 \equiv E(e_t e_t')$. To summarise, the N-dimensional vector of returns r_t is generated by $m \ll N$ global and region factors, time-varying factor loadings $\Lambda_t = (\lambda_{1t}, \ldots, \lambda_{it}, \ldots, \lambda_{Nt})'$ and normally distributed idiosyncratic errors $e_t = (e_{1t}, \ldots, e_{it}, \ldots, e_{Nt})'$.

With known factors, Specification (3.1) - (3.2) can be written and estimated in a linear state-space form as in Harvey (1990). With unknown factors, Mikkelsen et al. (2015) prove that the MLE of $\Phi_i, \bar{\lambda}_i, Q_i$ and ψ_i is consistent as $N, T \to \infty$. An alternative model specification is one where the loadings are static and the dynamics of the factors are estimated with the Kalman filter. However, since the Asset Pricing Theory assumes that stock prices are generated by a set of unpredictable factors, we follow the same rationale in the specification of our model. Furthermore, estimating stock-specific loadings dynamics allows one to sort stocks by loading variance or persistence, similarly to Armstrong et al. (2013). Estimating both the dynamics of the factors and the loadings is as possible extension of our model, but we leave this for future research[3].

In the next section, we present the procedure for the estimation of the model.

3.2.2 Estimation

The estimation procedure consists of two steps. First, global and regional factors are estimated using a sequential least squares estimator assuming constant loadings. Second, we treat the estimated factors as observed quantities in the likelihood function used to estimate the unknown loadings and variance parameters in Eqs. (3.1) - (3.2) via MLE.

Principal Component Estimation of the Latent Factors

The principal component estimator treats the loadings as constant over time and we use it to estimate the common factors in the following model:

$$y_t = \mathcal{B}^* F_t + u_t, \tag{3.6}$$

where y_t is the residual from the regression of r_t against O_t. Essentially, y_t is orthogonal to the observable factor and consequently the global and regional factors will not be influenced by the variation in O_t. This ensures that we can

[3]We thank Professor Catherine Doz for this suggestion.

disentangle two risk factors that affect all assets in the portfolio: the first is represented by the financial sector and the second is a non-financial latent global factor.

The estimated Principal Components (PC) are consistent in the presence of stationary fluctuations of the loadings around a constant mean. In particular, we make use of the results in both Bates et al. (2013), who prove the average convergence in t of the PCs to the true factor space and Mikkelsen et al. (2015), who prove also that the PCs uniformly converge in t when $\frac{T}{N^2} \to 0$.

The principal component estimator minimises the following sum of squared residuals:

$$S(F, \mathcal{B}^*) = \sum_{t=1}^{T} (y_t - \mathcal{B}^* F_t)'(y_t - \mathcal{B}^* F_t) \tag{3.7}$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{t=1}^{T} (y_{i,t} - b_i G_t - c_i F_{k,t} \mathbb{1}_{\{i \in R_k\}})^2. \tag{3.8}$$

Since both $\mathcal{B}^*$ and F_t are unobserved, we need to impose the following identifying restrictions to reach a unique solution.

IR1 $T^{-1} \sum_{t=1}^{T} G_t^2 = 1$ and $T^{-1} \sum_{t=1}^{T} F_{k,t}^2 = 1$.

IR2 $T^{-1} \sum_{t=1}^{T} F_{k,t} G_t' = 0$ for all j. This makes sure that regional factors are orthogonal to the global one.

IR3 $\sum_{t=1}^{T} F_{k,t} S_{k,t}' > 0$. This identifies the sign of the factors. Thus, F_k is normalised to ensure that $T^{-1} \sum_{t=1}^{T} F_{k,t} S_{k,t}' > 0$, where $S_{k,t}$ is the biggest country's stock market index return at time t in region k. This fixes the rotation indeterminacy and allows to interpret the sign of the factor loadings.

Wang (2010) shows that the restrictions IR1 and IR2 ensure that all parameters are identified, while IR3 is also used by Hirata and Otrok (2013) and Breitung and Eickmeier (2015). Note that we do not need to assume $T^{-1} \sum_{t=1}^{T} F_t F_t = \mathrm{I}_m$ as in standard factor analysis, thus the regional factors can be correlated across each other. The sparsity assumptions in Eq. (3.3) ensures that this does not create any multicollinearity issue. The following estimation algorithm follows from Breitung and Eickmeier (2015):

1. Initialise the global and regional factors in Eq. (3.6) with suitable values[4], $\widehat{F}_{t,(0)}$, and estimate the zero-stage loadings $\widehat{\mathcal{B}}^*_{(0)}$ from N time series OLS regressions of y_t on $\widehat{F}_{t,(0)}$.
2. The estimated loadings are then used as regressors in $y_t = F_{t,(1)}\widehat{\mathcal{B}}^*_{(0)} + \widetilde{u}_t$ to get the update of the factors at stage one, $\widehat{F}_{t,(1)}$, by OLS.
3. At stage s, update $F_{t,(s)}$ by regressing y_t on $\widehat{\mathcal{B}}^*_{(s-1)}$ and estimate $\mathcal{B}^*_{(s)}$ by regressing y_t on $\widehat{F}_{t,(s)}$.
4. Step 3 is repeated until convergence of $S(\widehat{F}_{(s)}, \widehat{\mathcal{B}}^*_{(s)})$ to a minimum.

Maximum Likelihood Estimation of the Time-varying Loadings

The last step is the estimation of Φ_i, λ_i, Q_i and ψ_i, for $i = 1, \ldots, N$, using the two-step maximum likelihood estimator proposed by Mikkelsen et al. (2015). The authors prove that the feasible likelihood function, which replaces the unobserved factors with PCs, convergences uniformly to the infeasible one, despite the presence of estimation error in the principal components and time-variation in the loadings.

In our model, the global and regional factors control for cross-sectional dependence in the returns, taking into account the fact that the market is partitioned in groups [Goyal et al. (2008)]. This allows to estimate the loadings parameters by a set on N univariate regressions where the latent factors are replaced by principal components. However, even if we do not fully capture the cross-sectional dependence in the errors, the MLE remains consistent. The framework is robust to the presence of temporal dependence in the errors with time-varying loadings capturing most of the volatility clustering. We refer to Mikkelsen et al. (2015) for details on the two-step estimation procedure.

Thus, conditional on the factors, r_i is uncorrelated among stocks and the likelihood can be analysed separately for each i. Thus, if r_i is the $T \times 1$ vector of time-series observations for stock i, we can write:

$$r_i = \widehat{\mathbf{F}}^* \Lambda_i + e_i, \tag{3.9}$$

where $\widehat{\mathbf{F}}^* = diag(\widehat{F}_1^{*\prime}, \ldots, \widehat{F}_T^{*\prime})$ is a $T \times mT$ block-diagonal matrix that stack to-

[4] We initialise the algorithm with the first PC of all stocks (for the global factor) and the first PC of each group of stocks clustered by region (for the regional factor).

gether the time series observations on the estimated factors, with diagonal elements $\widehat{F}_t^{*\prime} = (\widehat{O}_t, \widehat{G}_t, \widehat{F}_{1t}, \ldots, \widehat{F}_{Kt})$ representing the observations of each factor at time t. $\Lambda_i = (\lambda'_{i1}, \ldots, \lambda'_{iT})$ is a $Tm \times 1$ vector.

Assuming that the idiosyncratic errors are normally distributed, the likelihood function for r_i is Gaussian and, conditional on $\widehat{F}^* = (\widehat{F}_1^*, \ldots, \widehat{F}_T^*)'$, can be specified as follows:

$$\widehat{L}_T(r_i|\widehat{F}^*;\theta_i) = -\frac{1}{2}\log(2\pi) - \frac{1}{2T}\log|\Sigma_i| - \frac{1}{2T}(r_i - E(r_i))'\Sigma_i^{-1}(r_i - E(r_i)), \quad (3.10)$$

with parameter vector $\theta_i = \{\Phi_i, \lambda_i, Q_i, \psi_i\}$ for each i. $E(r_i) = (F_1^{*\prime}\lambda_i^0, \ldots, F_T^{*\prime}\lambda_i^0)$ is the $T \times 1$ mean vector of r_i and its covariance matrix is $\Sigma_i \equiv Var(r_i) = \widehat{\mathbf{F}}^* Var(\Lambda_i)\widehat{\mathbf{F}}^* + \psi_i I_T$. Eq. (3.10) represents the feasible likelihood because the factors $\widehat{F}^*$ are estimated. Finally, the maximum likelihood estimator of θ_i is:

$$\widehat{\theta}_i = \underset{\theta}{\text{argmax}}\, \widehat{\mathcal{L}}_T(r_i|\widehat{F}^*;\theta_i), \quad (3.11)$$

for each i. In practice, Eqs. (3.2) and (3.9) can be expressed as a linear state-space model and the likelihood is maximised via the Kalman filter. Theorem 1 in Mikkelsen et al. (2015) shows that $\widehat{\theta}_i \xrightarrow{p} \theta_i^0$, so the estimates are consistent[5] and converge in probability to their true values θ_i^0.

3.3 Data and identification

In this section we present the database, the data preparation and the descriptive statistics, together with some figures that outline the region and sector trends of equity markets between 2006 and 2016.

[5]The theoretical result is based on cross-sectional independence of errors and on the uniform consistency of the principal component. In our modelling framework, errors can be assumed to be uncorrelated because the group factors capture the cross-sectional correlation. Furthermore, uniform consistency in t of the principal components holds if $\frac{T}{N^2} \to 0$ and $\max_t ||F_t|| \leq M$. The first condition is satisfied because of the structure of our panel. We have that $N > T$, but also $\frac{T}{n_k^2} \to 0$ for all k. The second condition does not affect our results because we are not interested in estimating the parameters of the dynamics of F_t. See Mikkelsen et al. (2015) for more details.

3.3.1 Database Details

The sample period starts on Friday 13 January 2006 and ends on Friday 8 January 2016, a total of 521 weekly observations. This period contains various shocks that are interesting to analyse: the great financial crisis (GFC) of 2007-2008; the European sovereign debt crisis of 2011-2012; the Arab spring of 2011; and the oil shocks of 2015. Our universe include 1815 stocks that have been part of the main stock market index of 55 countries during the sample period and have complete time series of prices. Table 3.1 reports details of the number of stocks that survived our screening. The data is downloaded from Bloomberg and the universe resembles the one used by Bekaert et al. (2014). For each firm, we download the following variables: share price, number of shares outstanding, total assets and total debt. Prices refer to the last transaction of the week reported by the exchange, adjusted for subsequent splits but not for subsequent dividends. Working with weekly prices is especially convenient to avoid the problems caused by trading asynchronism of stocks listed in countries with different time zones. The balance sheet data is available at quarterly frequency.

To reduce the look-ahead bias we consider all stocks that entered the index during our sample period. This procedure limits the survivorship bias, which could be particularly severe in our sample period, considering the changes in composition of indexes that occurred in 2008. Table 3.1 reports detailed information on each stock market index.

[Table 3.1 about here.]

Prices are expressed in US dollars, and returns in the model are calculated as the first difference of the natural logarithm of the share prices. Each stock return is demeaned by subtracting the sample mean and scaled by dividing each series by the standard deviation. Before performing PCA, the data have been winsorised at the 99% level, while the dependent variable is not modified, so that outliers are captured by the time-varying loadings.

3.3.2 Observed factor

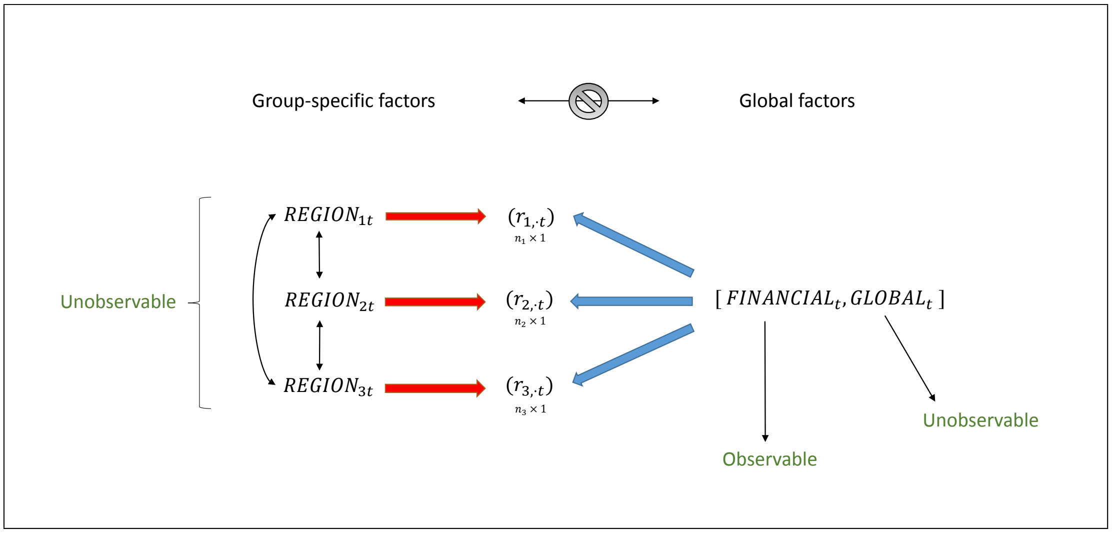
We define O_t as the US financial factor that captures a known source of risk that impacts all stocks in all regions. Other authors show that this factor plays an important role in explaining the cross-section of returns [e.g Bekaert et al. (2014)].

3.3.3 Region classification

There is evidence that shocks to share prices are more region specific than sector specific. Common currency, geographic proximity, similar stage of economic development or distribution of wealth are more resilient than sector membership. Heston and Rouwenhorst (1995), Griffin (2002), Bekaert et al. (2009) are some of the authors supporting the superiority of country factors. Recently, Ando and Bai (2017) let the group membership of a stock be an unknown parameter in the model and reach the same conclusion. We classify our universe of 2000 stocks into six geographical regions, in line with Hirata and Otrok (2013) and Breitung and Eickmeier (2015). The regions are: North America, Latin America, Asia-Pacific, Western Europe, Emerging Europe and Middle-East & Africa (MEA). This classification will be used to identify the region-specific factors in our model. The region composition is reported in Table 3.1. Furthermore, we aggregate the factor exposures by sector in order to understand if there is a sector-specific response (exposure) to the common factors. We use the following six sectors classified by Bloomberg: Basic Materials, Communications, Consumer Cyclical, Consumer Non-cyclical, Diversified, Energy, Financial, Industrial, Technology and Utilities.

Below we report a graphical representation of our model.

Visual representation of our model



3.3.4 Summary statistics

Table 3.2 report the summary statistics of simple returns (in %), market capitalisation, total assets and total debt of the 1815 companies that survived the cleaning procedures. Simple returns are non-Gaussian for all regions and sectors. In our empirical application, we use log-returns to make the distribution more similar to a Gaussian distribution. We report the average pair-wise Pearson correlation coefficient among the stocks in each group (region or sector), which gives a snapshot of the dependence between firms in each region and sector. Middle East & Africa has the lowest value, with a coefficient of 0.159, which can be expected given the economic diversity of this area. North America and Western Europe have a correlation of 0.378 and 0.424 respectively. The sector with the highest level of linear dependence is Energy, at 0.362. The balance sheet data are in line with expectations. For instance, North-American and Western European stocks have the highest average market capitalisation. The biggest companies that survived our screenings are: Apple (US); Vodafone (WestEur); GazProm (EmEur); China Petroleum (Asia). We exclude Financials from our analysis, given the peculiar nature of their balance sheet. Consequently, Energy stocks are the most capitalised, with a mean value of $25 billion and a median of $8 billion. Utilities and Energy stocks are the ones with larger assets and debt, in line with the infrastructures needed for the business.

[Table 3.2 about here.]

3.4 Estimation results

In this section, we present the estimated factors (one global and six regional), their corresponding time-varying loadings and we explore the benefits of allowing the loadings to vary over time in terms of model fit and misspecification.

3.4.1 Mapping the estimated factors to exogenous variables

Fig. 3.1 plots the estimated global factor and the regional factors for Asia-Pacific, Emerging Europe, Latin America, Middle-East Africa, North America and Western Europe. The factors are estimated by PCA from the model with static loading shown in Eq. (3.6) and they are rotated such that they are positively correlated with the stock market index of the biggest country in the region.

The global factor in Panel 3.1a captures a source of global risk uncorrelated with US Financials and its fluctuations are related to the overall trend of equity markets during our sample period. Together with the global factor, we plot a double-sided two-month moving average. Even after taking into account the shocks coming from US financial institutions, in 2008 the stocks of the major stock markets around the world experienced a substantial drop. This can be attributed to financial contagion [Bekaert et al. (2014)]. Other periods of consecutive negative returns can be seen in 2011 (around the sovereign-debt crisis) and between the end of 2014 and the beginning of 2015 (the period corresponding to a large drop in oil price).

Our model also allows to disentangle global and region-specific factors. The regional factors capture unknown shocks that affect only the firms in that region. We find that the impact of the Global Financial Crisis of 2008 was so pervasive that, even if the regional factors are orthogonal to the global factor and the observable US financial risk, in 2008 there were very large region-specific negative shocks in all six regions. However, the magnitude of the negative shocks in 2008 in the regional factors is smaller than the one of the global factor, and other shocks become more evident. This is an advantage of our model. All the regional factors display heteroscedasticity, in particular around periods of market turmoil. For instance, our estimated factors for Latin America and Emerging Europe experienced consecutive large negative shocks and excess volatility during the oil shocks of 2014-2015. Specifically, countries of Latin America with large oil reserves (e.g. Mexico) may have been particularly hit by the decrease in oil price that started at the end of 2014. Lower oil prices have a negative impact on oil producers and on countries whose GDP highly depend on oil exports, while they have a positive effect on companies whose costs depend on oil price (e.g. airlines) and for net importers countries. This could also explain the divergence of performance between Western Europe and Emerging Europe (which includes Russia) in the last part of the sample.

[Figure 3.1 about here.]

Table 3.3 reports, in Panel A, the correlation among the six regional factors. The sparsity of the loading matrix allows the estimated factors to be correlated between one another because they do not interact in the model. The North America factor is the most correlated with Latin America (which includes Mexico) and Western Europe, with a Pearson correlation coefficient of 0.429 and 0.483, respectively. Emerging Europe has the highest correlation with Latin America with a Pearson correlation coefficient of 0.441. Since the former region includes Russia and the latter Brazil, this connection may be due to the presence of large oil companies in the stock market indexes of these countries. The Middle-East Africa factor is uncorrelated with North America, Latin America and Western Europe, while it is mildly correlated with Asia-Pacific and Emerging-Europe, with a correlation coefficient of 0.210 and 0.131, respectively.

As a robustness check, in Panel B of Table 3.3 we report the Pearson correlation coefficients between the first principal component of the six regional portfolios and the global PC, the S&P500 index and S&P500 Financials index, respectively, where the factors are extracted separately from portfolios of stock returns in the relevant region. As expected, the first PC for North America correlates almost perfectly with the S&P500 index and the S&P500 Financials, with a correlation coefficient of 0.962 and 0.847, respectively. The firms based in Western Europe also closely mimic the pattern of the US stock markets, with correlations as high as 0.821 for the S&P500 index. The Middle-East&Africa factor is the least correlated with the fluctuations of the US stock market. All factors, except the MEA, are highly correlated with a global factor, with coefficients ranging from

0.889 to 0.968. This evidence supports the main feature of our model which allows to disentangle global and regional variation.

Finally, in Panel C of Table 3.3 we report the correlation coefficients between the estimated global and regional factors, the S&P500 index and S&P500 Financials index, without orthogonalising the dependent variables against the S&P500 Financials index. We notice that in this case there is a high correlation between Global and North America factors and both S&P500 and S&P Financials, which implies that the presence of contagion effects of shocks coming from the US stock markets. This makes the orthogonalisation with respect the S&P Financials a natural approach to disentangle the global risk from the US Financials risk.

[Table 3.3 about here.]

The connection with Fama-French factors

To understand what type of risk our factors are capturing, We now use canonical correlation analysis (CCA) to map the estimated factors to the three (market, value and size) Fama-French (FF) factors. CCA finds linear combinations of sets of random variables that correlate with one another maximally. For instance, if we have two vectors of random variables, $X = (x_1, \ldots, x_n)'$ and $Y = (y_1, \ldots, y_m)'$, CCA seeks the weights a and b such that $U = a'X$ and $V = b'Y$ maximise the Pearson correlation between U and V. $\{U, V\}$ is called the first pair of canonical variables. The second pair maximises the same correlation with the constraint of being uncorrelated with the first pair, and so on, until $\min\{m, n\}$ pairs. The canonical correlations are estimated using the time series data on $x_{i,t}$ and $y_{j,t}$. Thus, CCA is a useful tool to map the estimated principal components with observable factors.

Table 3.4 reports the (squared) maximum canonical correlation between the linear combinations of estimated factors and FF market, value and size factors. Panel A reports the canonical correlation between all three factors, i.e. $[O_t, G_t, F_{j,t}]$, for $j = 1, \ldots, R$, while Panel B reports the results using only the regional factors. We use five sets of FF region-specific factors: Asia-Pacific (excluding Japan), Europe, Global, Global excluding US, and North America. Overall, we find that the spaces spanned by the FF factors and by the factor estimated in this paper are very similar. This leads us to conclude that our model is correctly capturing the risk to which firms are exposed. In particular, our three factors for Asia-Pacific, Western Europe and North America correctly identify the FF factors constructed with stocks in the respective regions, with canonical correlations up to 0.958 in North America's case. Only for Europe's case, the FF factors have a comparable canonical correlation with both our Western Europe and North America estimated factors, as reported in Panel B. This is possibly due to the integration between Europe and the US.

[Table 3.4 about here.]

3.4.2 Factor loadings

Our model allows to estimate, for each stock, the parameters (mean, variance and persistence) driving the dynamics of the stock's exposure to each factor. In this section, we present the results for two selected stocks, and for the aggregation by either region or sector.

The factor exposure of two large firms

Fig. 3.2 reports the estimated time series of the three factor loadings for two large firms: IBM and Tenaris. We choose to individually analyse these firms because they are likely to be exposed to global determinants that are difficult to quantify and our model can provide guidance in their identification.

Figs 3.2a - 3.2c plot the financial, global and regional factor loadings for IBM, respectively. On average, this firm is most exposed to financial risk, with a factor loading fluctuating around a level of 0.6[6]. The global and regional factor loadings fluctuate around a similar long-run mean of about 0.3 but different AR(1) parameters (0.5 for the global and -0.22 for the regional factors, respectively). A large AR(1) implies that the process spends a long time away from the long-run average.

Figs 3.2d - 3.2f plot the financial, global and regional factor loadings for Tenaris, respectively. Tenaris is a global company, headquartered in Luxem-

[6]Since the dependent variable is standardised, the factor loadings' economic magnitude corresponds to β standard deviations for one standard deviation increase in the factor.

bourg and with business in over 20 countries. It deals with the construction, distribution and service of steel pipes. Our model correctly identifies this company as highly exposed to both observed and unobserved global factors. However, these exposures have very different dynamics. On the one hand, the exposure to the financial factor is constant around a value of 0.537 with a negligible variance that makes the AR(1) not identifiable. On the other hand, the exposure to the global factor follows an AR(1) coefficient of 0.94 close to the unit root case suggesting a persistent exposure of Tenaris to global shocks.

Our factor loadings follow closely the idiosyncratic variations of stock returns and identify some firm-specific events that the factors are not able to capture. For instance, the financial factor in Fig. 3.2a shows some negative spikes which appear regularly in October from 2011 to 2015. These can be caused by either third quarter earning announcements or the expiration of stock options. We can rule out that this could be caused by dividend payments, as these happened quarterly in that period of time. Furthermore, the regional factor in Fig. 3.2c exhibits a large drop in the third week of October 2014, which corresponds to the announcement by IBM of a large fall in sales, [7] while the stock market was rallying upwards. In that moment, the covariance between the market and IBM return switched from positive to negative. These events caused the loading to become temporarily negative, before reverting to its long run mean.

[Figure 3.2 about here.]

Aggregate results

Table 3.5 reports the average magnitude of the factor loadings, their persistence (AR(1) parameter) and their volatilities, aggregated by either region or sector. The loading magnitude is estimated via OLS from a static loading model, while the AR(1) parameter and the variance are estimated via maximum likelihood estimation from Eq. (3.1). The table also reports the percentage of stocks whose loadings vary so little that they are indistinguishable from the OLS ones. We set a loading volatility threshold to 0.01[8], under which it is very difficult to iden-

[7] Source: "IBM shares tumble as profits and sales fall", Financial Times.

[8] We consider volatilities smaller than this threshold numerically undistinguishable from zero.

tify the autoregressive parameter, and consider such loadings static. Assuming that the goal of an equity investor is to estimate the systematic risk of a set of stocks correctly, factor loadings with large AR(1) coefficients and high variance indicate stocks with very persistent exposure to factors.

First, we analyse the results by regions. The firms listed in North America are the most exposed to the financial factor. This result might be due to some endogeneity since some of the constituents of the S&P500 Financials are also in our portfolio (19 firms). The second most exposed group is Western Europe, with an average factor loading of 0.461. These firms, together with firms listed in Latin America and Emerging Europe are relatively more exposed to financial shocks than to global or regional shocks. Conversely, firms listed in MEA are more exposed to regional than global shocks. Firms listed in the Asia-Pacific region are the least exposed to the US financial risk. We report the percentage of firms, within each group, with an AR(1) parameter larger than 0.5, defining them as firms with very persistent factor loadings. Table 3.5 shows that the regional factor loadings are the most persistent, implying higher predictability of regional systematic risk compared to financial and global risk, in particular for Asia-Pacific and MEA.

Second, we analyse the results by sectors. As to be expected, financial firms are the most exposed to the financial factor. However, we do not find much variation across sectors. Utilities stocks are the least exposed to the financial factor. Energy stocks are the most exposed and Consumer Non-cyclical are the least exposed to the estimated global factor. This is in line with expectations, since Consumer Non-Cyclical stocks (e.g. food, beverages and tobacco) should not follow global trends.

[Table 3.5 about here.]

3.5 The benefits of using time-varying factor loadings

In this section, we compare the residuals obtained from a multi-level factor model with static loadings with those obtained from our model. First, we provide evidence that our model has an unambiguously higher goodness of fit. We also analyse the significant deviations between our factor loadings and the static ones estimated by OLS. Second, we show that allowing the loadings to be time-varying has an impact on standard residual-based misspecification tests and on the estimation of the number of factors following Bai and Ng (2002).

3.5.1 Model fit

The model specification in Eq. (3.5) nests the two-level factor model with static loadings as special case, where it is assumed that $\Lambda \equiv \Lambda_t$. The static specification for the N stocks takes the form:

$$r_t = \Lambda F_t^* + e_t, \tag{3.1}$$

where Λ is an $N \times m$ matrix of parameters and F_t^* contains m factors, some global and some region-specific. The model in Eq. (3.1) has been used by Breitung and Eickmeier (2015) to study the comovements of real economy variables and by Goyal et al. (2008) to estimate NYSE- and NASDAQ-specific factors for stock returns.

Panel A of Table 3.6 reports the goodness of fit of specifications (3.5) and (3.1), measured by the R^2 coefficient of the regressions averaged within regions and sectors. We choose this method over a joint likelihood-ratio test of the overall fit because we want to unveil in which regions/sectors the time-varying loading model provides the largest improvements. We find that there is an improvement of around 25% by using time-varying parameters. The biggest improvement, of 28%, is for the firms listed in Middle East & Africa, which are the ones where the idiosyncratic component plays a prominent role. As showed in Table 3.2, this group of stocks has the lowest average pair-wise correlation, which is caused by the economic and political differences of these countries that might have limited their economic integration. Regions where the OLS estimator provided the highest average R^2 are the ones that are more integrated (North America and Western Europe). Thus, the time-varying specification can be viewed as a bridge between the common factors and the idiosyncratic characteristics of each stocks.

Furthermore, we can compare the distribution of $\lambda_{i,t}$ with the theoretical distribution of the OLS estimator of λ_i in Eq. (3.1). Since the least squares estimator is normally distributed, we expect 5% of the observations ($\lfloor T(0.05) \rfloor = 26$) to lie outside the 95% confidence interval (CI). Let us define the number of times the factor loadings a_i, b_i and c_i are outside the 95% confidence interval of the OLS estimator as:

$$n_i^F = \sum_{t=1}^{T} \mathbb{1}_{\left(a_{i,t} \notin \{a_i \pm 1.96\sqrt{Var(a_i)}\}\right)} \tag{3.2}$$

$$n_i^G = \sum_{t=1}^{T} \mathbb{1}_{\left(b_{i,t} \notin \{b_i \pm 1.96\sqrt{Var(b_i)}\}\right)} \tag{3.3}$$

$$n_i^R = \sum_{t=1}^{T} \mathbb{1}_{\left(c_{i,t} \notin \{c_i \pm 1.96\sqrt{Var(c_i)}\}\right)} \tag{3.4}$$

for $i = 1, \ldots, N$ stocks. We then average n_i^F, n_i^G and n_i^R by region and sector and report the results in Panel B of Table 3.6. We find that in all groups of stocks there is a large number of significant deviations from OLS.

[Table 3.6 about here.]

Since we estimate the loadings conditionally at each time t, we can also identify in which year and for which factors there are more deviations from the OLS estimates, i.e. which is the time when a static specification is mostly underestimating or overestimating the true factor exposure. Fig. 3.3 reports, for every year, the cross-sectional average of the number of significant deviations from OLS in the six world regions.

We find that in North America, in 2008, the exposure of the average firm to financial risk would have been underestimated or overestimated 25 out of 52 weeks, if it was estimated assuming constant loadings. In 2009, the number is very similar and the same can be stated for stocks listed in Western Europe, Emerging Europe and Asia-Pacific. Even though the Great Financial Crisis is affecting the estimation of the factor loadings of all stocks, in the Middle-East&Africa and Latin America, the estimation of the regional factor loadings is the most affected.

[Figure 3.3 about here.]

3.5.2 Misspecification tests

In this section we explore the effect of not accounting for time-varying factor loadings on the Bai and Ng (2002) estimator of the number of factors and on various misspecification tests, namely residual heteroscedasticity tests and serial correlation tests.

First, we estimate the number of factors using the Bai and Ng (2002) criteria on the returns matrix, the residuals from a static model and the residuals from a time-varying loading model. Bai and Ng (2002) propose various modifications of information criteria for model selection with an additional penalty that is a function of both N and T. We use

$$IC_{p1}(k) = \ln(V(k, \widetilde{F}^k)) + k\left(\frac{N+T}{NT}\right)\ln\left(\frac{N+T}{NT}\right) \tag{3.5}$$

where $V(k, \widetilde{F}^k)$ is the average residual variance of a factor model with k factors. $IC_{p1}(k)$ can be used for estimating the number of factors as a standard information criterion:

$$\hat{k} = \underset{0 \leq k \leq kmax}{\arg\min}\; PC_1(k) \tag{3.6}$$

Second, we test the null of homoscedasticity in the residuals using a White-type test, estimating the following auxiliary regression:

$$\hat{e}_{i,t}^2 = \alpha_i + \gamma_i \hat{F}_{i,t}^{*2} + u_{i,t}, \quad u_{i,t} \sim N(0, \sigma_{u,i}^2) \tag{3.7}$$

for $i = 1\ldots, N$ stocks, where $\hat{F}_{i,t}^{*}$ contains the three factors specific to stock i, estimated by PCA. The test statistics is equal to the R^2, times the sample size T, and it is distributed as a χ^2 with degrees of freedom equal to the number of factors. The properties of the test are studied by Mikkelsen (2017), who proves that this test is also equal to testing for constant loadings. Finally, we test the null hypothesis of no serial correlation of the error term $\hat{e}_{i,t}^2$ up to the p-th lag using the Breusch-Godfrey test (up to lags two and five). The test statistics is equal to R^2 times the sample size and it is distributed as a χ^2 with degrees of freedom equal to $T - p$.

Table 3.7 reports the results. The Bai and Ng (2002)'s IC_{p1} criterion finds that ten factors are needed to describe the variation in our panel of 1815 firms from 50 countries. Given that we include eight factors, we expect the residuals to have two omitted factors. Instead, the Bai and Ng (2002) criterion suggests to use five factors and only when allowing the factors to vary over time this number decreases to three. Thus, the structural instability of the loadings has an influence on the Bai and Ng (2002) number of factors estimator.

For the heteroscedasticity and serial correlation tests, in Table 3.7 we report the percentage of stocks for which we reject the null at 99% confidence level. We find that 51% of firms have time-varying loadings, which implies that the volatility of returns is not entirely captured by the static-loadings model. When loadings are allowed to change over time, only 5% of stock returns have time-varying volatility. The results on the serial correlation tests are less strong. The percentage of firms that exhibits residual serial correlation up to lag five is reduced from 31% to 21%.

[Table 3.7 about here.]

3.6 Global comovements of stock returns: new evidence

In this section we use variance decompositions methods to analyse the comovements of a large panel of stock returns. Variance decomposition has been extensively used to interpret the estimates of factor models. Among others, Breitung and Eickmeier (2015) assess the degree of comovement of groups of variables. In particular, the higher the (average) share of variance explained by common factors compared to the idiosyncratic variance, the higher the comovements. However, this method provides one number for the whole sample and often two or more sample periods are compared to assess whether there was an increase in commonality [see, e.g., Hirata and Otrok (2013)]. Our model allows to overcome this limitation, estimating a conditional variance decomposition at each time t. First, we calculate the traditional variance decomposition. Second, we introduce a time-varying variance decomposition that allows us to assess the importance of the risk factors of our model at every point in time.

3.6.1 Static variance decomposition

Table 3.8 reports the static variance decomposition. The firms listed in North America and Western Europe have the highest commonality, while the ones listed in the MEA region have the largest idiosyncratic component. This is in line with the results in Table 3.5.

[Table 3.8 about here.]

3.6.2 Time-varying variance decomposition

Our specification allows one to calculate the share of variance explained by the factors at each point in time. Hence, we can capture possible shifts in the importance of some factors and connect them to macro events. The variance of the returns on stock i at time t, conditional on the estimated factor loadings, can be written as:

$$\mathrm{var}_t(r_{i,t} \,|\, \hat{\lambda}_{i,t}) = \hat{a}_{i,t}^2 \,\mathrm{var}(O_t) + \hat{b}_{i,t}^2 \,\mathrm{var}(G_t) + \hat{c}_{i,t}^2 \,\mathrm{var}(F_{k,t}) \mathbb{1}_{\{i \in R_k\}}, \tag{3.1}$$

for $i = 1, \ldots, N$, $t = 1, \ldots, T$, and assuming that the factors and the errors are conditionally orthogonal. O_t is an observable global factor, G_t is an unobservable (latent) global factor and $F_{k,t}$ is an unobservable factor specific to stocks in region R_k, for $k = 1, \ldots, K$. All these factors have unconditional variance equal to one. Thus, if we are interested in the share of variance explained by each of the factors at each time t, we can define the following quantities:

$$FV_{i,t} = \frac{\hat{a}_{i,t}^2}{\text{var}_t(r_{i,t} \mid \hat{\lambda}_{i,t})} \qquad \text{(Financial)},$$
$$GV_{i,t} = \frac{\hat{b}_{i,t}^2}{\text{var}_t(r_{i,t} \mid \hat{\lambda}_{i,t})} \qquad \text{(Global)},$$
$$RV_{i,t} = \frac{\hat{c}_{i,t}^2}{\text{var}_t(r_{i,t} \mid \hat{\lambda}_{i,t})} \qquad \text{(Regional)},$$

where $FV_{i,t}$ is the share of variance explained by the financial factor at time t, and $GV_{i,t}$ and $RV_{i,t}$ are defined accordingly. Calculating the cross-sectional average of the quantities above provides a measure of the importance of different drivers for the comovements of groups of stocks (e.g. inside one region, one sector, one country).

Fig. 3.4 shows the share of variance explained by each factor, averaged across all N stocks in the portfolio. The graphs shows that, on average, the financial factor is the most pervasive, with a share of variance explained around 15%. However, during the GFC, there was a considerable increase in the exposure of stocks to financial shocks, which corroborates the evidence of contagion from the financial sector to other areas of the economy. Regional and sectoral figures shed more light on the heterogeneity of this effect. Furthermore, Fig. 3.4b shows that the share of variance explained by all the common factors increased by 10% by the end of 2008.

[Figure 3.4 about here.]

Figs. 3.5 - 3.6 plot the share of variance explained by the three factors for each region. For the stocks in all regions, the contribution of all factors increases during the Financial Crisis, hence the comovements increase. The increase in comovement at the outset of the crisis varies across regions and, for instance in Asia-Pacific and Western Europe, some stocks start to become more sensitive to global shocks already at the end of 2007 or the beginning of 2008, We find evidence of increased comovements during the European sovereign-debt crisis, recording positive spikes in the financial factor contribution in 2011 for Western

Europe and for Emerging Europe. In Middle East & Africa, the regional factor is the most important throughout the sample.

[Figure 3.5 about here.]

[Figure 3.6 about here.]

Figs. 3.7 - 3.9 plot the share of variance explained by the three factors in each sector. The most interesting cases are the ones where the variance shares increase during the sample. For example, the variance of the stocks in the Energy sector start to be highly explained by the global factor from the beginning of 2015. This is the direct effect of shocks from the oil market: Energy companies were hit by a very low price of oil, due to over production, shale-gas substitute product, and a reduced demand from China. In addition, the Utilities and Basic Materials sectors experience a large increase in the importance of global factor from 2015. As expected, these firms, providing for instance gas and electricity, are subject to demand shocks that are specific to their area and they are also very sensitive to interest rate changes due a high debt/equity ratio.

[Figure 3.7 about here.]

[Figure 3.8 about here.]

[Figure 3.9 about here.]

3.7 The connection with the profile of the firm

In this section we investigate if the firm-specific estimates that we obtain from our model - in particular loading persistence and variance - can identify different types of firm. This research question is motivated by Ang et al. (2017), who show that creating portfolios for asset pricing tests destroys information and leads to larger standard errors than using individual stocks. Our goal is to test whether the dynamics of the beta of the factors is related to the size and the leverage of the firm and to its expected returns. This information, gathered only from market prices, could be used in asset allocation models.

3.7.1 Size effect

Fig. 3.10 reports the median market capitalisation (at the end of the sample) by loading variance, persistence and magnitude quantiles. The figure is composed of three panels and each one reports three sets of bars, one for each factor. Since our universe of securities comprises the constituents of large stock market indexes, we do not expect this signal to be strong for every quantile.

We find that bigger firms have larger exposure to financial and regional common factors, while there is no clear difference across global factor loading quantiles [Fig. 3.10c]. For instance, the firms in the bottom quantile of financial factor loadings have a median market capitalisation of $1 billion, while the ones in the top quantile have 15$ billion. This evidence is in line with the finding of Fama and French (2017) that a global version of their factor model is not able to price the cross-section of stock returns. Furthermore, we find that stocks with loading uncertainty, approximated by loading variance, tend to be larger (from two to three times bigger) than firms with little variation of factor loadings. In Fig. 3.10a we can see that the effect exists for the financial and global factors only. This is an extension of Armstrong et al. (2013), who also analyse the cross-section of firms with loading uncertainty, but using a single-factor CAPM with US stocks. There is no difference in size between firms with difference persistence parameters [Fig. 3.10b].

In conclusion, we find that large firms tend to have large exposures to US financial and regional factors, and these exposures are also more volatile than small firms.

[Figure 3.10 about here.]

3.7.2 Leverage effect

Fig. 3.11 reports the expected change in the leverage ratio from January 2010 and January 2016, by loading variance, persistence and magnitude quantiles. Each panel reports three sets of bars, one for each of the factors. Financial stocks are excluded. We use the average of the change in leverage ratio (at quarterly frequency) instead of the leverage ratio because various authors have shown that firms adjust their leverage towards a target ratio [see, e.g. Halling et al.

(2016)], and we find that the average leverage does not vary substantially across stocks.

We find that leverage is increasing in the variance of the regional factor loading, which implies that firms with the highest leverage have a higher loading uncertainty. Fig. 3.11c shows that the firms with high leverage have a low exposure to the regional factor and a relatively high exposure to the financial factor. From Fig. 3.10c we know that these are small firms. This result is consistent with a leveraged firm that is exposed to regional shocks such as interest rates shocks.

[Figure 3.11 about here.]

3.7.3 Expected returns

Fig. 3.12 reports the expected weekly returns, expressed in basis points, as a function of loading variance, persistence and magnitude. Expected returns are the average log-return for each stock in the universe, excluding Financials, from January 2010 until January 2016.

We find that expected returns are increasing in the variance of financial and global factor loadings, while they are decreasing in the variance of the regional factor loading [Fig. 3.12a]. The decreasing relationship is in line with the finding of Armstrong et al. (2013) for US stocks. However, our model suggests that there is a premium for holding stocks with large variance in the exposure in the global factor. This pattern cannot be explained by cross-sectional differences in returns volatility. Furthermore, expected returns are decreasing in the persistence of financial and global factor loadings, implying that there is no premium for holding firms with highly persistent factor exposures.

[Figure 3.12 about here.]

3.8 Conclusions

In this paper, we studied the dynamics of the systematic risk in a large portfolio of firms from 54 countries from January 2006 to January 2016. We proposed a two-level factor model with time-varying loadings that captures financial, global and regional risk to estimate common components in stock returns. The global and regional factors are latent and estimated via principal component analysis. The loadings evolve as autoregressive processes and are estimated via maximum likelihood.

Our analysis yields three main findings. First, we find that our estimated factors are linear combinations of Fama and French's market, value and size factors. Thus, we are able to capture the same source of risk. Second, we find that the relative importance of unobserved regional and global factors is time-varying, with the global factor becoming more relevant when firms are exposed to global shocks. For instance, Energy stocks were exposed to global shocks, both during the Great Financial Crisis (GFC) and from the beginning of 2015. Finally, the dynamics of the factor loading are related to the profile of a company. In line with Armstrong et al. (2013), we find that expected returns are lower when the variance of the regional factor loading is large. However, we also find that they are higher when the variance of financial and global factor loadings is large. Furthermore, our model suggests that there is a premium for holding stocks whose global systematic risk is more volatile. Finally, expected returns are decreasing in the persistence of financial and global factor loadings, implying that there is no premium for holding firms with highly persistent factor exposures.

Table 3.1: Universe of securities

This table reports the countries that make up each region, and for each country it reports the following variables: *Ticker* is the Bloomberg ticker that identifies the stock market index; *#Stocks* is the number of companies that became members of the index during the sample period, from 10 January 2003 to 19 May 2017; *Avg.Active* is the average number of index members at the beginning of every month in the sample period; *Full* is the number of stocks with complete price time series - missing values are filled with the previous value as long as there are no more than four consecutive missing; *Jan06-Jan16* is the number of complete price series, when restricting the sample from 13 January 2006 to 8 January 2016.

					#Complete	
Region	Country	Ticker	#Stocks	Avg.Active	Full	Jan06-Jan16
North America	Canada	SPTSX60 Index	101	60	61	65
North America	US	OEX Index	167	100	120	123
Latin America	Mexico	MEXBOL Index	72	35	23	33
Latin America	Argentina	MERVAL Index	53	16	32	34
Latin America	Brazil	IBOV Index	129	66	43	56
Latin America	Chile	IPSA Index	71	40	41	55
Latin America	Peru	SPBLPGPT Index	89	33	20	27
Latin America	Venezuela	IBVC Index	18	15	0	12
Asia-Pacific	Japan	TPXL70 Index	125	70	96	101
Asia-Pacific	China	SSE50 Index	151	50	24	33
Asia-Pacific	HongKong	HSCEI Index	83	40	28	48
Asia-Pacific	India	SENSEX Index	48	30	34	39
Asia-Pacific	Indonesia	LQ45 Index	122	45	46	59
Asia-Pacific	Korea	KOSPI50 Index	86	50	49	57
Asia-Pacific	Taiwan	TW50 Index	83	50	66	73
Asia-Pacific	Thailand	SET50 Index	110	50	61	75
Asia-Pacific	NewZealand	NZSX15G Index	33	15	16	20
Asia-Pacific	Australia	AS31 Index	95	50	49	57
Western Europe	Austria	ATX Index	39	20	23	24
Western Europe	Belgium	BEL20 Index	37	20	20	26
Western Europe	Denmark	KFX Index	34	20	24	27
Western Europe	Finland	HEX25 Index	37	25	23	26
Western Europe	France	CAC Index	63	40	43	47
Western Europe	Germany	DAX Index	46	30	36	38
Western Europe	Ireland	ISEQ Index	91	53	15	22
Western Europe	Luxembourgh	LUXXX Index	21	10	5	5
Western Europe	Netherlands	AEX Index	47	20	20	23
Western Europe	Norway	OBX Index	67	26	22	34
Western Europe	Portugal	PSI20 Index	38	20	20	22
Western Europe	Spain	IBEX Index	61	35	30	33
Western Europe	Sweden	OMX Index	41	30	35	35
Western Europe	Switzerland	SMI Index	34	21	26	26
Western Europe	UK	UKX Index	203	102	110	133
Emerging Europe	Croatia	CRO Index	61	26	7	21
Emerging Europe	CzechRepublic	CCTX Index	14	8	6	7
Emerging Europe	Estonia	TALSE Index	25	16	5	7
Emerging Europe	Hungary	BUX Index	28	13	11	14
Emerging Europe	Latvia	RIGSE Index	44	29	7	9
Emerging Europe	Malta	MALTEX Index	26	18	4	4
Emerging Europe	Lithuania	VILSE Index	46	29	9	13
Emerging Europe	Poland	WIG20 Index	46	20	20	32
Emerging Europe	Romania	ROTXEUR Index	24	12	3	6
Emerging Europe	Russia	CRTX Index	49	16	5	7
Emerging Europe	Serbia	BELEX15 Index	26	15	0	15
Emerging Europe	Turkey	XU030 Index	77	30	44	51
Emerging Europe	Ukraine	PFTS Index	27	19	0	7
MEA	Egypt	HERMES Index	80	39	42	55
MEA	Qatar	DSM Index	38	20	2	25
MEA	UAE	ADSMI Index	69	61	11	19
MEA	Morocco	MOSEMDX Index	81	51	27	35
Total			3256	1709	1464	1815

Table 3.2: Descriptive statistics

This table reports the summary statistics for the 1815 companies that survived the cleaning procedures. The original number of firms in each country is reported in Table 3.1. Panel A reports the cross-sectional average of the summary statistics of simple returns. Panel B reports average market capitalisation, total assets and debt. *Min* and *Max* are the minimum and maximum over time and across all stocks in a group (i.e. the absolute min and max). The remaining statistics are N-averages of the relevant coefficient: *Mean*, *Med* and are the cross-sectional averages of mean and median; *StDev, Skw* and *Krt* are the average standard deviation, skewness and kurtosis; $\rho(1)$ is the OLS estimate of the first autocorrelation coefficient; *ADF* is the statistics for the Augmented Dickey-Fuller test, which is run with a constant, time trend and one lag. The critical value at 95% significance is -3.41 and the null hypothesis is "the series contains a unit root". Finally, *Pearson* is the average pair-wise correlation of the stocks in the relevant group.

Panel A: Stock returns

	Mean	Med	Min	Max	StDev	Skw	Krt	$\rho(1)$	ADF	Pearson	#
Returns(%)											
North America	0.202	0.223	-35.903	45.019	4.298	-0.007	4.440	-0.057	-16.535	0.378	188
Latin America	0.275	0.082	-28.376	48.919	5.353	0.311	4.671	-0.023	-15.923	0.265	217
Asia-Pacific	0.265	0.123	-35.674	42.507	5.248	0.181	4.131	-0.027	-16.037	0.237	562
Western Europe	0.186	0.213	-58.848	55.163	5.087	-0.020	4.369	-0.046	-16.590	0.424	521
Emerging Europe	0.099	0.029	-31.425	39.595	5.863	0.169	4.697	0.004	-15.360	0.269	193
MEA	0.143	-0.065	-25.218	38.627	4.997	0.345	5.247	-0.004	-16.158	0.159	134
Basic Materials	0.190	0.055	-35.674	48.919	5.954	0.208	4.490	0	-15.823	0.268	208
Communications	0.173	0.119	-30.810	47.807	4.886	0.119	4.335	-0.047	-16.260	0.246	147
Energy	0.158	0.096	-35.714	48.309	5.527	0.035	4.275	-0.035	-16.382	0.362	122
Consumer, Cyclical	0.261	0.170	-28.809	38.201	5.348	0.153	4.358	-0.026	-16.015	0.245	212
Financial	0.205	0.112	-40.133	55.163	5.294	0.156	4.924	-0.032	-16.159	0.272	366
Technology	0.159	0.165	-25.818	34.320	5.067	0.050	3.960	-0.036	-16.162	0.267	79
Industrial	0.196	0.123	-58.848	48.148	5.357	0.126	4.364	-0.018	-16.038	0.256	293
Consumer, Non-cyclical	0.283	0.211	-31.425	41.759	4.343	0.129	4.288	-0.050	-16.465	0.206	262
Utilities	0.176	0.131	-26.030	28.125	4.276	0.029	4.075	-0.048	-16.551	0.225	103
Diversified	0.210	0.048	-23.529	30.195	5.211	0.181	4.436	-0.013	-15.876	0.256	23

Panel B: Balance sheet

	Market Cap ($bil.)	**Tot Assets ($bil.)**	**Tot Debt ($bil.)**
North America	48.914	123.886	32.957
Latin America	7.167	25.520	8.288
Asia-Pacific	8.613	30.583	7.715
Western Europe	16.858	97.682	29.290
Emerging Europe	2.844	10.154	2.135
MEA	1.892	4.852	1.056
Basic Materials	8.542	12.783	3.379
Communications	17.890	24.288	7.650
Energy	24.144	39.474	7.284
Consumer, Cyclical	10.100	17.669	6.051
Financial	14.880	198.829	56.687
Technology	24.436	16.209	2.529
Industrial	8.835	14.648	4.461
Consumer, Non-cyclical	17.451	13.819	3.540
Utilities	10.148	28.427	10.190
Diversified	4.394	17.717	3.129

Table 3.3: Pearson correlation between factors and exogenous variables

The table reports the correlation coefficients of the global and regional factors. Panel A reports the matrix of the Pearson correlation between the six estimated factors. Panel B reports the correlation between the regional factors (each estimated by the first PC of a portfolio of the relevant stocks) and a global factor, the S&P500 index and S&P500 Financials index. Panel C reports the correlation between our estimated factors and the S&P500 index and S&P500 Financials index, i.e. before orthogonalising against the Financial factor.

Panel A: Correlation between regional factors

	(1)	(2)	(3)	(4)	(5)	(6)
North America (1)	1					
Latin America (2)	0.429	1				
Asia-Pacific (3)	0.294	0.460	1			
Western Europe (4)	0.483	0.374	0.361	1		
Emerging Europe (5)	0.215	0.441	0.406	0.422	1	
MEA (6)	-0.044	0.097	0.210	-0.025	0.131	1

Panel B: First PCs

	Glob PC	S&P500	S&P Fin
North America	0.902	0.962	0.847
Latin America	0.909	0.766	0.638
Asia-Pacific	0.904	0.685	0.557
Western Europe	0.968	0.821	0.708
Emerging Europe	0.889	0.688	0.597
MEA	0.348	0.225	0.177

Panel C: Factors

	S&P500	S&P Fin
Global	0.718	0.570
North America	0.663	0.689
Latin America	0.291	0.277
Asia-Pacific	0.183	0.158
Western Europe	0.402	0.445
Emerging Europe	0.212	0.250
MEA	0.009	0.007

Table 3.4: Mapping estimated factors and Fama-French 3 factors

This table reports the maximum squared canonical correlations between the market, size and value factors constructed by Fama and French, and our three orthogonal factors (US Financials, a global factor and a regional factor). Panel B reports the same but using only the regional factors.

Panel A: Three factors and Fama-French

	North America	Latin America	Asia-Pacific	Western Europe	Emerging Europe	MEA
Asia_Pacific_ex_Japan_3_Factors	0.765	0.780	0.876	0.760	0.763	0.746
Europe_3_Factors	0.830	0.814	0.817	0.961	0.834	0.790
Global_3_Factors	0.922	0.875	0.894	0.927	0.874	0.856
Global_ex_US_3_Factors	0.840	0.816	0.857	0.927	0.817	0.777
North_America_3_Factors	0.958	0.891	0.888	0.895	0.886	0.885

Panel B: Regional factor and Fama-French

	North America	Latin America	Asia-Pacific	Western Europe	Emerging Europe	MEA
Asia_Pacific_ex_Japan_3_Factors	0.056	0.064	0.265	0.055	0.078	0.088
Europe_3_Factors	0.214	0.078	0.051	0.204	0.047	0.015
Global_3_Factors	0.308	0.090	0.115	0.133	0.030	0.014
Global_ex_US_3_Factors	0.129	0.060	0.133	0.179	0.044	0.034
North_America_3_Factors	0.371	0.098	0.072	0.122	0.024	0.004

Table 3.5: Model estimates

The table reports the average magnitude of the factor loadings and the average of their volatilities, aggregated by either region or sector. Within each group, we also report the percentage of stocks with AR(1) parameter larger than 0.5 and the percentage of stocks whose loadings vary so little that we consider them constant. The loading magnitude is estimated via OLS from a static loading model, while the AR(1) parameter and variance are estimated via maximum likelihood estimation from Eq. (3.1).

$$r_{i,t} = a_{i,t}O_t + b_{i,t}G_t + \sum_{j=1}^{R} c_{i,t}F_{j,t}\mathbb{1}_{\{i\in\mathcal{J}_j\}} + e_{i,t}$$

$$\begin{aligned} a_{i,t} &= (1-\phi_i^O)\bar{a}_i + \phi_n^O a_{i,t-1} + \eta_{i,t}^O, \quad \eta_{i,t}^O \sim iidN(0, q_i^O) \\ b_{i,t} &= (1-\phi_i^G)\bar{b}_i + \phi_n^G b_{i,t-1} + \eta_{i,t}^G, \quad \eta_{i,t}^G \sim iidN(0, q_i^G) \\ c_{i,t} &= (1-\phi_i^j)\bar{c}_i + \phi_n^j c_{i,t-1} + \eta_{i,t}^j, \quad \eta_{i,t}^j \sim iidN(0, q_i^j) \end{aligned}$$

	Financial				Global				Regional				
	Avg a_i^{OLS}	AR(1)>0.5	Std $a_{i,t}$	#static(%)	Avg b_i^{OLS}	AR(1)>0.5	Std $b_{i,t}$	#static(%)	Avg c_i^{OLS}	AR(1)>0.5	Std $c_{i,t}$	#static(%)	tot
North America	0.519	16	0.222	5	0.225	20	0.193	5	0.248	29	0.179	10	188
Latin America	0.330	19	0.156	20	0.285	29	0.126	29	0.282	33	0.140	15	217
Asia-Pacific	0.266	17	0.234	5	0.267	28	0.187	11	0.317	36	0.156	9	562
Western Europe	0.461	18	0.244	4	0.354	30	0.148	12	0.302	28	0.154	14	521
Emerging Europe	0.312	13	0.241	8	0.307	24	0.154	15	0.276	28	0.168	16	193
MEA	0.088	11	0.203	13	0.158	29	0.157	28	0.353	40	0.223	9	134
Basic Materials	0.322	11	0.227	11	0.370	30	0.167	13	0.260	27	0.171	11	208
Communications	0.357	20	0.202	8	0.255	26	0.149	20	0.299	33	0.149	17	147
Energy	0.369	22	0.190	7	0.432	45	0.145	7	0.239	29	0.173	11	122
Consumer, Cyclical	0.358	14	0.243	5	0.249	21	0.176	15	0.325	35	0.150	15	212
Financial	0.387	11	0.263	6	0.269	27	0.175	13	0.313	36	0.165	8	366
Technology	0.352	23	0.219	5	0.250	24	0.154	18	0.304	34	0.157	18	79
Industrial	0.350	15	0.213	10	0.301	27	0.164	15	0.297	32	0.160	12	293
Consumer, Non-cyclical	0.306	23	0.221	5	0.234	23	0.171	16	0.300	30	0.165	10	262
Utilities	0.295	22	0.194	13	0.249	34	0.140	17	0.339	29	0.177	15	103
Diversified	0.346	17	0.188	13	0.264	30	0.146	13	0.365	39	0.117	17	23

Table 3.6: Goodness of fit

The table reports, in Panel A, the goodness of fit of our model compared with a static model where the loadings are estimated using OLS; and in Panel B the number of times the time-varying λ_{it} is outside the 95% confidence interval of the static λ_i, estimated with ordinary least squares. The numbers reported are averages of the total number in each group. Note that $T = 521$ and $T \times 0.05 = 26$.

Panel A: R^2 comparison

	R^2	R^2-OLS	Δ
North America	0.707	0.459	0.248
Latin America	0.493	0.325	0.168
Asia-Pacific	0.550	0.279	0.271
Western Europe	0.686	0.464	0.222
Emerging Europe	0.586	0.329	0.256
MEA	0.506	0.226	0.280
Basic Materials	0.623	0.381	0.242
Communications	0.538	0.327	0.211
Energy	0.642	0.427	0.215
Consumer, Cyclical	0.607	0.354	0.253
Financial	0.668	0.401	0.268
Technology	0.537	0.310	0.228
Industrial	0.588	0.361	0.227
Consumer, Non-cyclical	0.547	0.289	0.258
Utilities	0.542	0.322	0.219
Diversified	0.574	0.389	0.185

Panel B: Significant deviations from OLS

	Fin	Glob	Reg
North America	168	177	146
Latin America	89	81	138
Asia-Pacific	130	134	124
Western Europe	187	141	136
Emerging Europe	124	108	118
MEA	77	82	193
Basic Materials	133	153	133
Communications	131	99	112
Energy	128	172	134
Consumer, Cyclical	149	131	139
Financial	176	141	178
Technology	133	110	102
Industrial	121	111	121
Consumer, Non-cyclical	139	113	119
Utilities	121	108	136
Diversified	122	100	133

Table 3.7: Misspecification tests

The table reports in the first column the number of factors implied by the Bai and Ng (2002)'s IC_{p1} criterion for the returns matrix, the residual matrix derived from a static loading factor model and the residual matrix implied by our time-varying factor loading model; in the second column the percentage of stocks for which we reject the null at 99% confidence level using the White's test; in the last two column, the Breusch and Godfrey with 2 and 5 lags, respectively

	Bai-Ng02 (#)	White (%)	BG 1-2 (%)	BG 1-5 (%)
Returns	10			
Static loadings	5	51	12	31
TV loadings	3	5	9	21

Table 3.8: Static variance decomposition

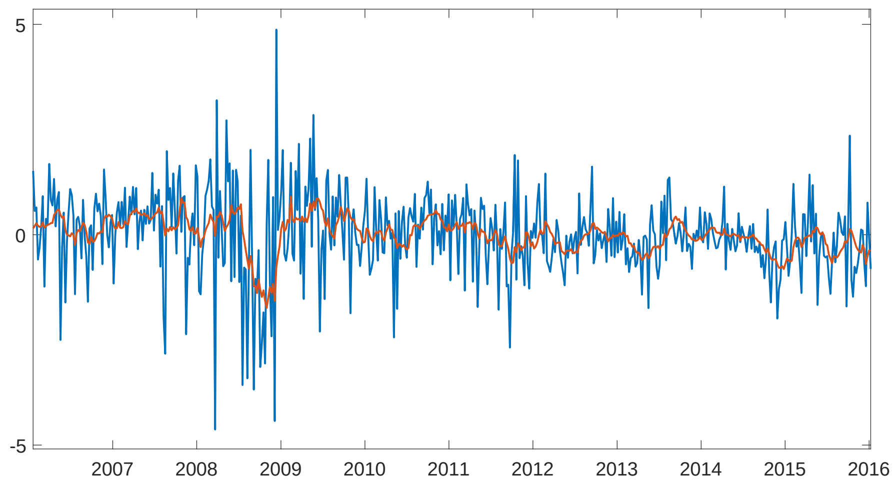
This table reports the average share of variance explained by the common factors in the relevant region or sector.

	Fin	Glob	Reg	Idio
North America	29.419	8.905	7.590	54.086
Latin America	12.938	9.354	10.159	67.549
Asia-Pacific	8.513	8.723	10.600	72.163
Western Europe	22.573	13.707	10.080	53.641
Emerging Europe	10.976	10.250	11.657	67.118
MEA	1.052	3.114	18.421	77.412
Basic Materials	13.261	15.854	8.929	61.956
Communications	15.152	7.534	10.012	67.302
Energy	15.263	20.461	6.888	57.387
Consumer, Cyclical	15.513	7.603	12.264	64.620
Financial	19.084	8.531	12.405	59.980
Technology	14.354	6.824	9.759	69.063
Industrial	15.246	10.538	10.272	63.943
Consumer, Non-cyclical	11.400	6.992	10.454	71.155
Utilities	10.704	8.423	13.081	67.792
Diversified	14.379	7.898	16.624	61.099

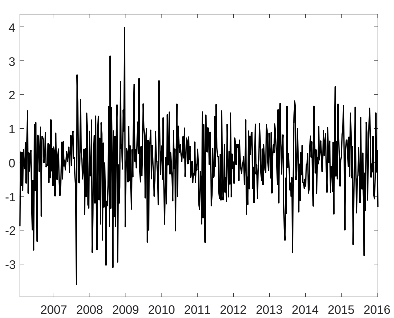
Figure 3.1: Estimated global and regional factors

The figure plots the estimated global factor and the regional factors for Asia-Pacific, Emerging Europe, Latin America, Middle-East Africa, North America and Western Europe. Together with the estimated factor, we plot a double-sided two-month moving average. The factors are estimated by PCA from the model with static loadings in Eq. (3.6). The factors are rotated to ensure that they are positively correlated with the stock market index of the biggest country in the region.

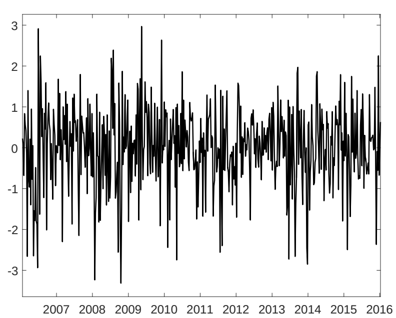
(a) Global



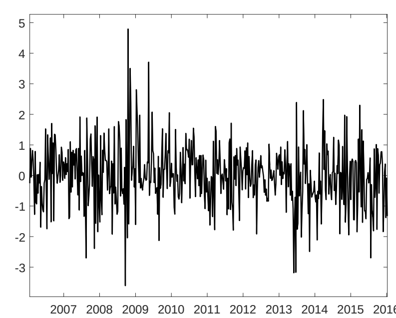
(b) Asia-Pacific



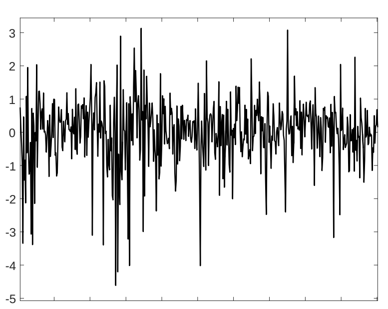
(c) Emerging Europe



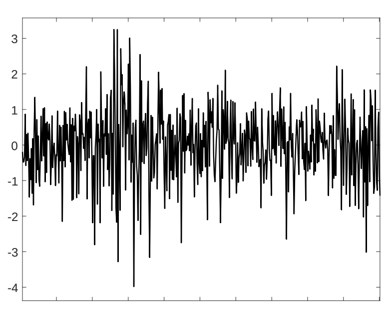
(d) Latin America



(e) Middle-East and Africa



(f) North America



(g) Western Europe

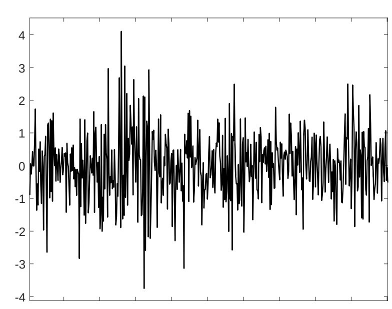
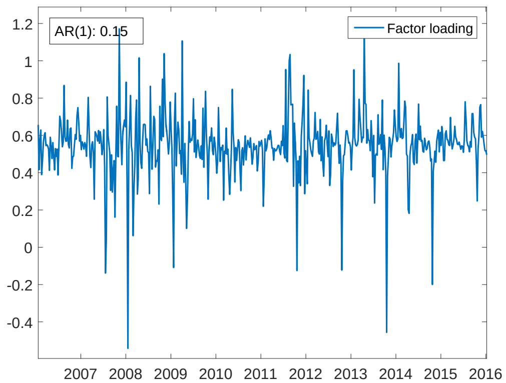


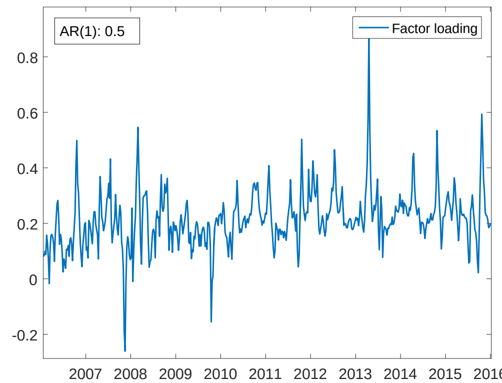
Figure 3.2: The conditional factor exposure of two large firms

The figure plots time-varying loadings estimated for IBM and Tenaris, respectively. The loadings are exposures of each stock's returns to financial, global and regional factors.

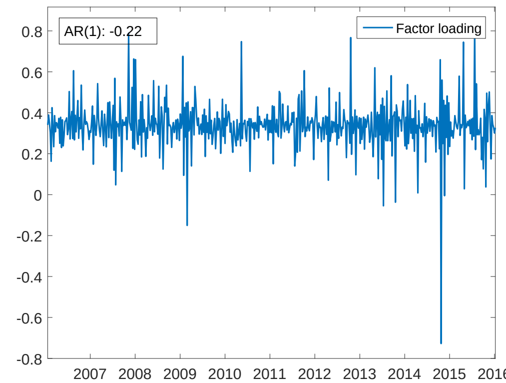
(a) IBM, financial factor



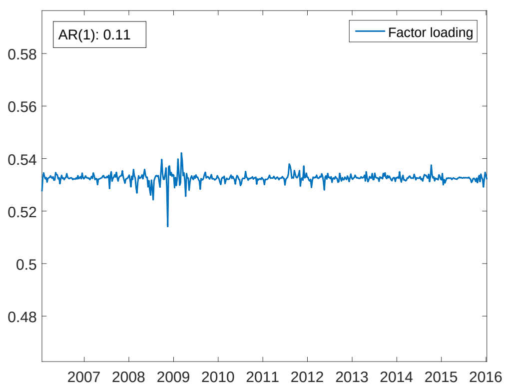
(b) IBM, global factor



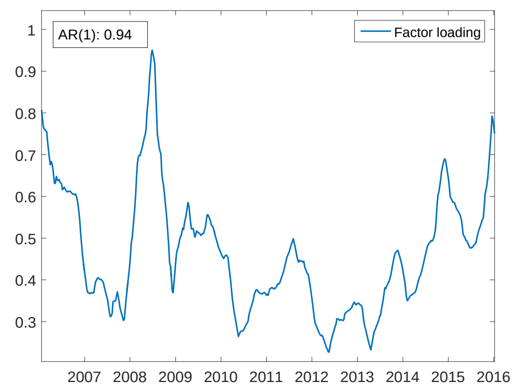
(c) IBM, regional factor



(d) Tenaris, financial factor



(e) Tenaris, global factor



(f) Tenaris, regional factor

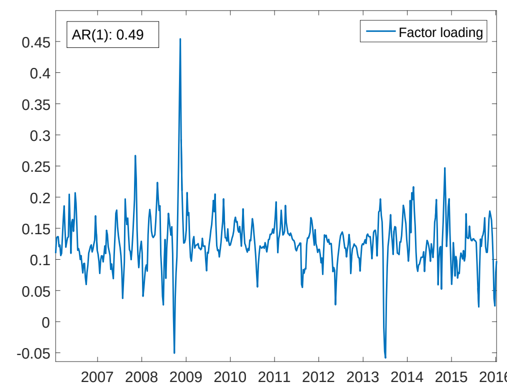
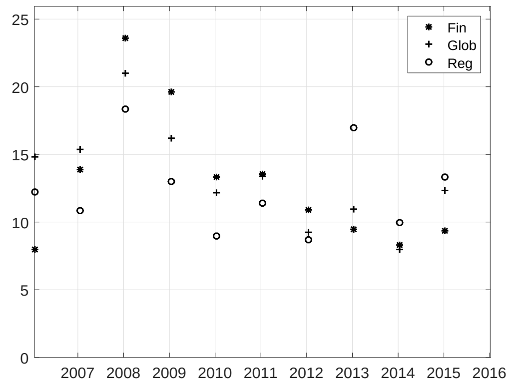


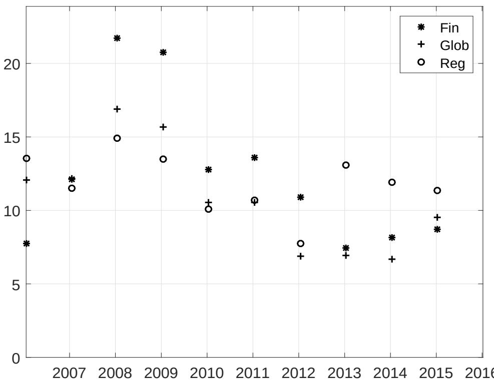
Figure 3.3: Average deviations from OLS per year

The figure plots, for every year, the cross-sectional average of the number of significant deviations from OLS in our six world regions (Asia-Pacific, Emergin Europe, Latin America, Middle-East and Africa, North America, Western Europe).

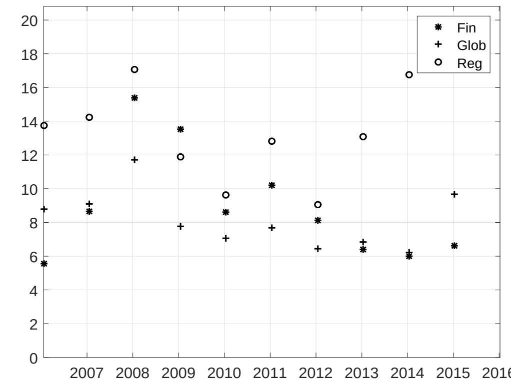
(a) Asia-Pacific



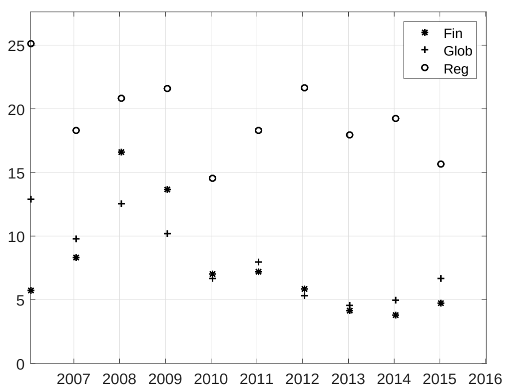
(b) Emerging Europe



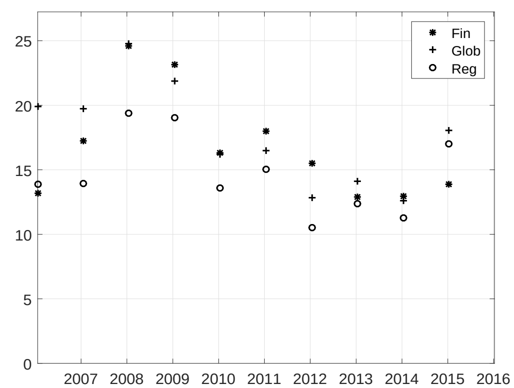
(c) LatinAmerica



(d) Middle-East and Africa



(e) North America



(f) Western Europe

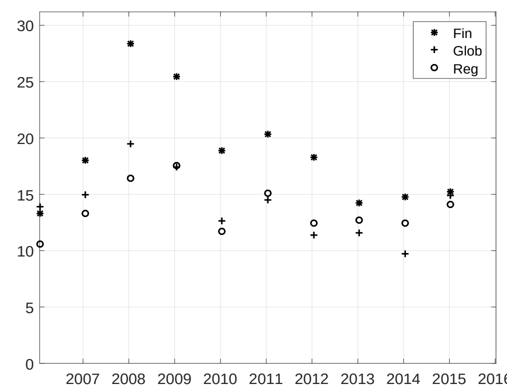
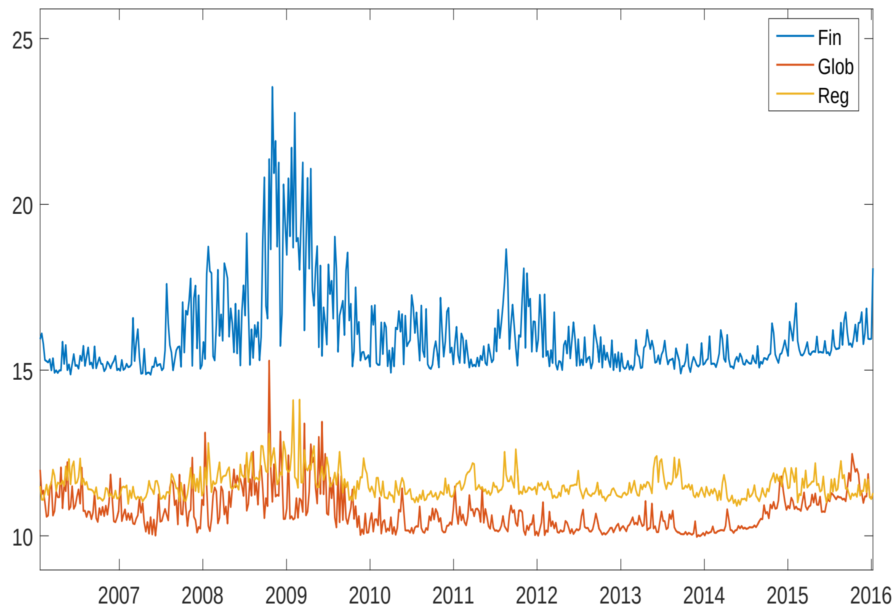


Figure 3.4: Time-varying variance decomposition

The figure reports the average estimated conditional variance decompositions. Panel (a) shows the cross-sectional average of the share of variance explained by each factor at each point in time. Panel (b) reports the total share of variance explained by the factors as the sum of the three series in Panel (a).

(a) Variance decomposition



(b) Variance decomposition

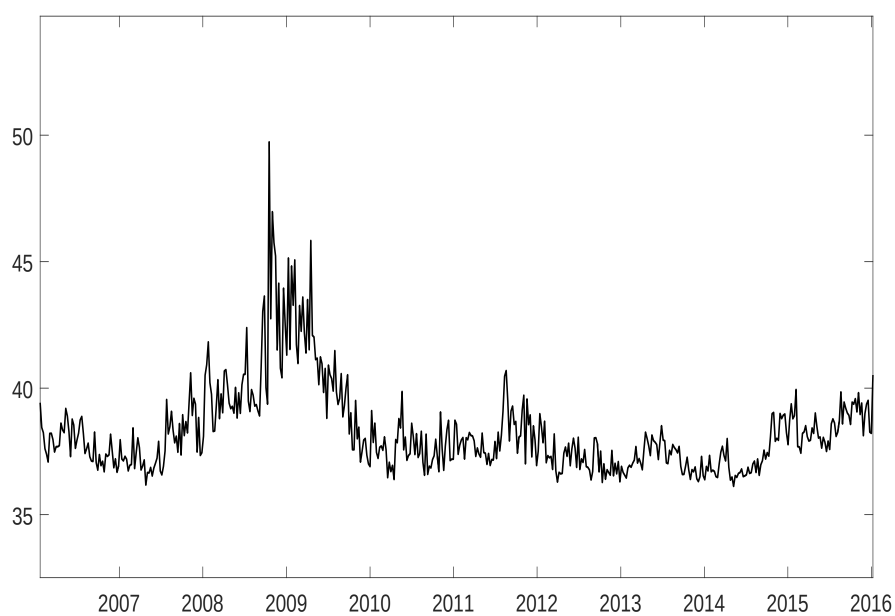
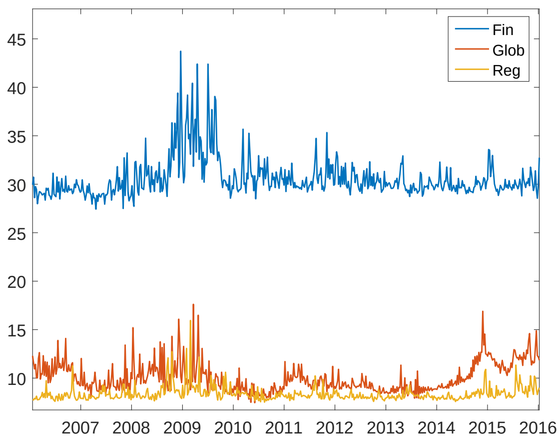


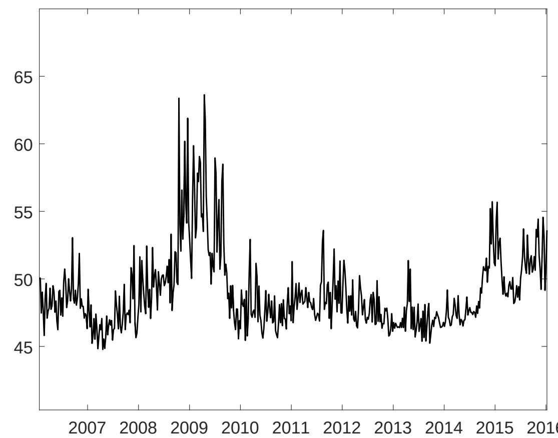
Figure 3.5: Time-varying variance decomposition (by region)

The figure reports the average estimated conditional variance decompositions, aggregated by region. In the left column of panels, the blue line represents the percentage of variance explained by the financial factor. The yellow line by the regional and the orange line by the global factor. The right column of panels shows the total share of variance explained by the factors, the sum of the left panel.

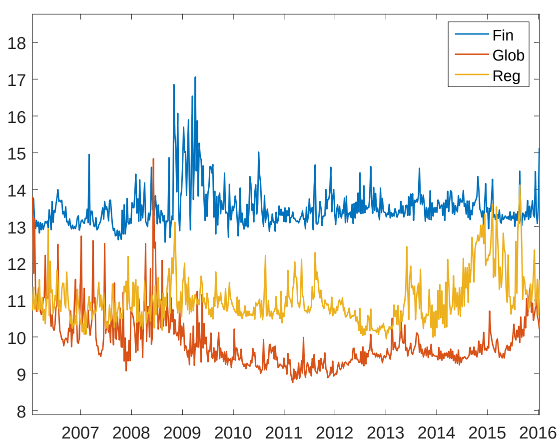
(a) North America



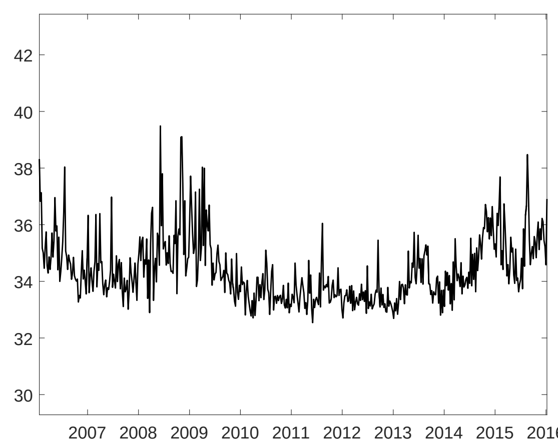
(b) North America total



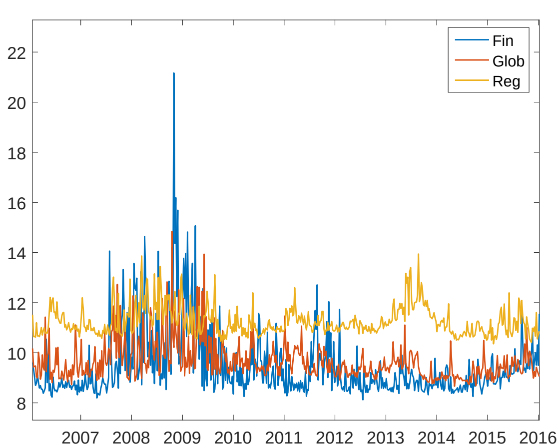
(c) Latin America



(d) Latin America total



(e) Asia-Pacific



(f) Asia-Pacific total

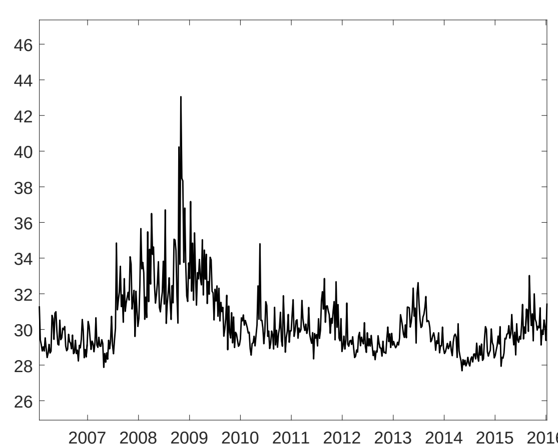
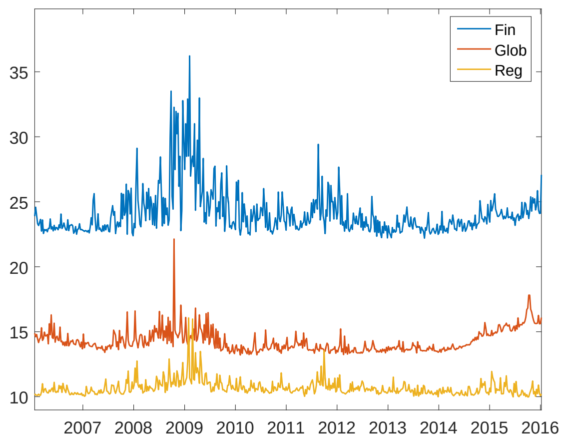


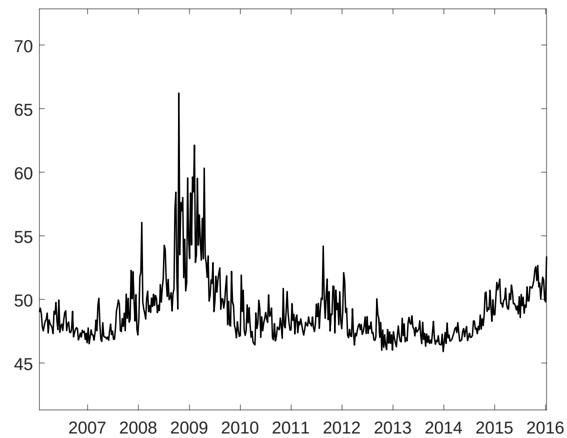
Figure 3.6: Time-varying variance decomposition (by region) - continued

The figure reports the average estimated conditional variance decompositions, aggregated by region. In the left column of panels, the blue line represents the percentage of variance explained by the financial factor. The yellow line by the regional and the orange line by the global factor. The right column of panels shows the total share of variance explained by the factors, the sum of the left panel.

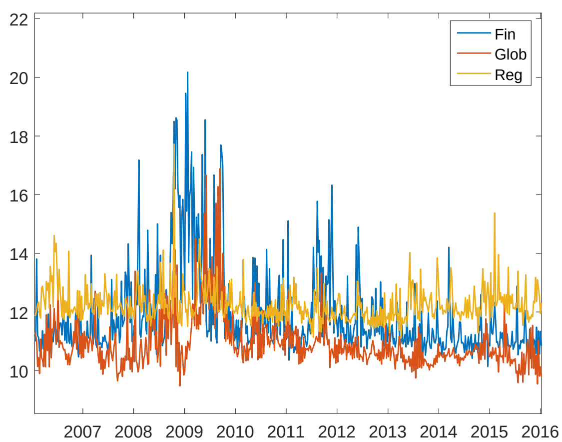
(a) Western Europe



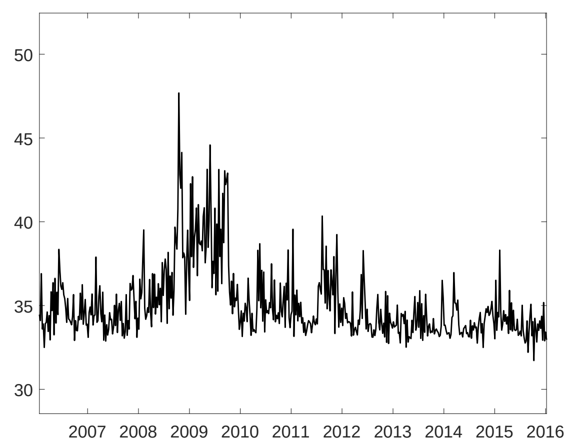
(b) Western Europe total



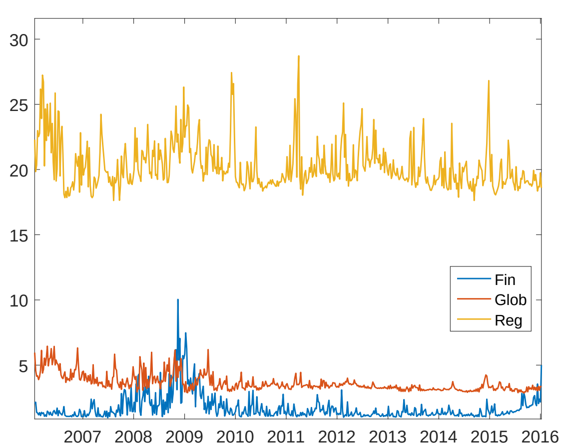
(c) Emerging Europe



(d) Emerging Europe total



(e) Middle-East Africa



(f) Middle-East Africa total

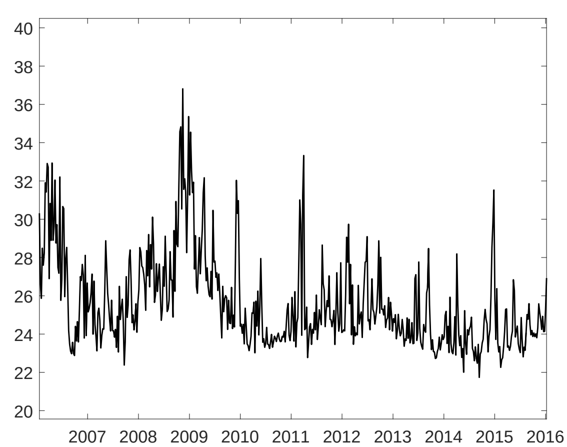
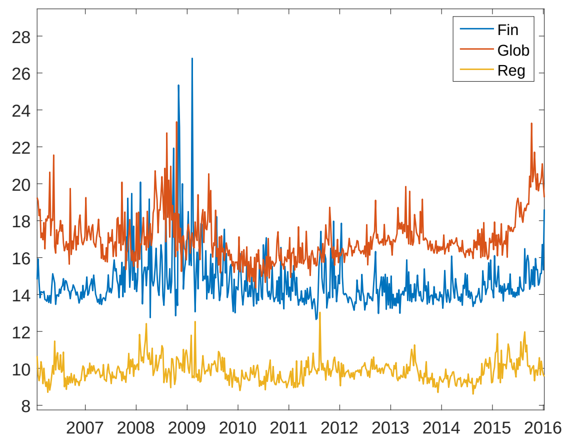


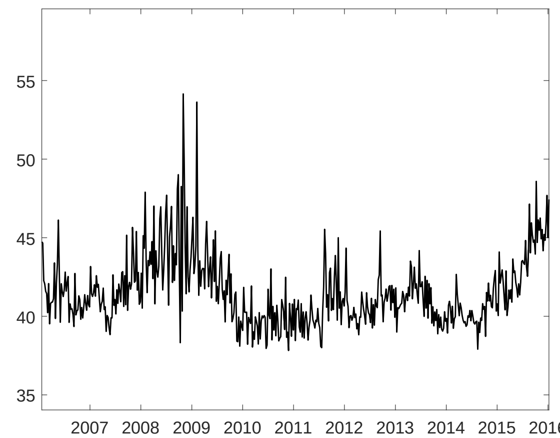
Figure 3.7: Time-varying variance decomposition (by sector)

The figure reports the average estimated conditional variance decompositions, aggregated by sector. In the left column of panels, the blue line represents the percentage of variance explained by the financial factor. The yellow line by the regional and the orange line by the global factor. The right column of panels shows the total share of variance explained by the factors, the sum of the left panel.

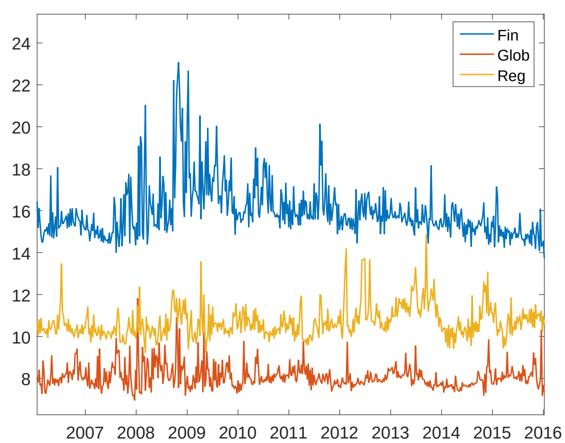
(a) Basic Materials



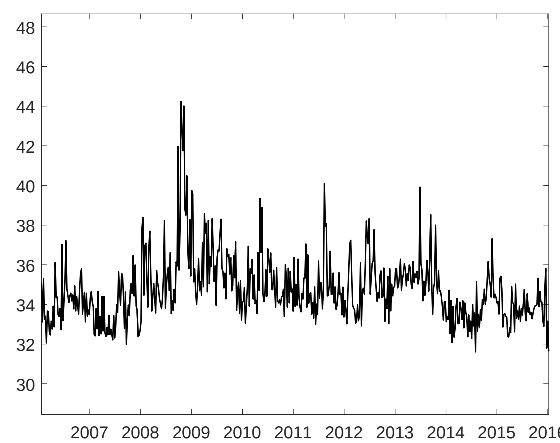
(b) Basic Materials total



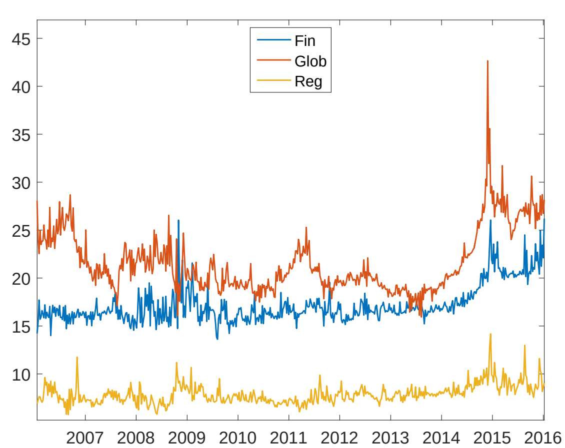
(c) Communications



(d) Communications total



(e) Energy



(f) Energy total

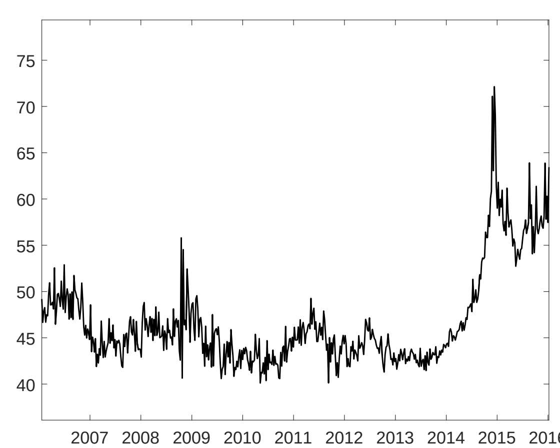
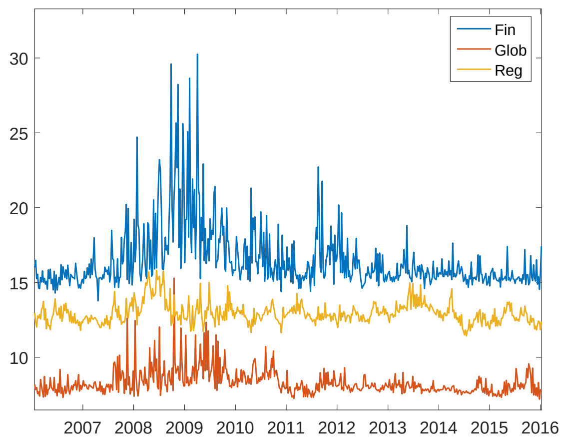


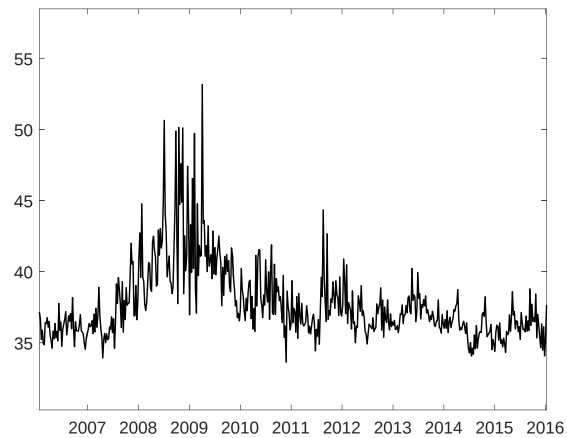
Figure 3.8: Time-varying variance decomposition (by sector) - continued

The figure reports the average estimated conditional variance decompositions, aggregated by sector. In the left column of panels, the blue line represents the percentage of variance explained by the financial factor. The yellow line by the regional and the orange line by the global factor. The right column of panels shows the total share of variance explained by the factors, the sum of the left panel.

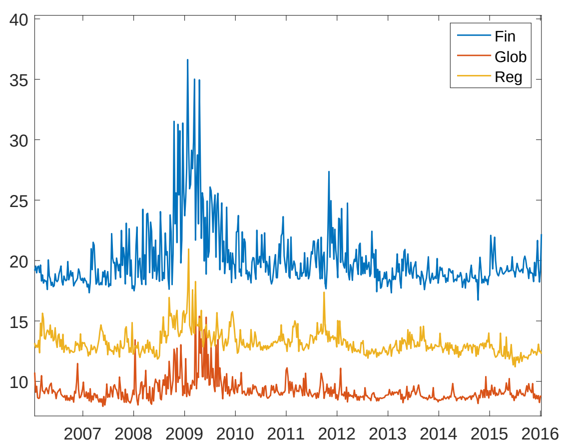
(a) Consumer Cyclical



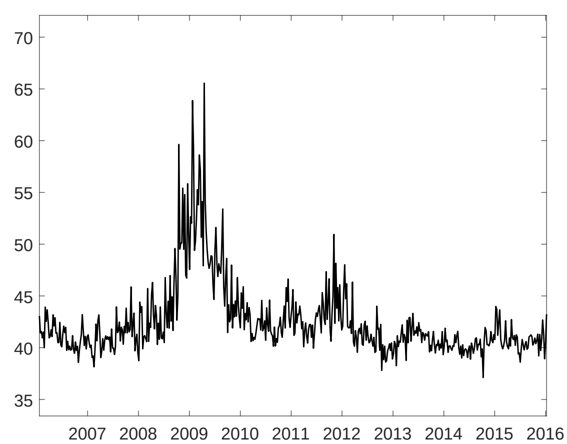
(b) Consumer Cyclical total



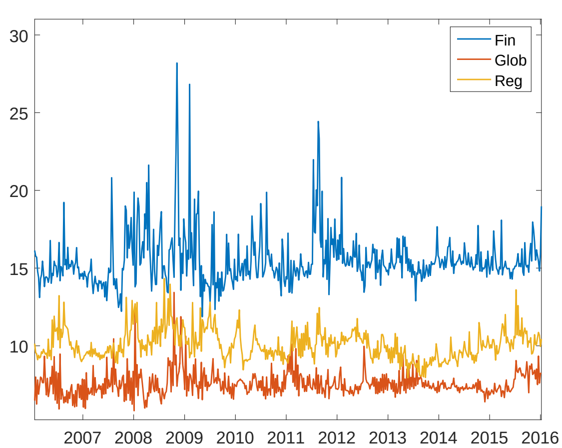
(c) Financial



(d) Financial total



(e) Technology



(f) Technology total

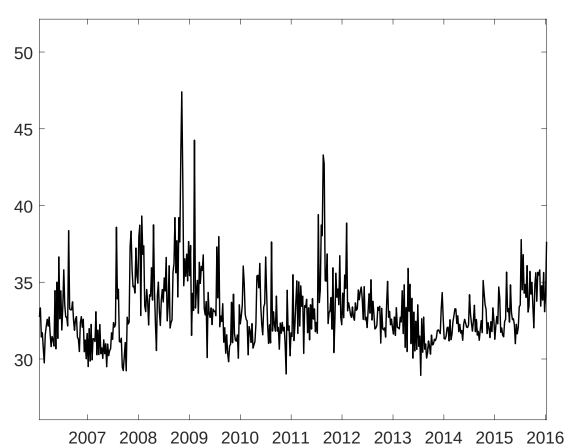
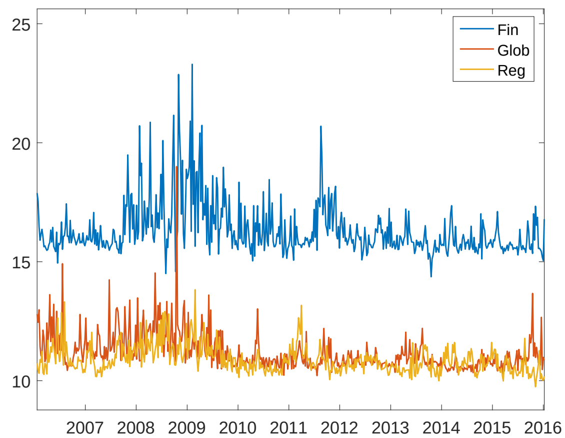


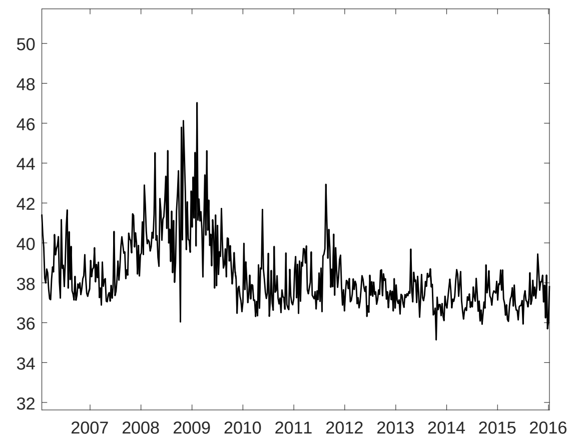
Figure 3.9: Time-varying variance decomposition (by sector) - continued

The figure reports the average estimated conditional variance decompositions, aggregated by sector. In the left column of panels, the blue line represents the percentage of variance explained by the financial factor. The yellow line by the regional and the orange line by the global factor. The right column of panels shows the total share of variance explained by the factors, the sum of the left panel.

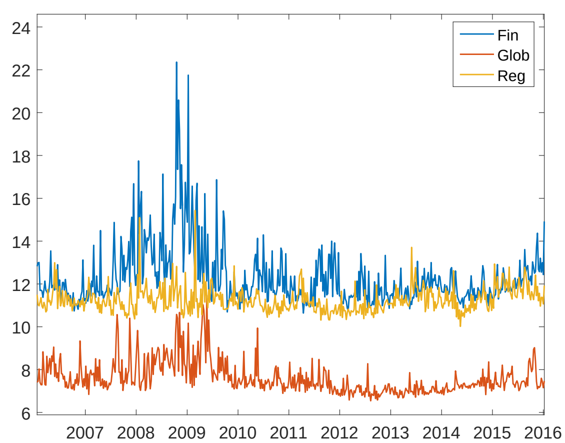
(a) Industrial



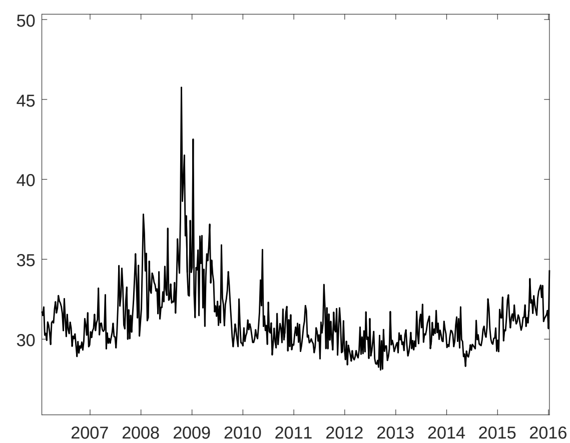
(b) Industrial total



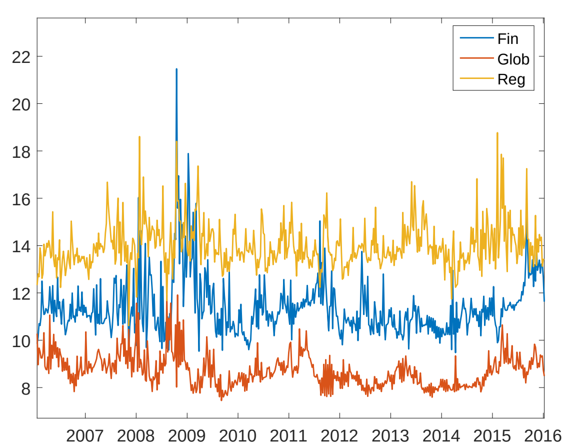
(c) Consumer Non-cyclical



(d) Consumer Non-cyclical total



(e) Utilities



(f) Utilities total

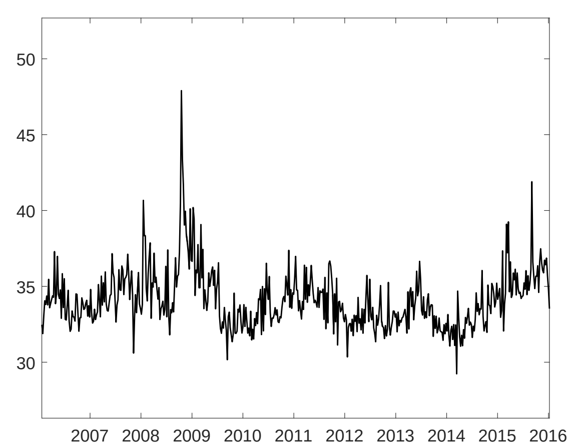
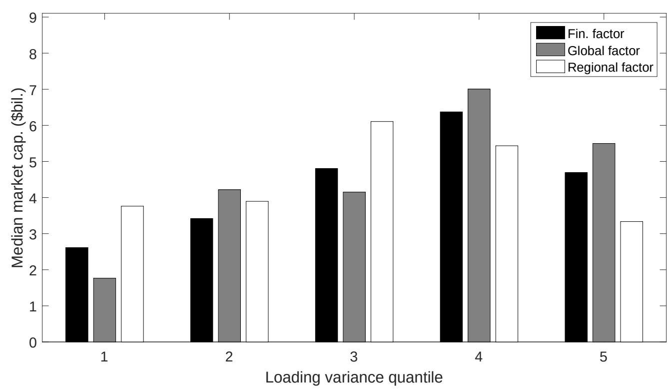


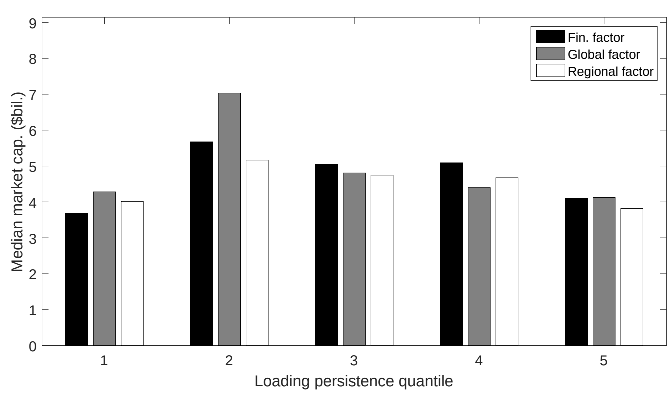
Figure 3.10: Relationship with firm size

The figure reports the median market capitalisation (at the end of the sample) ordered by factor loading variance (Panel (a)), persistence (Panel (b)) and magnitude (Panel (c)). At the end of the sample, stocks are sorted in quantiles of either loading variance, or persistence or magnitude. Quantile five contains the larger value. Then, for each quantile we calculate the median market market capitalisation and we plot it against the quantile number.

(a) Variance of loading



(b) Persistence of loading



(c) Loading magnitude

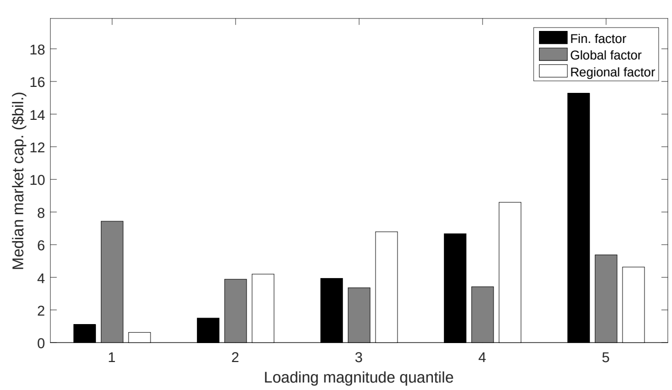
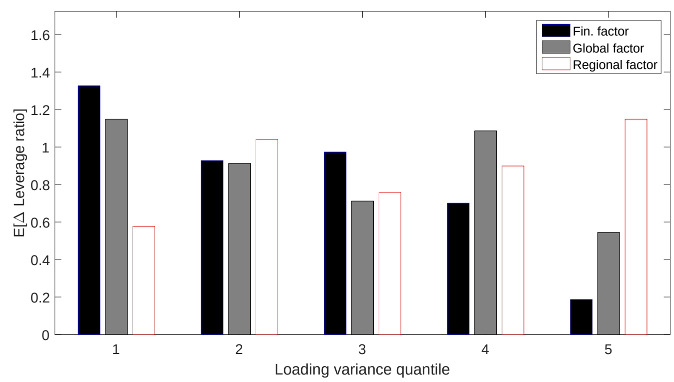


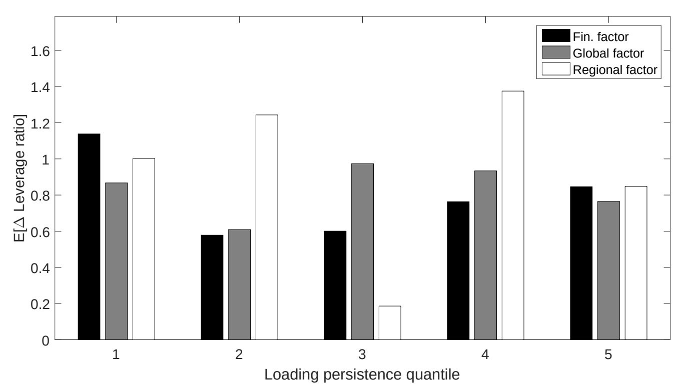
Figure 3.11: Relationship with firm leverage

The figure reports the average quarterly change in leverage ratio (from January 2010 and January 2016) ordered by factor loading variance (Panel (a)), persistence (Panel (b)) and magnitude (Panel (c)). At the end of the sample, stocks are sorted in quantiles of either loading variance, or persistence or magnitude. Quantile five contains the larger value. Then, for each quantile we calculate the average quarterly change in leverage ratio (debt over assets) and we plot it against the quantile number. Financials are excluded.

(a) Variance of loading



(b) Persistence of loading



(c) Loading magnitude

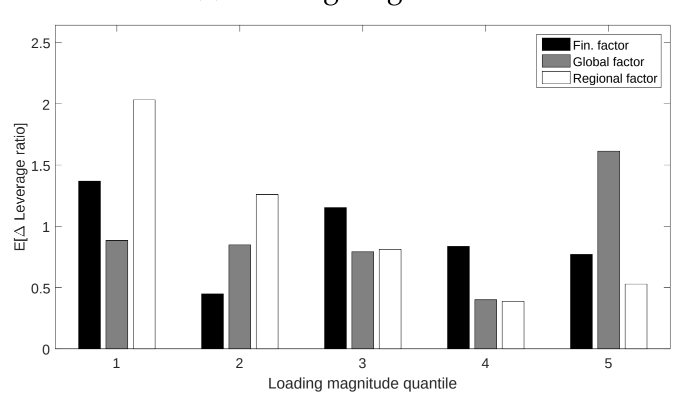
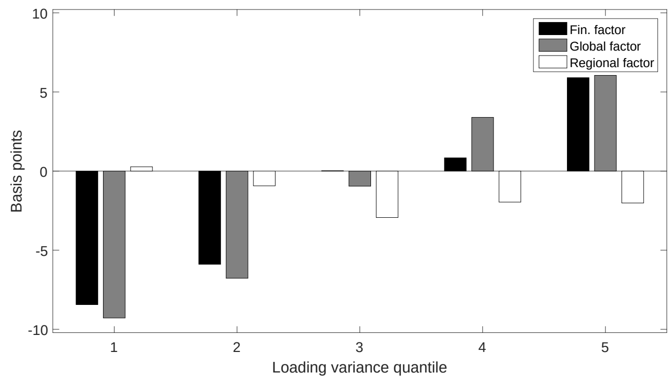


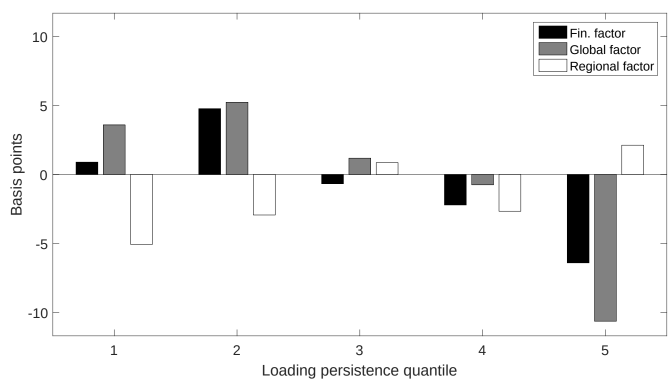
Figure 3.12: Relationship with expected returns

The figure reports the average weekly stock returns from January 2010 until January 2016, expressed in basis points (one basis point $= 0.01\%$), ordered by factor loading variance (Panel (a)), persistence (Panel (b)) and magnitude (Panel (c)). Financials are excluded. At the end of the sample, stocks are sorted in quantiles of either loading variance, or persistence or magnitude. Quantile five contains the larger value. Then, for each quantile we calculate the average log-returns and we plot it against the quantile number.

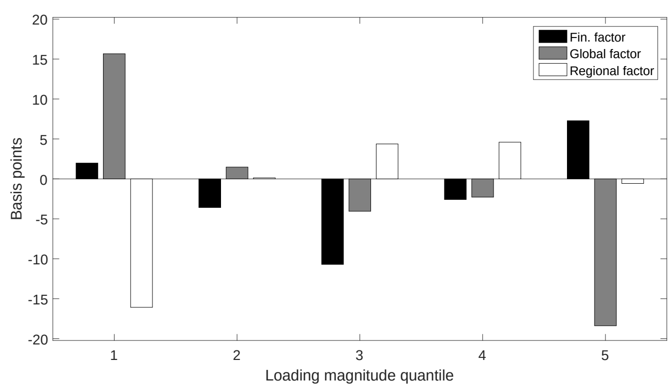
(a) Variance of loading



(b) Persistence of loading



(c) Loading magnitude



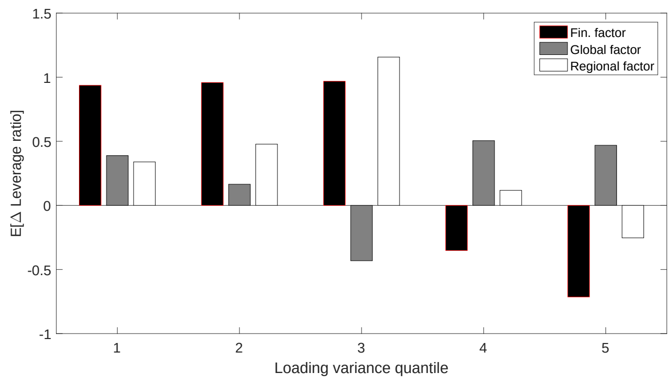
Appendix 3.A Leverage for Financials

Figure 3.A.1 reports the average quarterly change in leverage ratio (from January 2010 and January 2016) for financial firms ordered by factor loading variance (Panel (a)), persistence (Panel (b)) and magnitude (Panel (c)).

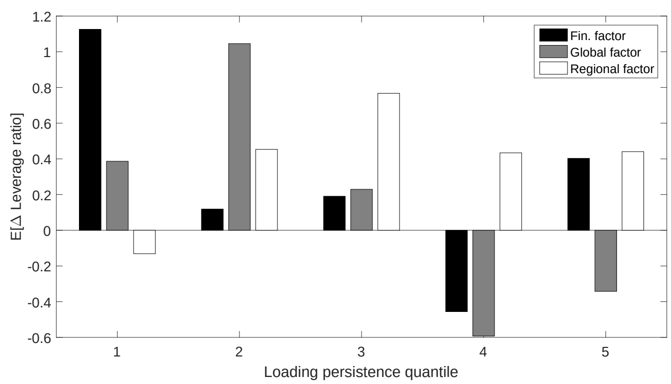
Figure 3.A.1: Relationship with firm leverage for Financials

The figure reports the average quarterly change in leverage ratio (from January 2010 and January 2016) for financial firms ordered by factor loading variance (Panel (a)), persistence (Panel (b)) and magnitude (Panel (c)). At the end of the sample, stocks are sorted in quantiles of either loading variance, or persistence or magnitude. Quantile five contains the larger value. Then, for each quantile we calculate the average quarterly change in leverage ratio (debt over assets) and we plot it against the quantile number.

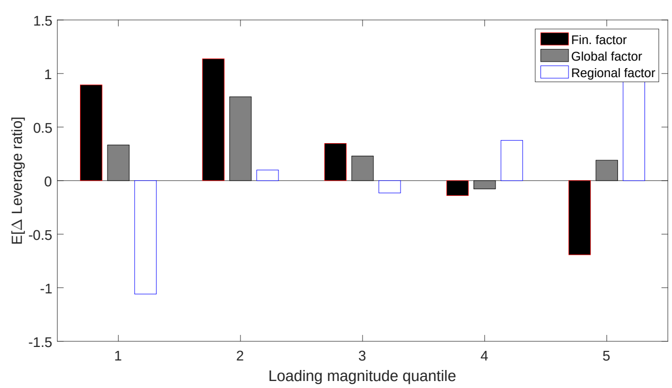
(a) Variance of loading



(b) Persistence of loading



(c) Loading magnitude



Appendix 3.B Data cleaning

For religious reasons, some countries have a different reference weekday. Thus, we need to pay attention at how we synchronize data across country. The length of the complete time series for industrialised countries is $T = 835$.

Egypt. Following the "Egyptian Revolution of 2011" that started on the 25th January 2011, the stock exchange closed from the 27th January until the 23rd March, which results in 7 consecutive missing data cells for all stocks. Thus, we "carry last value", which will result in 7 zero-return observations.

Russia, Ukraine and India. Table 3.1 shows that both countries' time series start on the 23rd January 2000 and finish on the the 10th January 2016. This comes unexpected because both days are Sundays and Bloomberg should be assigning to the weekly observation the last working day of the week. However, in these countries the stock exchange operates normally, from Monday to Friday. Thus, prices refer to the Friday close (or the last available data point of the week) but Bloomberg reports the Sunday date (for no particular reason). We checked with other data providers (Datastream) that this is the case. Thus, since we are downloading Friday to Friday data, the last observation is missing and $T = 834$, where the last observation refers to the penultimate week of the industrialised countries. Note that we do not need to shift the time series because all the other data points match across countries.

Egypt, Israel, Qatar, UAE. Weekday reference is Thursday, for religious reasons. I.e. weekends are on Fridays and Saturdays. $T = 835$

Korea and Taiwan. Time series start on the 22nd January 2000 and they end on the 9th January 2016, which are both Saturdays. Hence, an observation is missing at the end the sample. $T = 834$.

In both countries the stock exchange operates a traditional trading calendar. So, the weekly data refers to the last working-day traded price.

Serbia. For all companies, the time series of prices starts in December 2005. Thus, if we want to include this country in the study we have to trim all time series. Note that, according to Bekaert et al. (2014), Table IV, pag. 2618, the Serbian equities experienced the most negative return over the crisis period so it is a country worth including.

In conclusion, we are going to eliminate the last observation for countries with 835 observations, and the final time series length will be $T = 834$. The countries are 54 and not 55, since there is a repetition of Slovenia in Table A.1 of Bekaert et al. (2014).

Appendix 3.C Replication of Bekaert et al. (2014)

In this section we report the results of the replication of the main results of Bekaert et al. (2014). First, we prepare the data and construct the country-sector portfolios using our sample. Second, we estimate the interdependence model.

Bekaert et al. (2014) downloaded the composition of the index at one point in time only: December 2008. This could be assessed interpolating the number of components of "top-capitalisation" indexes in Table A.1 with the time series of the *number* of index components provided by Bloomberg. In fact, these stock market indexes, capturing the top quantile of market capitalisation of the country, see their number of components varying substantially over time.

Prices (hence returns) are expressed in \$-terms. The data have been resampled at weekly frequency. For the balance sheet data this is trivial because they are usually provided every quarter. The price data is resampled at weekly frequency with the following rule: the last available (daily) closing price in the week represents the observation for the week. For instance, for the industrialised countries, where most daily closing prices are available, the weekly price would be represented by the observation recorded by the exchange on Friday. For religion reasons, some countries trade until Thursday and the first day of the week is Sunday (see Table 3.1 and Section 3.B).

- The full sample consists of $T = 834$ weeks, from the week ending on Sunday 23^{rd} January 2000 until the week ending on Sunday 10^{th} January 2016.

- To use a higher number of complete time series for each country, we restrict the sample from **January 2006 to January 2016**. We consider a time series to be "full" if the maximum number of consecutive missing data points is four.

Returns are calculated as $p_t/p_{t-1} - 1$, instead of using logarithmic approximation. This is because the log approximation is valid only when returns are really small, which might be true for short time intervals but invalid for weekly returns. Besides, log-returns overestimates returns and that would highly affect returns during crisis. Comparing returns in crisis calculated with our database to the ones in Bekaert et al. (2014), we concluded that they also used simple returns. The use of simple returns implies that the value of portfolios at the end of a certain period must be calculated as $100(1 + r_1)(1 + r_2) \cdots (1 + r_T)$.

We winsorise returns at 99% level. It is worth noting that Bekaert et al. (2014) remove outliers (although it is not mentioned at all in the paper, we found this information in earlier drafts and in similar work done by the authors). We think this practice would highly bias the crisis returns results.

Our results are in line with the country trends in prices during crisis and could differ from Bekaert et al. (2014) also because we consider the a wider sample of companies. See Appendix B for more details on the database and choice of countries.

Country-sector portfolios construction

Given the time-varying nature of the index composition, there are two options to construct our country-sector indexes:

1. Complete time series index. i.e. exclude the non-complete time-series. Tables 3.1 reports the number of companies with complete time series for each country and each variable. In this case, the lower bound is one series, in order to be able to represent that country.
2. Rolling window index. When one analyses long time series, it is possible that many stocks have missing data for the first years of the sample. This could happen for two reasons: (1) they were not listed, (2) they are

new entities that have been created after a merge or an acquisition[9]. In the second case, we would not be able to retrieve the old data because Bloomberg deletes it. Rebalancing our country index every year would give the opportunity to use more stocks, at the risk of becoming unstable.

We estimate the full model - option 1 - using only the complete time series.

The country-sector portfolios are calculated as capitalisation-weighted indexes of the stocks belonging to sector s of country c. Not all sectors are represented in each country. Precisely, we obtain a total of 373 country-sector portfolios (out of 540 if all 10 sectors were represented in each of the 54 countries).

The capitalisation-weighted average ensures that the main stock market index constituents of the beginning of the sample have a weight which is proportional to their value. For instance, if a company exited the main stock market index, it will have a marginal contribution in the country-sector portfolio. Also the three factors - US, global and domestic - are capitalisation-weighted portfolios.

The index is constructed as follows:

$$Index_t = \frac{\sum_{i=1}^{N} p_{it} s_{it}}{Divisor}. \tag{3.1}$$

The numerator is the total market capitalisation of the index at time t. The divisor is a constant number and it is chosen such that the index is equal to 100 (or 1000) in the base year.

Since the three factors are computed in the same way, we can compare the US factor with the time series of the S&P500 Index. Note that the S&P500 is updated every 15 seconds and the numerator in (3.1) is adjusted for corporate events. The correlation between our index and the S&P500 Index is 0.99.

Because the model wants to embed different CAPMs, all portfolios and factors returns are calculated as premiums, in excess of the US 3-month T-bill rate. The latter must be expressed in weekly units, which can be calculated from the annualised counterpart which we download. We divide the annualise interest rate by 52 to get an approximation of the weekly rate.

[9] sometimes companies change name after an acquisition, hence they also change ticker.

Estimation results

We estimate the following model on the $N = 373$ country-sector portfolios:

$$r_{i,t} = \underbrace{\alpha_{i,0} + \alpha_{i,1} r_{i,t-1} + \alpha_{i,2} dy_{i,t-1}}_{\mathbb{E}_{t-1}[r_{i,t}]} + \beta'_{i,0} F_{i,t} + e_{i,t}. \tag{3.2}$$

$r_{i,t}$ is the return on the i$-th$ sector-country portfolio, in excess of the 3-Month US T-bill rate. $F_{i,t} = [U_t, G_t, D_t^{\setminus i}]$, where U_t is the U.S. factor, G_t denotes the global factor (financial sector), and $D_t^{\setminus i}$ the domestic factor. The factors have been orthogonalised. In line with Bekaert et al. (2014), first we orthogonalise the global factor against the US factor, taking the residuals from $G_t = bUS_t + g_t$ as the global factor, $\hat{g}_t$. Then we orthogonalise the domestic factor against both U and $\hat{g}_t$, for every i. For the time being, we do not include an intercept in the orthogonalisation regression, in order to be as close as possible to Bekaert et al. (2014) (who do not mention any demeaning of variables).

dy_{it} is the cap-weighted average of the dividend yield of the stocks in the portfolio. The dividend yield is the ratio between the *annual* total cash dividend paid to shareholders and the stock price. Since not all companies pay dividends, the time series of dividend yield might be zero in certain sub-samples, or indeed for the whole sample period.

Table 3.C.1 and 3.C.2 report the un-weighted average of the factor loadings across all portfolios, including the zero-βs of the portfolios that are not represented inside one country. The rationale for this procedure is the following: if a country's main stock market index does not include stocks that belong to certain sectors, it means that country is exposed to fewer "channels" of contagion. Thus, it makes sense to include the zero βs in the calculation of the average exposure of the country.

Table 3.C.1: Table II Bekaert et al (2014)

The table reports the estimates of the following model:

$$r_{i,t} = \underbrace{\alpha_{i,0} + \alpha_{i,1} r_{i,t-1} + \alpha_{i,2} dy_{i,t-1}}_{\mathbb{E}_{t-1}[r_{i,t}]} + \beta'_{i,0} F_{i,t} + e_{i,t}.$$

$r_{i,t}$ is the return on the i$-th$ sector-country portfolio, in excess of the 3-Month US T-bill rate. $F_{i,t} = [U_t, G_t, D_t^{\backslash i}]$, where U is the U.S. factor, G denotes the global factor (financial sector), and $D_t^{\backslash i}$ the domestic factor. The factors have been orthogonalised. The table reports the unweighted average of the loadings. Because not all the countries have stocks in all 10 sectors, some betas are equal to zero, which do not enter the average in this table.

	Coef	St.Err.
Const.	0,000	0,001
AR(1)	-0,009	0,021
dy_{t-1}	-0,010	0,367
β^U	0,573	0,038
β^G	0,476	0,048
$\beta^{D/i}$	0,466	0,039
Observations	192249	
Portfolios	369	
R^2	0,528	

Table 3.C.2: Table III Bekaert et al (2014)

The table reports the estimates of the following model:

$$r_{i,t} = \underbrace{\alpha_{i,0} + \alpha_{i,1} r_{i,t-1} + \alpha_{i,2} dy_{i,t-1}}_{\mathbb{E}_{t-1}[r_{i,t}]} + \beta'_{i,0} F_{i,t} + e_{i,t}.$$

$r_{i,t}$ is the return on the i$-th$ sector-country portfolio, in excess of the 3-Month US T-bill rate. $F_{i,t} = [U_t, G_t, D_t^{\setminus i}]$, where U is the U.S. factor, G denotes the global factor (financial sector), and $D_t^{\setminus i}$ the domestic factor. The factors have been orthogonalised. The table reports the unweighted average of the loadings. Because not all the countries have stocks in all 10 sectors, some betas are equal to zero, which do not enter the average in this table.

Region	β^U	β^G	$\beta^{D/i}$
Asia-Pacific	0,566	0,580	0,675
Emerging Europe	0,427	0,427	0,373
Western Europe	0,726	0,531	0,421
Latin America	0,775	0,459	0,593
Middle-East/Africa	0,153	0,207	0,410
Sectors	β^U	β^G	$\beta^{D/i}$
Basic Materials	0,748	0,574	0,595
Communications	0,570	0,486	0,517
Consumer, Cyclical	0,629	0,488	0,482
Consumer, Non-cyclical	0,532	0,453	0,468
Diversified	0,192	0,160	0,186
Energy	0,589	0,482	0,492
Financial	0,921	0,873	0,604
Industrial	0,785	0,608	0,653
Technology	0,343	0,253	0,234
Utilities	0,426	0,379	0,428

As a robustness check, we also average the estimates of the interdependence model when the zeros are excluded. It is possible to notice that the results are **not** in line anymore with those reported previously.

Furthermore, we compare the model performance during the financial crisis (August 2007 to March 2009), which is one of the overlapping periods of our two databases. Figure 3.C.1 shows the regression line between actual and predicted country returns during crisis.

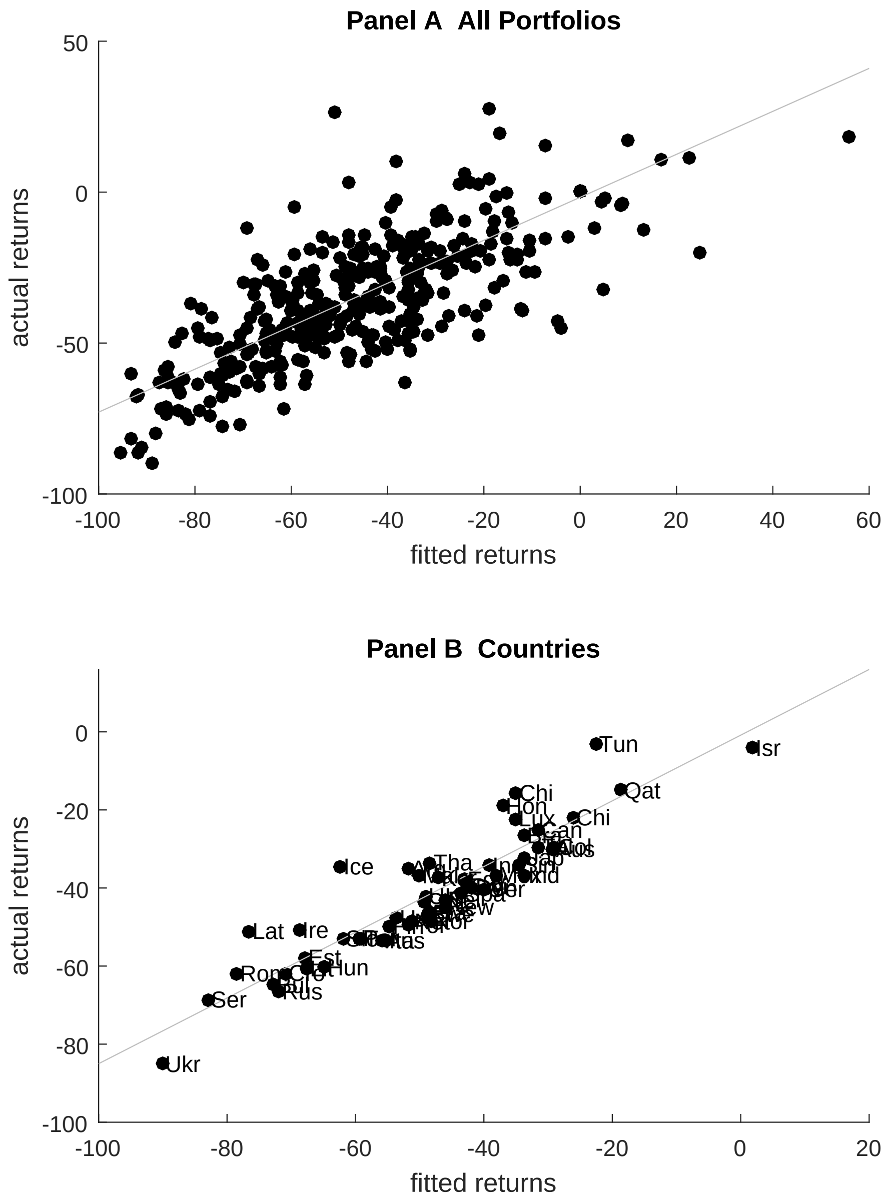


Figure 3.C.1: Goodness of fit interdependence model

Unweighted averages of cumulative returns within countries (average of sectors portfolios) during crisis (from August 2007 to March 2009). Goodness of fit regression:

$$r_i = \underset{(0.0088)}{-0.0584} + \underset{(0.0259)}{1.0697}\hat{r}_i + v_i, \quad R^2 = 0.761$$

Table 3.C.3 reports the actual and predicted country returns during crisis, with their ranking.

Table 3.C.3: Table IV Bekaert et al (2014)

The table reports unweighted average of cumulative return within country during the crisis period (from August 2007 to March 2009). Actual returns are compared to fitted returns obtained from interdependence model.

	Actual			
Country	**Returns**	**Rank**	**Returns**	**Rank**
Ukraine	-90,1	1	-85,2	1
Serbia	-83,0	2	-68,6	2
Romania	-78,5	3	-62,0	6
Latvia	-76,5	4	-51,0	5
Bulgaria	-72,8	5	-64,9	3
Russia	-71,9	6	-66,6	7
Croatia	-70,9	7	-61,8	10
Ireland	-68,8	8	-50,8	11
Estonia	-67,9	9	-57,9	9
Lithuania	-67,6	10	-60,7	17
Hungary	-64,9	11	-60,4	16
Iceland	-62,4	12	-34,5	13
Slovenia	-62,0	13	-53,0	15
Poland	-59,3	14	-52,9	14
Turkey	-59,1	15	-52,9	4
Austria	-55,8	16	-53,5	8
Italy	-55,3	17	-53,6	18
Finland	-54,7	18	-49,8	21
UAE	-53,6	19	-47,7	22
Argentina	-51,9	20	-35,0	29
Portugal	-51,8	21	-49,6	19
Netherlands	-51,3	22	-48,6	26
Malta	-50,1	23	-36,8	27
CzechRepublic	-49,4	24	-43,4	32
UK	-49,1	25	-42,1	24
Sweden	-48,8	26	-46,8	31
France	-48,5	27	-46,3	25
Thailand	-48,3	28	-33,6	33
Norway	-48,3	29	-48,5	37
Korea	-47,1	30	-37,4	36
Belgium	-46,1	31	-42,9	35
NewZealand	-46,0	32	-44,9	34
Spain	-43,4	33	-41,4	30
Egypt	-43,1	34	-37,9	23
Switzerland	-42,5	35	-39,6	46
Denmark	-41,8	36	-39,9	39
Germany	-40,0	37	-40,3	20
India	-39,2	38	-34,3	12
Mexico	-38,2	39	-36,7	38
HongKong	-37,1	40	-19,0	43
Luxembourg	-35,1	41	-22,3	28
China	-35,0	42	-15,7	45
Singapore	-34,5	43	-34,1	49
Brazil	-33,8	44	-26,6	50
Japan	-33,7	45	-32,3	48
Indonesia	-33,6	46	-36,7	44
Canada	-31,6	47	-25,1	47
Taiwan	-31,5	48	-29,5	41
Australia	-29,3	49	-30,0	51
Colombia	-29,0	50	-29,6	40
Chile	-26,1	51	-22,0	42
Tunisia	-22,6	52	-3,2	53
Qatar	-18,7	53	-15,0	54
Israel	1,8	54	-4,2	52

Conclusions and further research

This thesis' primary goal was to model, estimate and explain liquidity commonality in equity markets. The thesis started with the exploration of traditional supply- and demand-side determinants of the correlation between execution costs over a long time series. Then, we moved to the analysis of high-frequency data, proposing high-frequency quoting (HFQ) as a new supply-side explanation for liquidity commonality. Motivated by the importance of factor models in our research, we then proposed a factor model with time-varying loadings and analysed its implications for stock returns comovements.

In Chapter 1, we studied the impact of supply-side variables (funding liquidity of intermediaries) and demand-side variables (trading behaviour of investors) on liquidity commonality using a novel factor model where loadings and volatilities are time-varying, and liquidity is exposed to global, regional and sectoral factors. We define three types of commonality driven by either loadings or variances or both. Using weekly data on 1909 firms from the US, Japan, the UK and Euro zone countries, from January 2000 to January 2017, we found that the common drivers of the cross-section of liquidity are mainly global and region-specific, while liquidity of stocks in the same sector show little tendency to co-move. Furthermore, we found that both demand and supply shocks play a role in explaining liquidity commonality. On the supply side, when funding constraints are binding, liquidity commonality increases, in line with Brunnermeier and Pedersen (2009) and Hameed et al. (2010). On the demand side, index-related trading (proxied by ETF trading volume) is positively related to liquidity commonality, supporting the results of Kamara et al. (2008)

and Koch et al. (2016). We found that the correlated liquidity demand of institutional investors is the strongest regional economic force that makes stock liquidity co-move. The effects are less consistent in the supply-side.

In Chapter 2, we investigated the relationship between HFT activity and liquidity commonality. Our sample comprises all trades and best quotes' updates for the FTSE100 stocks from January 2010 to December 2011, traded on the London Stock Exchange. An upgrade of the trading systems of the London Stock Exchange on February 2011 is used to identify an exogenous positive shock to HFT. We found, both using an eigenvalues analysis and a panel data analysis, that liquidity commonality increased after the introduction of the Millennium Exchange and this result is robust to different measures of liquidity. Furthermore, we analysed the intraday relationship between HFT's liquidity provision activity, measured by a type of Order-to-Trade ratio, and other microstructure variables. In general, we found that HFTs increase their participation from market opening until 12pm, when stocks are most illiquid and market makers are needed the most. We found that intraday RV has a U-shaped pattern. The high volatility at the end of the day together with the evidence that quoted spread are very narrow is puzzling but it can be reconciled with the same U-shaped pattern in the size of the order flow. Finally, we found that liquidity providers absorb demand/supply shocks more efficiently at the end of the day, which could be due to the use of marketable orders.

In Chapter 3, we investigated the dynamics of the systematic risk in a large portfolio of firms from 54 countries. We proposed a two-level factor model with time-varying loadings that captures financial, global and regional risk to estimate common components in stock returns from January 2006 to January 2016. The factors are latent and estimated via principal component analysis. Using canonical correlation analysis, we found that the estimated factors are linear combinations of the Fama-French three factors, which leads us to conclude that our model is correctly capturing the risk at which firms are exposed. We found that the relative importance of unobserved regional and global factors is time-varying, with the global one becoming more relevant when firms are exposed to global shocks. Furthermore, the dynamics of factor loading are related to the profile of a company. We found that bigger firms have larger exposure to financial and regional common factors, while there is no clear difference across global factor loading quantiles. This evidence is in line with the finding of Fama and French (2017) that a global version of their factor model is not able to price the cross-section of stock returns. We found that expected returns are increasing in the variance of financial and global factor loadings, while they are decreasing in the variance of the regional factor loading. The decreasing relationship is in line with the finding of Armstrong et al. (2013) for US stocks. However, our model suggests that there is a premium for holding stocks whose global systematic risk is very volatile. Furthermore, expected returns are decreasing in the persistence of financial and global factor loadings, implying that there is no premium for holding firms with highly persistent factor exposures.

Future research can be developed in various directions. In Chapter 1 we introduced three versions of liquidity commonality. It would be interesting to test if exposure-driven commonality is related to permanent shifts in co-movements and volatility-driven commonality related to temporary. We expect VD to be related to short-term variations because factor volatility returns to its long run mean, while the product of two stationary variables (the factor loadings) does not have to be stationary. Various authors have associated changes in factor loadings as permanent changes in comovement statistics and changes in factor volatilities as temporary [Bekaert et al. (2014) and Dungey and Renault (2017)].

In Chapter 2 we analysed the order-driven market of the London Stock Exchange and our results suggest that HFT activity increases liquidity commonality. This could be explained either by informational reasons [Cespa and Foucault (2014)] or by lack of capital [Gromb and Vayanos (2002)]. Interestingly, we think that these two mechanisms work at different frequencies, i.e. it is hard to think of reasons why the funding costs of market makers should vary at high frequencies, but these can still play a role at lower frequencies. Our framework allows us to directly test the informational channel of Cespa and Foucault (2014) on high-frequency data. In particular, we have identified a computer glitch (on 25th February 2011) that can be used as an unexpected shock to the flow of information, which can be used to measure the resilience of liquidity commonality and the speed at which it reverts to its long-run level. Furthermore, since the European equity markets are very fragmented, it would be interesting to measure liquidity commonality using the order books provided by competing trading venues (e.g. BATS, Chi-X Europe, Turquoise). It would be interesting both to compare the commonality within exchanges and to compute the explanatory power of common factors across exchanges.

The results of Chapter 3, in particular the connection between the dynamics of factor loadings and the profile of the firm, naturally extend to an out-of-sample asset pricing test. The estimation need to be conditioned at time t, when a long-short portfolio can be formed using signals based on the persistence or the variance of the factor loadings. It would also be interesting to build an "enhanced" beta strategy with a double sorting, where we long stocks with low beta and low variance of beta. Finally, the model of Chapter 3 can be also estimated on US data only, assuming a factor structure featuring US and sector (instead of regional) factors.

Bibliography

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77*(2), 375–410.

Adrian, T. and H. S. Shin (2008). Financial intermediaries, financial stability, and monetary policy. *Federal Reserve Bank of New York Staff Reports 346*.

Adrian, T. and H. S. Shin (2010). Liquidity and leverage. *Federal Reserve Bank of New York Staff Reports 328*.

Aït-Sahalia, Y. and D. Xiu (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, Forthcoming.

Alexander, C. (2001). *Mastering Risk Volume II (C.O. Alexander, Ed.)*. Financial Times-Prentice Hall.

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *Journal of Finance 57*(3), 1047–1091.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets 5*(1), 31–56.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*(2), 223–249.

Anand, A. and K. Venkataraman (2016). Market conditions, fragility, and the economics of market making. *Journal of Financial Economics 121*(2), 327–349.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and G. Wu (2006). Realized beta: Persistence and predictability. *Advances in Econometrics 20*(1), 1–39.

Ando, T. and J. Bai (2015). Asset pricing with a general multifactor structure. *Journal of Financial Econometrics 13*(3), 556–604.

Ando, T. and J. Bai (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association 112*(519), 1182–1198.

Ang, A., J. Liu, and K. Schwarz (2017). Using stocks or portfolios in tests of factor models. *Available at SSRN: https://ssrn.com/abstract=1106463*.

Armstrong, C. S., S. Banerjee, and C. Corona (2013). Factor-loading uncertainty and expected returns. *The Review of Financial Studies 26*(1), 158.

Baca, S. P., B. L. Garbe, and R. A. Weiss (2000). The rise of sector effects in major equity markets. *Financial Analysts Journal 56*(5), 34–40.

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

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 61*(4), 1645–1680.

Bali, T., L. Peng, Y. Shen, and Y. Tang (2014). Liquidity shocks and stock market reactions. *Review of Financial Studies 27*(5), 1434–1485.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

Bates, B. J., M. Plagborg-Moller, J. H. Stock, and M. W. Watson (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics 177*(2), 289–304.

Beaupain, R., P. Giot, and M. Petitjean (2010). Volatility regimes and liquidity co-movements in cap-based portfolios. *Revue de l'association française de finance 31*(1), 55–79.

Bekaert, G., M. Ehrmann, M. Fratzscher, and A. Mehl (2014). The global crisis and equity market contagion. *Journal of Finance 69*(6), 2597–2649.

Bekaert, G., R. J. Hodrick, and X. Zhang (2009). International stock return comovements. *Journal of Finance 64*(6), 2591–2626.

Belvisi, M., R. Pianeti, and G. Urga (2016). Modelling financial markets comovements during crises: A dynamic multi-factor model approach. *Advances in Econometrics Vol. 35, Dynamic Factor Models 35*, 317–360.

Berk, J. B., R. C. Green, and V. Naik (1999). Optimal investment, growth options, and security returns. *The Journal of Finance 54*(5), 1553–1607.

Boehmer, E., K. Y. L. Fong, and J. Wu (2013). Algorithmic trading and changes in firms' equity capital. *http://ssrn.com/abstract=2348730*.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Bollerslev, T. and B. Y. B. Zhang (2003). Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance 10*(5), 533–558.

Brandt, M. W., A. Brav, J. R. Graham, and A. Kumar (2010). The idiosyncratic volatility puzzle: Time trend or speculative episodes? *Review of Financial Studies 23*(2), 863–899.

Breitung, J. and S. Eickmeier (2015). Analyzing business cycle asymmetries in a multi-level factor model. *Economics Letters 127*, 31–34.

Brockman, P. and D. Y. Chung (2002). Commonality in liquidity: Evidence from an order-driven market structure. *Journal of Financial Research 25*(4), 521–539.

Brockman, P., D. Y. Chung, and C. Pérignon (2009). Commonality in liquidity: A global perspective. *Journal of Financial and Quantitative Analysis 44*(4), 851–882.

Brogaard, J., T. Hendershott, S. Hunt, and C. Ysusi (2014). High-frequency trading and the execution costs of institutional investors. *Financial Review 49*(2), 345–369.

Brogaard, J., T. Hendershott, and R. Riordan (2014). High-frequency trading and price discovery. *Review of Financial Studies 27*(8), 2267–2306.

Brooks, R. and M. Del Negro (2005). Country versus region effects in international stock returns. *Journal of Portfolio Management 31*(4), 67–72.

Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics and Data Analysis 51*(4), 2232–2245.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies 22*(6), 2201–2238.

Carlson, M., A. Fisher, and R. Giammarino (2006). Corporate investment and asset price dynamics: Implications for SEO event studies and long-run performance. *The Journal of Finance 61*(3), 1009–1034.

Carlson, M., A. Fisher, and R. Giammarino (2010). SEO risk dynamics. *The Review of Financial Studies 23*(11), 4026–4077.

Cavaglia, S., C. Brightman, and M. Aked (2000). The increasing importance of industry factors. *Financial Analysts Journal 56*(5), 41–54.

Cespa, G. and T. Foucault (2014). Illiquidity contagion and liquidity crashes. *Review of Financial Studies 27*(6), 1615–1660.

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in liquidity. *Journal of Financial Economics 56*(1), 3–28.

Chordia, T., A. Sarkar, and A. Subrahmanyam (2005). An empirical analysis of stock and bond market liquidity. *Review of Financial Studies 18*(1), 85–129.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal 14*(1), C45–C90.

Cont, R., A. Kukanov, and S. Stoikov (2014). The price impact of order book events. *Journal Of Financial Econometrics 12*(1), 47–88.

Cooper, I. and R. Priestley (2011). Real investment and risk dynamics. *Journal of Financial Economics 101*(1), 182–205.

Corwin, S. A. and M. L. Lipson (2011). Order characteristics and the sources of commonality in prices and liquidity. *Journal of Financial Markets 14*(1), 47–81.

Coughenour, J. F. and M. Saad, Mohsen (2004). Common market makers and commonality in liquidity. *Journal of Financial Economics 73*(1), 37–69.

Del Negro, M. and C. Otrok (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *Staff report 326, Federal Reserve Bank of New York, New York*.

Domowitz, I., O. Hansch, and X. Wang (2005). Liquidity commonality and return co-movement. *Journal of Financial Markets 8*(4), 351–376.

Dungey, M. and L. Martin, Vance (2007). Unravelling financial market linkages during crises. *Journal of Applied Econometrics 22*(1), 89–119.

Dungey, M. and E. Renault (2017). Identifying contagion. *Journal of Applied Econometrics*, forthcoming.

Eickmeier, S., W. Lemke, and M. Marcellino (2015). Classical time varying factor-augmented vector auto-regressive models: estimation, forecasting and structural analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(3), 493–533.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics 20*(3), 339–350.

Fama, E. F. and K. R. French (1997). Industry costs of equity. *Journal of Financial Economics 43*(2), 153–193.

Fama, E. F. and K. R. French (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics 123*(3), 441.

Fan, J., A. Furger, and D. Xiu (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business and Economic Statistics 34*(4), 489–503.

Ferson, W. E., S. Kandel, and R. F. Stambaugh (2002). Where is beta going? the riskiness of value and small stocks. *Unpublished PhD dissertation, MIT*.

Flood, R. P. and A. K. Rose (2005). A new approach to asset integration: Methodology and mystery. *Pacific Economic Review 10*(1), 5–28.

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics 82*(4), 540–554.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica 84*(3), 985–1046.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1992). Stock prices and volume. *Review of Financial Studies 5*(2), 199–242.

Gatev, E. and P. E. Strahan (2006). Banks' advantage in hedging liquidity risk: Theory and evidence from the commercial paper market. *Journal of Finance 61*(2), 867–892.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *The Journal of Finance 53*(2), 549–573.

Gorton, G. B. and G. G. Pennacchi (1993). Security baskets and index-linked securities. *Journal of Business 66*(1), 1–27.

Goyal, A., C. Pérignon, and C. Villa (2008). How common are common return factors across the NYSE and nasdaq? *Journal of Financial Economics 90*(3), 252–271.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics 92*(2), 153–181.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific? *The Review of Financial Studies 15*(3), 783–803.

Gromb, D. and D. Vayanos (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics 66*(2), 361–407.

Hackbarth, D. and E. Morellec (2008). Stock returns in mergers and acquisitions. *The Journal of Finance 63*(3), 1213–1252.

Hafner, C. M. and A. Preminger (2009). Asymptotic theory for a factor garch model. *Econometric Theory 25*(2), 336–363.

Hallin, M., C. Mathias, H. Pirotte, and D. Veredas (2011). Market liquidity as dynamic factors. *Journal of Econometrics 163*(1), 42–50.

Halling, M., J. Yu, and J. Zechner (2016). Leverage dynamics over the business cycle. *Journal of Financial Economics 122*(1), 21.

Hameed, A., W. Kang, and S. Viswanathan (2010). Stock market declines and liquidity. *Journal of Finance 65*(1), 257–293.

Harvey, A. C. (1990). *The econometric analysis of time series* (2nd ed.). Philip Allan.

Hasbrouck, J. and G. Saar (2009). Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets 12*(2), 143–172.

Hasbrouck, J. and G. Saar (2013). Low-latency trading. *Journal of Financial Markets 6*(4), 646–679.

Hasbrouck, J. and D. J. Seppi (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics 59*(3), 383–411.

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *Journal of Finance 66*(1), 1–33.

Heston, S. L. and K. G. Rouwenhorst (1995). Industry and country effects in international stock returns. *Journal of Portfolio Management 21*(3), 53–58.

Hirata, H., K. M. and C. Otrok (2013). Regionalization vs. globalization. *IMF Working Paper 13/19*.

Huberman, G. and D. Halka (2001). Systematic liquidity. *Journal of Financial Research 24*(2), 161–178.

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics 196*(1), 55–67.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer-Verlag.

Kamara, A., X. Lou, and R. Sadka (2008). The divergence of liquidity commonality in the cross-section of stocks. *Journal of Financial Economics 89*(3), 444–466.

Karolyi, G. A., K.-H. Lee, and M. A. van Dijk (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics 105*(1), 82–112.

Kim, S. and D. Murphy (2013). The impact of high-frequency trading on stock market liquidity measures. *http://ssrn.com/abstract=2278428*.

Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun (2017). The flash crash: Highfrequency trading in an electronic market. *The Journal of Finance 72*(3), 967–998.

Koch, A., S. Ruenzi, and L. Starks (2016). Commonality in liquidity: A demand-side explanation. *Review of Financial Studies 29*(8), 1943–1974.

Korajczyk, R. A. and R. Sadka (2008). Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics 87*(1), 45–72.

Krishnamurthy, A. (2002). The bond/old-bond spread. *Journal of Financial Economics 66*(2), 463–506.

Kyle, Albert, S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1336.

Lee, K. (2011). The world price of liquidity risk. *Journal of Financial Economics 99*(1), 136–161.

Lewellen, J. and S. Nagel (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics 82*(2), 289–314.

Linton, O., M. O'Hara, and J.-P. Zigrand (2012). Economic impact assessments on mifid ii policy measures related to computer trading in financial markets. *Future of computer trading: working paper, Foresight, Government Office for Science, London, UK*.

Menkveld, A. (2013). High frequency trading and the new-market makers. *Journal of Financial Markets 16*(4), 712–740.

Menkveld, A. J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics 8*(1), 1–24.

Mikkelsen, G., J. (2017). Testing for time-varying loadings in dynamic factor models. *CREATES Research Paper 2017-22*.

Mikkelsen, G., J., E. Hillebrand, and G. Urga (2015). Maximum likelihood estimation of time-varying loadings in high-dimensional factor models. *CREATES Research Paper 2015-61*.

Mikkelsen, G., J., E. Hillebrand, and G. Urga (2017). Mamum likelihood estimation of time-varying loadings in high-dimensional factor models. *Manuscript, CREATES*.

Moshirian, F., X. Qian, C. Wee, and B. Zhang (2017). The determinants and pricing of liquidity commonality around the world. *Journal of Financial Markets 33*, 22–41.

Muravyev, D. and N. Pearson (2013). Is there price discovery in equity options? *Journal of financial economics 107*, 259–283.

Mykland, P. and L. Zhang (2016). Between data cleaning and inference: Pre-averaging and robust estimators of the efficient price. *Journal of Econometrics 194*(2), 242–262.

Ng, S. and P. Perron (1995). Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association 90*(429), 268–281.

Pastor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Patton, A. J. and M. Verardo (2012). Does beta move with news? Firm-specific information flows and learning about profitability. *Review of Financial Studies 25*(9), 2789–2839.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies 22*(1), 435–480.

Riordan, R., A. Storkenmaier, and M. Wagener (2011). European market integrity: Regulating equity trading in fragmented markets. *Working paper*.

Robichek, A. A. and R. A. Cohn (1974). The economic determinants of systematic risk. *The Journal of Finance 29*(2), 439–447.

Rosenberg, B. and J. Guy (1995). Prediction of beta from investment fundamentals. *Financial Analysts Journal 51*(1), 101.

Santos, T. and P. Veronesi (2004). Conditional betas. *Available at SSRN: http://ssrn.com/abstract=649462*.

Savor, P. and M. Wilson (2016). Earnings announcements and systematic risk. *The Journal of Finance 71*(1), 83–138.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business and Economic Statistics 7*(2), 147–159.

Shanken, J. (1990). Intertemporal asset pricing - an empirical investigation. *Journal of Econometrics 45*(1), 99–120.

Stiglitz, J. E. (2014). Tapping the brakes: Are less active markets safer and better for the economy? *Paper prepared for presentation at Atlanta Federal Reserve Conference*.

Stock, J. H. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instability. *Hendry, D.F., Castle, J., Shephard, N. (Eds.), The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry, Oxford University Press, Oxford.*, 173–205.

Tauchen, G., H. Zhang, and M. Liu (1996). Volume, volatility, and leverage: A dynamic analysis. *Journal of Econometrics 74*(1), 177–208.

van der Weide, R. (2002). Go-garch: A multivariate generalized orthogonal garch model. *Journal of Applied Econometrics 17*(5), 549–564.

Vanhatalo, E. and M. Kulahci (2016). Impact of autocorrelation on principal components and their use in statistical process control: Autocorrelation and pca-based spc. *Quality and Reliability Engineering International 32*(4), 1483–1500.

Wang, P. (2010). Large dimensional factor models with a multi-level factor structure. *Mimeo, Hong Kong University of Science and Technology*.

Westerlund, J. and J. Urbain (2015). Cross-sectional averages versus principal components. *Journal of Econometrics 185*(2), 372–377.