Dynamic models for financial and sentiment time series

Danilo Vassallo

Scuola Normale Superiore

Classe di Scienze

SCUOLA
NORMALE
SUPERIORE

PhD Thesis

Tesi di perfezionamento in Matematica per la Finanza

Internal supervisor:

Prof. Dr. Stefano Marmi

Supervisors:

Prof. Dr. Giacomo Bormetti

Prof. Dr. Fabrizio Lillo

Contents

Contents **i**

Acknowledgements **iv**

List of publications **vi**

1 Introduction **1**

2 Score-driven models and Kalman Expectation Maximization **21**

2.1 Kalman filter and Expectation Maximization 23

2.2 Score-driven models 29

3 Disentangling short-run and long-run components in sentiment dynamics **36**

3.1 Introduction 36

3.2 The Model 42

3.3 Data 50

3.4 Empirical analysis 51

3.5 Contemporaneous and lagged relations 61

3.6 Portfolio allocation with sentiment data 69

3.7 Conclusions 77

4 Modeling Realized Covariance with score-driven dynamics 79

4.1 Introduction 79

4.2 Framework 85

4.3 Monte-Carlo analysis 94

4.4 Empirical evidence 102

4.5 Conclusions 122

5 Asset allocation and Portfolio Instability in the Low for Long environment 124

5.1 Introduction 124

5.2 Framework 126

5.3 The role of monetary policy in the portfolio instability . . . 130

5.4 Risk Parity Investor 144

5.5 Conclusions 151

6 Conclusive remarks 153

Bibliography 155

A Appendix of Chapter 3 173

A.1 List of stocks 173

A.2 Signal-to-noise ratio and comparison with MLNSL 175

A.3 Quantile regression: Contemporaneous effects 176

A.4 Market absortion of news 178

A.5 Portfolio allocation on February 2007-June 2017 without trading costs 182

A.6 Robustness check: Portfolio allocation on July 2017 - December 2019 183

B Appendix of Chapter 4 **185**

B.1 Statistical significance of the sentiment portfolios 185

B.2 Computation of the scaled score in the univariate models . . 188

B.3 Proposition 1 190

B.4 Proposition 2 192

B.5 Proposition 3 193

B.6 Proposition 4 193

B.7 Proposition 5 195

B.8 Proposition 6 195

B.9 Figures of Section (4.3) 195

C Appendix of Chapter 5 **199**

C.1 Proof Propositions 1 and 2 199

C.2 Econometric model with US Data 201

Acknowledgements

I would first like to thank my supervisors, Giacomo Bormetti and Fabrizio Lillo, whose guidance was invaluable. You helped me to always challenge my ideas and produce better results. Thanks to my coauthors Giuseppe Buccheri, Fulvio Corsi, Lieven Hermans and Thomas Kostka, your expertise helped me to sharpen my thinking and brought my work to a higher level. A huge thanks go to the head of the PhD program, Stefano Marmi, for giving me the opportunity to live this wonderful adventure.

I spent these last years in Pisa and Frankfurt and in both cities, I have found competent colleagues and great friends.
Thanks to Carlo, Giulia, Giuseppe, Maddalena, Tonio, Mateusz, Gael, Domenico and Daniele who helped me with their friendship. Thanks to my office mates Elisa, Clemente, Francesco and Giacomo for the great discussions that we had every morning.
I would also like to thank my colleagues in Frankfurt. Christian and Sante, I couldn't ask for better office mates and friends. Thanks also to all the other colleagues in MBF for inspiring me with their guidance and passion.

A huge thanks to my friends in Torino and my family for always being of invaluable support, in particular, I want to thank my lamented grandfather Evaristo. You taught me to love questions themselves and to challenge my beliefs.

Finally, my greatest luck has been having you Roberta by my side through good times and bad during all these years. I will never be grateful enough.

Pisa, 2021

List of publications

The following list includes all the publications that I have produced during the last 5 years as a Ph.D. student

1. Vassallo, D., Buccheri, G., Corsi, F., 2021. A DCC-type approach for Realized Covariance modelling with score-driven dynamics. *International Journal of Forecasting.*

2. Vassallo, D., Bormetti, G., Lillo, F., 2019. A tale of two sentiment scales: Disentangling short-run and long-run components in multivariate sentiment dynamics. *Available at SSRN: https://ssrn.com/abstract=3463691.*

3. Grothe, M., Helmersson, T., Quint, D., Vassallo, D., 2021. Risk of spillovers from US equity market corrections to euro area markets and financial conditions. *Box 3 of May 2021 Financial Stability Review. Available here*

4. Giuzio, M., Moldovan, C., Vassallo, D., 2020. Financial stability implications of private equity. *Box 1 of May 2020 Financial Stability Review. Available here*

5. Vassallo, D., Hermans, L., Kostka, T., 2020. Volatility-targeting strategies and the market sell-off. *Box 2 of May 2020 Financial Stability Review. Available here*

Chapter 1

Introduction

Asset allocation and investment decision are two of the main studied topics in finance and financial mathematics. In this thesis, I tackle some of the problems associated with the three main ingredients of portfolio allocation theory, i.e. conditional expected returns, conditional variance, and the risk-free asset from three different perspectives. The first ingredient is crucial in investment decisions since the conditional expected returns represent the expected profits coming from an investment, however, they are extremely difficult to predict and the shorter the investment horizon the harder is the prediction of future returns. Chapter 3 presents a new decomposition of the sentiment series to enrich the investors' information set and improve the forecasts of daily returns. The second ingredient represents one of the measures of risk used by investors to make their investment decisions. In addition, the conditional variance may be useful to compute more complex measures of risk as to the Value at Risk and the Expected Shortfall. Chapter 4 presents a new class of models to better forecast conditional covariance matrices of financial assets The third and last ingredient is used by investors to compare the performance of their investments with an investment free

of risk. The risk-free asset is an investment where the payoff is known in advance and it is not subject to any source of risk. This ingredient is exogenous to investors and strongly depends on the central bank monetary policy. This link is explained in Chapter 5 of the thesis and we show how the central bank monetary policy may affect the investors' decisions.

The first example of a quantitative investment approach that uses the mentioned three ingredients is the Mean-Variance portfolio allocation of Markowitz (1952), Sharpe (1966) and Merton (1972a). This one-period investment framework can be generalised in a multi-period framework by solving a sequence of single-period problems, where the optimal weights for any given time t are chosen such that the *conditional expected return* of the portfolio is maximized for a given level of *conditional variance* (Fama, 1970; Hakansson, 1971).

Sentiment analysis and asset returns.

Since the beginning of the financial markets, investors used public and private information to invest or disinvest in a specific stock. *News*, defined as the difference of the expectations in two subsequent times, was indicated as the main drivers in the changes in stock prices and volatilities. Campbell and Shiller (1988b) and Campbell and Shiller (1988a) theoretically showed how unexpected returns are a combination of news on dividends and future returns. Campbell and Hentschel (1992) and Engle and Ng (1993) showed how the volatility reacts asymmetrically to news.

In recent years, with the advent of big data and the increment of computational capacity, alternative datasets have been explored to assess if they contain market sensible information. In addition, the introduction of a new brand of machine learning techniques, called *Natural Language Processing*,

helped researchers and investors to categorize the *sentiment* of news, i.e. the polarity of information contained in a message. This approach allows us to analyze a huge number of newspapers and social media and automatically extract the mood of the investors.

Using these new datasets, researchers explore the value added by the information flow in the investor's information set. If the article in today's Financial Times affects the investor's behaviour, we should expect that the content of the article can affect today's realized return. More interestingly, if the market does not digest the news immediately, we can use today's information flow to predict future market movements. This lead-lag effect may be present in the market but it is not easy to exploit. Indeed the flow of news arrival is not homogeneous in time and we may not observe news for a given company for hours or days. The higher is the frequency considered, the stronger is the lead-lag effect but the higher is the number of missing observations in the sentiment time series. On the opposite, if a lower time scale is used, the sentiment time series is more tractable but the lead-lag effect between sentiment and return is weaker. Antweiler and Frank (2004) showed that the information contained in the messages posted online aggregated at a daily level is useful to statistically predict future returns, volatilities and volumes. Allen et al. (2015) and Audrino et al. (2020) showed that the sentiment and volumes of newspaper articles and tweets are important to predict market volatility. Tetlock (2007) and Garcia (2013) showed that the Dow Jones returns can be predicted using sentiment variables and that the forecasting power is stronger during a crisis.

In the economics and financial literature using sentiment analysis, the textual data used depends on two choices of the authors. The first choice is the definition of the interested sentiment variable and what is the metric

used to classify textual data, where the appropriate quantitative technique depends on the context. For instance, Antweiler and Frank (2004) and Picault and Renault (2017) use the same classification technique but they target different sentiment variables. The first paper classifies the sentiment data as bullish and bearish as it is referred to the stock market, while the second paper denotes the sentiment variable as dovish and hawkish as they study the ECB communications. The classification technique used in these two papers is based on the count of *positive* or *negative* words, but the definition of *positive* and *negative* words changes according to the context.

The second choice is the sources of news that the authors want to analyze. For instance, Tetlock (2007) uses only the articles from "Abreast of the Market" column in the Wall Street Journal while Peterson (2016) uses a very broad and diverse number of sources, ranging from the Financial Times to Twitter. When a low number of homogeneous sources are used, the sentiment signal represents the view of a small sample of individuals and it may be not representative of the population. At the same time, homogeneous sources tend to reflect the same opinions and there is a small dispersion in their views. This case represents a high-bias low-variance measure of sentiment. On the contrary, when a high number of heterogeneous sources are used, the signal is more representative of the view of the population. However different groups may have very different views on the same news, for this reason, the dispersion around the mean is higher than in the previous case. In this scenario, we have a low bias high variance measure of sentiment. The number and diversity of the used sources present a bias-variance tradeoff. When the number of sources is high and the sentiment series are affected by noise, econometric techniques have been proposed to filter the noise. Peterson (2016) uses the Moving

Average Convergence-Divergence methodology proposed in Appel (2003) and Borovkova and Mahakena (2015) Audrino and Tetereva (2019) and Borovkova and Lammers (2017) introduce the Local News Sentiment Level model (LNSL), a univariate method which takes inspiration from the Local Level model of Durbin and Koopman (2012). However, there is not a clear consensus of which technique should be used and how these filtering approaches can be interpreted.

Chapter 3 presents the first original work of the thesis. We contribute to the literature by proposing a new econometric technique to filter the noise from the sentiment data and we investigate the relationship between the sentiment series and market returns. In the first part of chapter 3, we model the noisy sentiment series as a composition of three components, one noisy component and two sentiment signals. The first sentiment signal, named *long-term sentiment*, captures the permanent shocks in the investor's mood caused by certain news. In Borovkova et al. (2017) and Audrino and Tetereva (2019), the long-term sentiment is the only relevant sentiment component. The second sentiment signal, named *short-term sentiment*, captures the short-lived trends in the investor's mood. This component is inspired by Gerber et al. (2011) who shows how these fast trends in the investor's preferences have strong impacts on the investor's behaviour. The second novelty of our approach is the multivariate nature of the model. While Audrino and Tetereva (2019) uses a univariate specification to filter the noise from the sentiment series, we propose a multivariate approach that exploits the cross-sectional information of the sentiment. In the empirical analysis, we see that the cross-correlation among the short-term sentiment component has the same structure as the cross-correlation component of asset returns and residuals. The *Multivariate Long Short Sentiment model*

(MLSS) reads

$$\begin{aligned} S_t &= \Lambda F_t + \Psi_t + \epsilon_t, \quad & \epsilon_t &\overset{d}{\sim} \mathcal{N}(0, R), \\ \Psi_t &= \Phi \Psi_{t-1} + u_t, & u_t &\overset{d}{\sim} \mathcal{N}(0, Q_{short}), \\ F_t &= F_{t-1} + v_t, & v_t &\overset{d}{\sim} \mathcal{N}(0, Q_{long}), \end{aligned} \tag{1.1}$$

where $R \in \mathbb{R}^{K \times K}$ is the diagonal covariance matrix of the observation noise ϵ_t, $\Phi \in \mathbb{R}^{K \times K}$ is the diagonal matrix of autoregressive coefficients, $Q_{short} \in \mathbb{R}^{K \times K}$ is the covariance matrix of the short-term sentiment innovations, and $Q_{long} \in \mathbb{R}^{q \times q}$ is the covariance matrix of the long-term sentiment innovations. In equation (1.1), F_t and Ψ_t are the latent processes of the long-term sentiment and short-term sentiment respectively. It's worth noticing that the long-term sentiment is weighted for a factor loading matrix $\Lambda \in \mathbb{R}^{q \times K}$ where $q \leqslant K$. The intuition behind the factor structure of the long-term sentiment is that the permanent shocks in sentiment are common across set of assets. Empirically, the significance of Λ can be statistically tested and the selection of the number q of common factors can be performed by means of AIC and BIC criteria. We show that the optimal number of common long-term sentiment series is 2.

We decompose and identify this three components using a combination of Kalman filter and Expectation Maximization as in Shumway and Stoffer (1982), Banbura and Modugno (2014) and Jungbacker and Koopman (2008) that allows us to jointly filter the sentiment components and estimate the parameters of model (1.1).

In the second part of the Chapter, we show that this new decomposition is of practical use to shed light on the relationship between sentiment and market returns. We use sentiment and market daily data from January 2006 to December 2017 on a subset of the *Dow Jones Industrial Index* and perform a series of quantile regressions with different filtering tech-

niques. We consider the MLSS model in (1.1), the univariate version of MLSS model (LSS), the *Local News Sentiment Level* (LNSL) of Audrino and Tetereva (2019) and Borovkova et al. (2017) and the multivariate version of the LNSL model (MLNSL). As a point of comparison, we consider also the case, denoted as *Obs*, where no filtering is applied to the data. The goal of the analysis is twofold. First, we analyse the relationship between the sentiment and asset returns distribution. Second, we investigate if the different filtering techniques affect the results. We show that news affects and predict better extreme returns, i.e. returns over the 95th or below the 5th quantiles, rather than returns in the centre of the distribution. In particular, we show that the explanatory and predictable power is mainly given by the novel short-term sentiment. Indeed when the LNSL and MLNSL filtering models are considered, which are the models which do not take into account the short-term sentiment, the statistical effect disappears. These results confirm the initial intuition that fast trends affect the future performances of an asset more than long-lasting trends, which is in line with Gerber et al. (2011). To conclude the second part of the chapter, we investigate the inverse relation between returns and future sentiment. Using again a quantile regression analysis, we see that the effect of present returns and the future sentiment is significant for extreme events, both positive and negative. Intuitively, the result shows that newspaper and social media slowly digests days of extreme returns, where the significance lasts for up to one week. The effect is stronger for extreme negative returns than for extreme positive returns. These findings suggest that the relationship between sentiment and market returns is on both time directions and non-linear, where only extreme parts of the return distribution are affected by the news.

Measures	BH	MLSS	LSS	MLNSL	LNSL	Obs
A. Sharpe ratio	0.469	**0.519**	0.385	0.495	0.419	0.35
A. Sortino ratio	0.578	**0.671**	0.487	0.614	0.53	0.435
Number of trades	1	553	161	81	73	93
Transaction costs ($)	50	37974	14866	5565	4085	7544

Table 1.1: Performances of the six strategies with transaction cost for the period February 2007 - June 2017. In bold, the best performance per row. BH is the buy-and-hold portfolio, while MLSS, LSS, MLNSL, LSNSL, and Obs correspond to portfolios built from the corresponding model for the sentiment time series.

In the third and last part of the chapter, we exploit the non-linear relation between extreme negative returns and long and short-term sentiment. We use a logistic regression model to predict abnormal negative returns, i.e. those returns which are negative of more than one standard deviation. The investment strategy consists of a long signal, i.e. a signal to buy or keep the asset, when the logistic regression does not predict an abnormal negative return and a short signal, i.e. a signal to buy or keep a short position in the asset when the logistic regression predicts an abnormal negative return. For each filtering technique, we estimate a different logistic regression using the sentiment series as exogenous variables. The performances of the portfolios are compared for all the different filtering techniques and are also compared with a simple Buy and Hold strategy (BH). Finally, transaction costs are considered. Table 1.1 reports the performances of the different strategies according to the *Annual Sharpe ratio* and the *Annual Sortino ratio*. We see that the only sentiment-based strategy which significantly outperforms the Buy and Hold is the strategy based on the Long and Short-sentiment decomposition. In the final part of the Chapter, we show the significance of the results and propose a simple technique to reduce the number of trades in the MLSS strategy which reduces the transaction costs and improves the

performance.

Forecasting the asset covariance matrix. A score-driven approach

The main point raised by Markowitz (1952) is the role of *diversification* in portfolio selection. Markowitz seminal paper opened the avenues for the studies of the second moment of asset returns. Volatilities and correlations, however, are not directly observable in the market and are considered latent. For these reasons, two streams of literature emerge in this context. The first stream explores the different modelling features and the techniques to effectively forecast the covariance matrix. The second stream studies the different estimators for the latent covariance matrix.

Concerning the modelling part, as for the mean of the asset returns, also the variance has been widely studied and exhibits well known stylized facts, see Cont (2001). Three key stylized facts of financial returns are the *conditional and unconditional fat tails*, i.e. extreme events occur much more often than as predicted by Gaussian's distribution both conditionally and unconditionally, the *volatility clustering*, i.e. high volatility events tend to follow high volatility events, and the *slow decay of autocorrelation in volatility*, i.e. volatility shows a very long-memory. On the volatility clustering property of financial returns, Engle (1982) and Bollerslev (1986) propose the first discrete-time model of financial returns where the variance is time-varying and capture the volatility clustering behaviour. ARCH and GARCH models reproduce the unconditional fat tails and the volatility clustering properties but they do not address the conditional non-normality and the long-memory. The conditional fat tail of asset returns is easily solved in Bollerslev (1987) who introduced the t-student distribution to model the

residuals of the ARCH-GARCH models. The problem of the long-memory property has been tackled by several researchers. The most popular long memory models are inspired by the ARFIMA(p, d, q), as the FIGARCH of where FI stands for Fractionally Integrated of Baillie et al. (1996). However, fractionally integrated models are hard to interpret economically and are hard to be extended in higher dimensions. Corsi (2009) proposed a simple model where he approximates the slowly decaying memory process with a restricted $AR(22)$ which perfectly reproduces the long-memory behaviour of volatility with a short memory process.

Covariance modelling introduces several extra technical challenges that are not present in the univariate case. The first one is the *curse of dimensionality*. Indeed, in practical asset allocation problems, the number of considered assets is very high and most of the models have the number of parameters that scale with the number of assets and it can lead to poorly estimated parameters. The second problem is the need to guarantee positive-definite matrices for any time. This condition can be hard to be guaranteed in many models. The first notable example of a model which addresses both problems is the DCC model of Engle (2002b).

Concerning the estimators for the covariance matrix, after the seminal paper Andersen and Bollerslev (1998), the use of high-frequency data became a standard in the financial literature. The *realized* measures have been proved to produce better models with higher forecasting performances. Engle and Gallo (2006), Corsi (2009) and Hansen et al. (2012) are notable examples of univariate models which exploit the daily realized variance measure to improve the forecasts of daily volatility. In the multivariate framework, Bauwens et al. (2012) improves the DCC model presented in Engle (2002b) to account for realized covariance measures. However, most

of the models which use the realized measures do not consider the estimation error of the realized measures or, if they do, they start from the Gaussian assumption which contradicts one of the previously presented stylized facts.

Chapter 4 presents the second original contribution of the thesis. We use a recent class of observation-driven models, named *score-driven models* introduced in Creal et al. (2013) and Harvey (2013), to model and forecast high dimensional covariance matrices. This procedure allows us to create a dynamic model with possibly any distribution. Gorgi et al. (2018) and Opschoor et al. (2017) use this technique to model high dimensional covariance matrix with a score-driven dynamic, where they produce two different models. The *Realized Wishart GARCH* (RWG) of Gorgi et al. (2018) and the *GAS-F* model of Opschoor et al. (2017). However, the parametrization used in these models is often too restrictive to capture the whole dynamics of the covariances.

In the first part of the Chapter, we take inspiration from the DCC model of Engle (2002b) to create a two-step procedure that is more flexible. We start from a series of positive-definite unbiased measures $\{X_t\}_{t=1}^T$ of the latent covariance matrices $\{V_t\}_{t=1}^T$ and we assume that

$$X_t|\mathcal{F}_{t-1} \sim \mathcal{G}_k\left(V_t\right) \tag{1.2}$$

where $\mathcal{G}_k$ is a matrix-variate distribution with expected value $V_t \in \mathbb{R}^{k\times k}$. We separately estimate variances and correlations and, considering a given specification for the return's distribution, we include the estimations of the volatilities to improve the estimation of the correlations' dynamics. The main idea for the decomposition is based on the variance-correlation decomposition $V_t = D_t R_t D_t$. D_t is the $k \times k$ standard deviation matrix for any time t with the standard deviations in the diagonal elements and 0 off-diagonal while R_t is the correlation matrix for any time t. In the first

step, we use the conditional distributions of the diagonal elements $x_t^{(i)}$ of the Realized measure X_t, i.e. $x_t^{(i)}|\mathcal{F}_{t-1} \sim g(v_t^{(i)})$. For any asset i, we estimate the following model

$$\begin{aligned} D_t(i,i) &= \sqrt{v_t^{(i)}} \\ v_t^{(i)} &= e^{\lambda_t^{(i)}} \\ \lambda_{t+1}^{(i)} &= \omega_i + \alpha_i s_t^{(i)} + \beta_i^{(d)} \lambda_t^{(i)} + \beta_i^{(w)} \lambda_{t-1|t-5}^{(i)} + \beta_i^{(m)} \lambda_{t-6|t-22}^{(i)}, \end{aligned} \tag{1.3}$$

where $\lambda_{t-m|t-n}^{(i)} = \frac{1}{n-m+1} \sum_{j=t-n}^{t-m} \lambda_j^{(i)}$ and $s_t^{(i)}$ denotes the scaled score of the i-th univariate density. $s_t^{(i)}$ is the update given by the score-driven models (Creal et al., 2013) and depends on the choice of the probability density function g.

In the second step, we use the estimation of the matrix D_t given by (1.3) to estimate the dynamics of the correlation matrix as

$$\begin{aligned} R_t &= \Delta_t^{-1} Q_t \Delta - 1_t \\ \mathsf{vech}(Q_{t+1}) &= \Omega + A s_t + B \mathsf{vech}(Q_t), \end{aligned} \tag{1.4}$$

where $\Delta_t = \mathsf{diag}\,(Q_t)^{1/2}$ and $\mathsf{vech}(\cdot)$ is the half-vectorization operator. As in the univariate case, the score driven update s_t depends on the density function $\mathcal{G}_k$ in (1.2).

Two possible parametrizations of $\mathcal{G}_k$ are proposed. The first one assumes a Gaussian density for the asset returns which leads to a Wishart distribution for the Realized Covariance matrix. The score-driven specification, in this case, reduces to the Realized DCC model presented in Bauwens et al. (2012). The second specification assumes a student-t distribution for the asset returns, which leads to a matrix-F distribution for the realized covariance matrix. In this case, we obtain a different and novel update rule. We name the models as 2-*step* W and 2-*step* F when the Wishart and matrix-F parametrizations are assumed respectively. We prove the positive

definiteness of the covariance matrices estimated with both models which also are not severely affected by the curse of dimensionality. In addition we show that the 2-*step* W is a special case of the 2-*step* F under some parameter restrictions.

The proposed models satisfy the stylized facts presented above, the only exception is that the Wishart specification does not account for the conditional fat tails while the matrix-F distribution does. This characteristic allows us to test whether and to which extent the conditional fat tailless affects the fit and forecast. In addition, both the specifications account for possible estimation errors of the realized covariance matrix.

The two new models present potential advantages with respect to the existing ones. The 2-*step* W *(*F*)* has a more flexible parameterization with respect to the score-driven models of Gorgi et al. (2018) and Opschoor et al. (2017). Indeed the 2-*step* W *(*F*)* approximate the long-term memory using a restricted AR(29) inspired by Corsi (2009) and capture the non-linearity of the volatility processes using an exponential link function. In comparison with the other multivariate models inspired by the DRD decomposition, the score-driven dynamics of the 2-*step* W *(*F*)* create a consistency between the univariate and multivariate distributions, where the univariate models follow the marginal distribution of the multivariate specification.

In the second part of the Chapter, we perform an extensive Monte Carlo exercise where we show three important features of the models. i) The maximum-likelihood estimator of the two-step routine has an increasing power with the cross-asset dimension and with the number of observations. In addition, the parameters can be accurately estimated even in low-dimension and/or with few data points. The finite-sample properties of the estimators are crucial to understanding the behaviour of the models

for different dimensions of the time series. ii) Both models can accurately capture the correlation dynamic also when the Data Generating Process is misspecified. This feature is crucial to understand whether the models are flexible enough to capture the correlation dynamics of real financial data. iii) We compare the performances of the 2-step model with respect to the joint models of Gorgi et al. (2018) and Opschoor et al. (2017) according to the *Root Mean Square Error* (RMSE) and the *Quasi-Likelihood* (Qlike) measures. The Diebold-Mariano test shows that the 2-step estimation models significantly outperform the joint estimation models both in and out of sample.

In the third and last part of the Chapter, we report the empirical exercise using 5 different portfolios from the NYSE (with 5, 10, 25, 50 and 100 assets) and 5 different portfolios from the Russel 3000 (again with 5, 10, 25, 50 and 100 assets). Firstly, we perform an in-sample exercise where we compare the 2-step models with the RWG and GAS-F models of Gorgi et al. (2018) and Opschoor et al. (2017). We show that the 2-step F model is the best one in terms of AIC and BIC. In addition, the choice of the matrix-F distribution is more important, in terms of performance, than the flexibility given by the 2-step estimator. Secondly, an out of sample exercise is performed to compare the 2-step estimators with the joint estimators and the HAR-DRD model of Oh and Patton (2018) which is a well-accepted benchmark in the literature. The HAR-DRD model is a two-step estimation model which does not take into account the measurement error of the realized covariance matrix. The 2-step F model outperforms all the models and the difference with the other models increase with the dimension. Finally, using an economical experiment inspired by the Markowitz (1952) framework and introduced in Fleming et al. (2001), we prove that our matrix-F model

generates economical gain over the other competing models.

Central Bank monetary policy and the Portfolio Instability

The primary goal of modern central banks is to keep the price of goods stable. For the European Central Bank, the practical definition of price stability is the year on year inflation: "Price stability is defined as a year-on-year increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below, but close to, 2% over the medium term"[1]. The main tool that the European Central Bank, as most of the other central banks, uses to keep prices stable is the *short-term interest rate*, or *deposit facility*. The deposit facility is the rate at which banks make overnight deposits with the Eurosystem. In recent years, due to a prolonged period of low inflation, several advanced economies such as Europe, Japan and the US lived in a prolonged period of low interest rates. In addition, the current state of the economic cycle and the recent inflation forecasts suggest that the low-rate period, which is now called *Low for Long*, will be prolonged even in the medium-term future.

The implications and side effects of the Low for Long for bonds (Lemke and Vladu, 2017) and banks (Altavilla et al., 2018; Brunnermeier and Koby, 2018) have been studied. The deposit facility rate is the main driver of the different proxies of *risk-free rate* used in the market, such as AAA short-term government bonds or the EURIBOR. A low level of interest rate decreases the yields of government bonds and other safe assets, which are strictly related to the interest rate, and pushes the investor to move into

[1]This is the quantitative definition of price stability of the ECB given in 2003 from the Governing Council

riskier asset classes. This mechanism is called *Risk-taking channel of monetary policy* (Borio and Zhu, 2012) and it is a desired effect of the monetary policy. The ways in which low rates can influence risk-taking are mainly two. Firstly, a reduction in the policy rate decreases the discount rate and increases stock valuations (Campbell and Shiller, 1988b). In addition, it increases the value of the collateral reducing the borrowing constraints for the firms. Secondly, low interest rates lower the expected returns for safe assets, creating the "search for yield" environment described by Rajan (2005), where economic agents move to higher risk assets to search for better performances. The risk-taking channel has been widely studied for banks, as they are the main liquidity providers. Altunbas et al. (2014) and Neuenkirch and Nöckel (2018) empirically show the effect of the risk-taking channel in the Euro Area Banks. Drechsler et al. (2014) theoretically and empirically shows the impact of low rates in the risk premium. In this thesis, we use the investors' perspective to show possible side effects of the risk-taking behaviour induced by low rates.

Chapter 5 presents the third and last contribution of the thesis. The main purpose of the chapter is to analyze possible undesired effects of the low rates environment for financial stability. Initially, we investigate this topic using a Mean-Variance framework, where the investor's portfolio is affected by the market conditions, i.e. market returns, variances and his/her risk-aversion, and the risk-free rate level. In this framework, the interest rate fixed by the central bank determines the risk-free rate level that affects the market conditions. The intuition is that a low level of risk-free rate forces the investor to move towards riskier assets. The market conditions together with the risk-free rate level define an equilibrium with high weights on the riskier assets. However, this equilibrium is mainly supported by a

small interest rate level. For this reason, a deterioration of the market conditions due to exogenous shocks can dramatically change the optimal portfolio weights and the transition can create market frictions and liquidity issues. In Chapter 5 we claim that the lower the interest rate level, the higher is the rebalancing movement created by an exogenous shock on the market conditions.

In the first part of the Chapter, we introduce a measure of *Portfolio Instability* which captures the rebalancing volumes that follow an exogenous shock in the market conditions. If we define as θ the vector representing all the relevant market conditions, for instance expected returns, expected variances and investor's risk aversion, as $\omega(\theta)$ the optimal portfolio according to the market condition vector θ and as v an exogenous shock in the vector θ, the *Portfolio Instability* measure is defined as the directional derivative of the portfolio $\omega(\theta)$ with respect to the shock v. In other words

$$\mathrm{PI}_v(\theta) = D_v\omega(\theta) = \lim_{h\to 0} \frac{\omega(\theta + hv) - \omega(\theta)}{h} \in \mathbf{R}^K \tag{1.5}$$

where K is the number of the considered assets in the portfolio. $\mathrm{PI}_v(\theta)$ is a vector where the i-th entry quantifies the rebalancing volume of asset i after the shock v. In order to quantify the combined volumes of rebalancing after the shock, we introduce the *Total Portfolio Instability* measure ($\mathrm{TPI}_v(\theta)$) as the Euclidean norm of the vector $\mathrm{PI}_v(\theta)$. This measure helps us to summarize the distance between the optimal portfolio pre-shock and the optimal portfolio post-shock.

In the second part of the chapter, we mathematically show that in a mean-variance framework under very broad assumptions, there exists an inverse relationship between the risk-free rate level and the Total Portfolio Instability measure. Given that we are interested in the relation between the risky portfolio, i.e. the market portfolio, and the risk-free asset, we

restrict our computations in a two assets environment. This restriction is supported by the mutual fund separation theorem of Merton (1972a) which shows that the optimal portfolio is composed of two assets. The first asset represents the market portfolio, while the second asset is the risk-free asset. The sensitivity of the Total Portfolio Instability to the risk-free rate is negative, where a lower level of the interest rate level generates higher portfolio instability. In addition, the relation between the two quantities is non-linear and convex, and a decrease of the risk-free rate to lower and lower level generates an explosive dynamic in the Portfolio Instability. The non-linear behaviour is a warning to highlight that, while a low level of interest rate is maybe crucial to sustaining the financial market conditions, during crises these benefits may be reversed and create financial stability problems.

In our empirical analysis, we use European market data from January 1999 to October 2020 with a weekly frequency. The equity market proxy is the EUROSTOXX600 while the 1 week EONIA rate is the proxy for the risk-free rate. We consider an ICAPM framework introduced in Merton (1973), to model the dynamics of variances and returns. In particular, we rely on the results of Campbell (1987), Glosten et al. (1993) and Scruggs (1998) to include the impact of the risk-free rate level in the variance and return dynamic. The results statistically validate the negative sensitivity of the Total Portfolio Instability to the risk-free rate.

We test the economic significance of a low level of the risk-free rate in the market sell-off during crises with a semi-natural experiment. During the Covid-19 pandemic, the market conditions deteriorated rapidly with huge outflows from investment funds. This extreme condition created several instabilities in the market liquidity and potential financial stability issues.

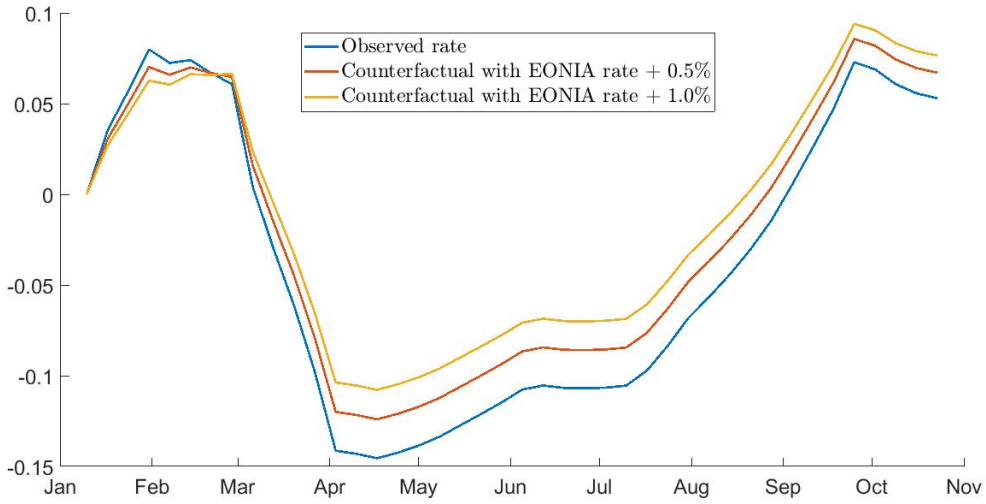

Figure 1.1: Observed sell-off (in blue) and counterfactual sell-off (in orange with 0.5% and yellow with 1%). The weights are rescaled to 0 at January 2020.

We use the CAPM framework to create a counterfactual analysis in mean-variance portfolio weights during the recent Covid-19 pandemic to test the economic impact of the risk-free rate in the market sell-off. Figure 1.1 shows the results of the counterfactual analysis. In blue we report the market sell-off, defined as the portfolio weight of the risky asset, for a mean-variance investor during the pandemic. The difference between the peak and the minimum reached 25% in April and started to recover straight after. In orange and yellow we report the counterfactual dynamic of the weights assuming the baseline EONIA +0.5% and in yellow the counterfactual weights assuming the baseline EONIA +1%. We see that the difference is economically relevant and, in the yellow line, the sell-off reduces to 10% less than the observed scenario.

Michaud (1989) and Chopra and Ziemba (1993) among others show that the mean-variance portfolio optimization suffers from extreme sensi-

tivity to the input parameters in practical use. For this reason, in the third part of the chapter, we assess whether the results are observed only in a Markowitz framework. We use the recent Risk Parity optimal portfolio (Maillard et al., 2010) with volatility targeting identification which presents a good robustness check for two reasons. From a modellistic perspective, we check whether the results obtained in the first two parts of the chapter depend on the Markowitz optimal portfolio. From a financial stability perspective, before the Covid-19 crisis around USD 2 trillion was invested in some form of volatility strategies among which USD 300 billion in risk-parity funds. The size of this market segment is macro-critical and possible flows induced by external shocks may thus be able to amplify the initial shock. In addition, we enrich the analysis using a different composition of the risky portfolio, which is now composed of equity, sovereign and corporate bonds. We use a DCC model to capture the correlation dynamics among these asset classes. The results are in line with the Mean-Variance investors and the Portfolio Instability dynamics are the same. In addition, we observe that the equity and corporate bonds asset classes are the most impacted during the Covid-19 crisis and that this impact partly depends on the low level of the interest rate. Finally, we check the robustness of our results using different market proxies and using US market data. The US universe offers a longer and well-studied time series. The results are in line with the findings of the previous parts of the chapter, where the analyses of longer time series confirm that the portfolio instability dependence on the interest rate level is not a new feature of the current macroeconomic situation but it is also present in other periods.

Chapter 2

Score-driven models and Kalman Expectation Maximization

In Chapter 3 and 4 of the thesis, two classes of econometric models are used. In this chapter, we focus on the historical and theoretical foundation of these two classes of models to assess similarities and differences.

Consider an n-dimensional latent variable of interest X which evolves through time and which is not directly observed. The variable X is linked with the observed k-dimensional variable Y via a link function g. In addition, assume that the latent variable may be expressed in a VAR form. The state space model reads

$$\begin{aligned} Y_t &\sim g(Y_t|X_t;\theta^Y), \\ X_{t+1} &= c + DX_t + \nu_t, \end{aligned} \tag{2.1}$$

where θ^Y is the set of (possibly unknown) static parameters which specifies the density function g, c is an n vector of mean values, D is the autoregressive $n \times n$ matrix and ν_t is a multivariate zero-mean noise of dimension n.

The innovation terms v_t are usually supposed to be independent.

In Chapter 3 of the thesis, the sentiment time series is modeled as a linear Gaussian state space model. In other words, the function g can be expressed as $g(Y_t|X_t;\theta^Y) = AX_t + \epsilon_t$ and both the error terms ν_t and ϵ_t are assumed to be normal, serially uncorrelated, and mutually independent. The parameters of the state space model are estimated using a combination of Kalman Filter and the expectation-maximization algorithm. The Kalman filter is used to extract the latent variable from the observed data and the expectation-maximization procedure is used to estimate the parameters.

In chapter 4, I use a recently introduced class of observation driven models know as *Generalized Autoregressive Score* (GAS) or *Dynamic Conditional Score* (DCS) model introduced in Creal et al. (2013) and Harvey (2013). In this type of models, even if the parameters are stochastic, they are perfectly predictable given the past information. Therefore the model reads

$$\begin{aligned} Y_t &\sim g(Y_t|f_t;\theta^Y), \\ f_{t+1} &= \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{i=1}^{q} B_i f_{t-i+1}, \end{aligned} \tag{2.2}$$

where ω is a vector of constants, coefficient matrices A_i and B_i relate the parameters f_{t+1} with the past values and observations. The scaled score update s_t is proportional to the score of the likelihood function g, then it depends on the measurement equation Y_t. The scaled score s_t is defined as

$$s_t = S_t \nabla_t, \qquad \nabla_t = \frac{\partial \ln g(Y_t|f_t;\theta^Y)}{\partial f_t}, \qquad S_t = S(t, f_t;\theta), \tag{2.3}$$

where $S(\cdot)$ is a matrix function.

Cox (1981) classifies the time series models with time-varying parameters into two broad classes, the *parameter-driven models* and the *observation-driven models*. The state space equation in (2.1) belongs to the class of

parameter-driven model. The state variable X_{t+1} is a stochastic process with its source of error, but it does not directly depend on past observations Y_t. In most cases, the estimation of these types of models requires Monte-Carlo simulations and it is typically unfeasible in high dimensions. One exception is the linear Gaussian state space model where the likelihood can be evaluated more easily. The score-driven models in (2.2) are in the class of *observation-driven models*, where the future's value of the parameter f_{t+1} is perfectly predictable given the present and the past, but the state variable is directly affected by past observations via the scaled score s_t.

In this chapter, I introduce the two classes of models as separated. However, recently Buccheri et al. (2018) have studied the properties of the score-driven models as approximate filters for state-space models, creating a bridge between the two approaches.

2.1 Kalman filter and Expectation Maximization

Consider a Gaussian linear version of the model (2.1), where the observed quantity $Y_t \in \mathbf{R}^k$ is a linear transformation of the unobserved variable of interest $X_t \in \mathbf{R}^n$ where $t = 1, \ldots, T$

$$\begin{aligned} Y_t &= BX_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, R) \\ X_{t+1} &= DX_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, Q), \end{aligned} \tag{2.4}$$

where B is a $k \times n$ matrix which expresses the relation between the measurement and the latent variable and D is the $n \times n$ matrix of autoregressive coefficients. In addition, we suppose that the initial value X_0 is normally

distributed with mean μ and variance Σ. Recall that both the error terms ϵ_t and ν_t are Gaussian, serially uncorrelated and independent processes. If the parameters of the model (2.4) are known, then the Kalman filter presented in Kalman (1960) provides the forecasts, filtering and smoother equations for the state variable and its variance. Denote with $\hat{X}_{k|m} = E\left[X_k|Y_1,\ldots,Y_m\right]$ and with $P_{k|m} = E\left[\left(X_k - \hat{X}_{k|m}\right)\left(X_k - \hat{X}_{k|m}\right)' |Y_1,\ldots,Y_m\right]$ the covariance matrix of $\hat{X}_{k|m}$. We define $\hat{X}_{k|k-1}$ as *prediction* of the state variable at time k, $\hat{X}_{k|k}$ as *filter* of the state variable at time k and $\hat{X}_{k|T}$ as *smoother* of the state variable at time k. Following Shumway and Stoffer (1982) the Kalman filter can be recursively evaluated. In particular, we divide the evaluation in prediction, filtering and smoothing. The prediction part is evaluated as

$$\begin{aligned} \hat{X}_{t|t-1} &= D\hat{X}_{t-1|t-1} \\ P_{t|t-1} &= DP_{t-1|t-1}D' + Q, \end{aligned} \tag{2.5}$$

where $\hat{X}_{t-1|t-1}$ and $P_{t-1|t-1}$ are evaluated in the filtering part as

$$\begin{aligned} K_t &= P_{t|t-1}B'\left(BP_{t|t-1}B' + R\right)^{-1} \\ \hat{X}_{t|t} &= \hat{X}_{t|t-1} + K_t\left(Y_t - B\hat{X}_{t|t-1}\right) \\ P_{t|t} &= P_{t|t-1} - K_tBP_{t|t-1}, \end{aligned} \tag{2.6}$$

where we take $\hat{X}_{0|0} = \mu$ and $P_{0|0} = \Sigma$. Notice that also the filtering recursion depends on the results of the previous predictions. For this reason, the two steps should be evaluated contemporaneously.

Now, using backward recursions $t = T,\ldots,1$ we derive the smoothing part

as

$$\begin{aligned} J_{t-1} &= P_{t-1|t-1} D' \left(P_{t|t-1}\right)^{-1} \\ \hat{X}_{t-1|T} &= \hat{X}_{t-1|t-1} + J_{t-1} \left(\hat{X}_{t|T} - D\hat{X}_{t-1|t-1}\right) \\ P_{t-1|T} &= P_{t-1|t-1} + J_{t-1} \left(P_{t|T} - P_{t|t-1}\right) J'_{t-1} \\ P_{t-1,t-2|T} &= P_{t-1|t-1} J'_{t-2} + J_{t-1} \left(P_{t,t-1|T} - DP_{t-1|t-1}\right) J'_{t-2} \end{aligned} \tag{2.7}$$

where $P_{T,T-1|T} = (I - K_T B)\, DP_{T-1|T-1}$.

Equations (2.5), (2.6) and (2.7) represent the three stages and refinement of the use of a Kalman filter. In the first stage, we use the prediction step (2.6) to estimate the likely future observation $\hat{X}_{t|t-1}$ of the latent variable. Once we make the observation, we update the predicted value using the new available observation Y_t. From equation (2.6), we get that the filtered value $\hat{X}_{t|t}$ depends on the *Kalman gain* K_t and on the prediction error $Y_t - B\hat{X}_{t|t-1}$. In particular, the Kalman gain is inversely proportional to the observation error matrix R. The higher is the variance of the observation equation Y, the lower is the adjustment of the Kalman filter. This result is intuitive if we think of the filtering procedure as an updating with respect to the previous forecast. The forecast is adjusted proportionally to the prediction error and the reliability of the observation Y_t. The same reasoning holds for the covariance matrices $P_{t|t-1}$ and $P_{t|t}$. The higher the value of the Kalman gain, the lower is the update of the mean square error $P_{t|t}$. In other words, the Kalman gain matrix measures the *information added* by the extra observation Y_t on the identification of the latent variable X_t.
Once the prediction and filtering procedure is over at time T, equation (2.7) shows how to update the full path of the latent variable given all history of the process. In equation (2.7) we see that the smoothing update depends, at every observation, on the updating matrix J_{t-1} which is proportional to the inverse of the *information added* by the future observation $P_{t|t+1}$ relative to

the present $P_{t|t}$. This matrix represents the equivalent of the Kalman gain matrix K_t in the updating procedure and the interpretation is analogous.

In most of the applications, the parameters B, D, R, Q, μ and Σ are unknown and need to be estimated. The maximum likelihood estimation is often unfeasible in high dimension as the number of parameters increases dramatically as the dimension increases. Dempster et al. (1977) and Shumway and Stoffer (1982) introduced a two step approach named Expectation Maximization (EM) to evaluate and maximize the likelihood using closed-form equations. Under the normality assumptions of the error terms and denoting with θ the set of parameters μ, Σ, B, D, R and Q , the log-likelihood may be written as

$$\begin{aligned} l\left(Y, X; \theta\right) =& \log f(X_0) + \sum_{t=1}^{T} \log f(X_t|Y_{t-1}) + \sum_{t=1}^{T} \log f(Y_t|X_t) \\ =& -\frac{1}{2}\log\left|\Sigma\right| - \frac{1}{2}\left(X_0 - \mu\right)\Sigma^{-1}\left(X_0 - \mu\right)' \\ & - \frac{T}{2}\log\left|Q\right| - \frac{1}{2}\sum_{t=1}^{T}\left(X_t - DX_{t-1}\right)Q^{-1}\left(X_t - DX_{t-1}\right)' \\ & - \frac{T}{2}\log\left|R\right| - \frac{1}{2}\sum_{t=1}^{T}\left(Y_t - BX_t\right)R^{-1}\left(Y_t - BX_t\right)'. \end{aligned} \tag{2.8}$$

Given that the likelihood strictly depends on the latent process X, an expectation-maximization approach is applied on the likelihood function (2.8). The maximization recursive routine is divided into two parts. Consider a general j step.

Expectation step

Consider θ_j as the vector of estimated parameters at step j, then we consider the expected log-likelihood

$$G(\theta_j) = E\left[l\left(Y, X; \theta_j\right)|Y_1, \ldots, Y_T\right]. \tag{2.9}$$

The quadratic form of the likelihood (2.8) allows us to use the Kalman smoother (2.7) in the formula and the expected likelihood function is

$$\begin{aligned} G(\theta) = &- \frac{1}{2}\log|\Sigma| - \frac{1}{2}tr\{\Sigma^{-1}\left[P_{0|T} + (X_0 - \mu)(X_0 - \mu)'\right]\} \\ &- \frac{T}{2}\log|Q| - \frac{1}{2}tr\{Q^{-1}(A_3 - A_2D' - DA_2' + DA_1D')\} \\ &- \frac{T}{2}\log|R| - \frac{1}{2}tr\{R^{-1}(E_3 - BE_2' - E_2B' + BE_1B')\}, \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} A_1 &= \sum_{t=1}^{T}\left(X_{t-1|T}X_{t-1|T}' + P_{t-1|T}\right), \\ A_2 &= \sum_{t=1}^{T}\left(X_{t|T}X_{t-1|T}' + P_{t,t-1|T}\right), \\ A_3 &= \sum_{t=1}^{T}\left(X_{t|T}X_{t|T}' + P_{t|T}\right), \\ E_1 &= \sum_{t=1}^{T}P_{t|T} + X_{t|T}X_{t|T}', \\ E_2 &= \sum_{t=1}^{T}Y_tX_{t|T}', \\ E_3 &= \sum_{t=1}^{T}Y_tY_t'. \end{aligned} \tag{2.11}$$

Maximization step

The expected likelihood (2.10) is function of the parameters only and the latent process X in equation (2.8) has been substituted by the Kalman smoother. The maximization step uses the likelihood function (2.10) to

update the vector of parameters θ_{j+1} in closed form as

$$\begin{aligned}
B(m+1) &= E_2 E_1^{-1} \\
D(m+1) &= A_2 A_1^{-1} \\
Q(m+1) &= \frac{1}{T}\left(A_3 - A_2 D\,(m+1)' - D\,(m+1)\,A_2' + D\,(m+1)\,A_1 D\,(m+1)'\right) \\
R(m+1) &= \frac{1}{T}\left(E_3 - B(m+1)E_2' - E_2 B(m+1)' + B(m+1)E_1 B(m+1)'\right) \\
a(m+1) &= X_{0|T} \\
\Sigma(m+1) &= P_{0|T}\,.
\end{aligned} \tag{2.12}$$

Dempster et al. (1977) show that the likelihood is non-decreasing following this routine. The procedure stops after a converging criterion is satisfied.

Imposing parameter restrictions

In many applications, including the model presented in Chapter 3, the matrices B, D, Q and R may be not full matrices. For instance, in many applications in physics where the observed equation is a measure of a unobserved quantity, researchers assume that the errors are cross-sectionally independent, then the matrix R should be diagonal. It is not trival how to optimally restrict the parameters estimated in the *Maximization* step. Wu et al. (1996) and Bork (2009) find a close form solution to optimally restrict the parameter matrices. Define as vec the vectorization operator and as $\otimes$ the Kronecker product. Consider a number of f restrictions k_D for the matrix D and s restrictions k_B for the matrix B. If the restrictions are linear and may be expressed in matrix form as $M\mathsf{vec}(D) = k_D$ and $G\mathsf{vec}(B) = k_B$ with restriction matrices $M \in \mathbf{R}^{f\times n^2}$ and $G \in \mathbf{R}^{s\times nk}$. In addition, denote

as B_r, D_r, R_r and Q_r as the restricted parameter matrices, we get that

$$\mathsf{vec}(D_r) = \mathsf{vec}(D) + \left(A_1^{-1} \otimes Q\right) M'(M(A_1^{-1} \otimes Q)M')^{-1}(k_D - M\mathsf{vec}(D)) \tag{2.13}$$

where A_1 is defined in (2.11).

Equivalently, for the restricted B_r:

$$\mathsf{vec}(B_r) = \mathsf{vec}(B) + \left(E_1^{-1} \otimes R\right) G'(G(E_1^{-1} \otimes R)G')^{-1}(k_B - G\mathsf{vec}(B)) \tag{2.14}$$

where E_1 is defined in (2.11).

The matrices Q_r and R_r may be restricted element-wise.

In Chapter 5 of the thesis, the Kalman Expectation-Maximization routine is exploited to extract the short-term sentiment and the long-term sentiment components from the observed sentiment.

2.2 Score-driven models

The score-driven models are a recently introduced class of observation-driven models where the update mechanism depends on the scaled score of the likelihood (Creal et al., 2013; Harvey, 2013). In these papers, the authors show that several observation-driven models are included in this general class, for instance, the ARCH and GARCH models of Engle (1982) and Bollerslev (1986), the dynamic copula model of Patton (2006) and Poisson count models of Davis et al. (2003). Starting from the general equation (2.2), Creal et al. (2013) suggest an update of f_{t+1} which depends on the observation density g for a given level of the variable f_t. Doing so, the update rule depends on the observation Y_t as defined in equation (2.3). In most of the applications, it is natural to consider the scale matrix S as the

inverse of the covariance matrix or its square root, then the matrix S reads

$$S_t = \mathcal{I}_{t|t-1}^{-1} \text{ or } S_t = \mathcal{I}_{t|t-1}^{-1/2}, \text{ with } \mathcal{I}_{t|t-1} = E_{t-1}\left[\nabla_t \nabla_t'\right]. \tag{2.15}$$

Notice that, if the model is well-specified, the scaled score s_t has a zero conditional mean by construction.

The main advantage of using this class of models is threefold. Firstly, as in the other observation driven models, the estimation problem can be solved using likelihood-based estimators. Secondly, the update rule directly comes from the density function of the observation equation. Indeed, the score-driven models define the scaled score as the update for the latent process. Thirdly, the update suggested by the scaled score is the optimal information-theoretic update (Blasques et al., 2015). Therefore the score-driven model, even if the model is severely misspecified, is conditionally Kullback–Leibler optimal.

The order of autoregressive lags q and innovation lags p are denoted as GAS(p,q).

Example: GARCH, GARCH-t and t-GAS

Score-driven models give an update that depends on the score of the observation density function. For this reason, it is interesting to compare the GARCH update rule, which does not directly depends on the observation density function, with the GAS update implied by equations (2.3) and (2.15). Consider the univariate model for demeaned asset returns $r_t = \sigma_t \epsilon_t$, where ϵ_t is random process with conditional mean equal to 0 and conditional variance equal to 1. Bollerslev (1986) assumes a Gaussian conditional distribution for the asset returns with a linear updating rule for the parameter

σ_t^2. The GARCH model reads

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0,1), \\ \sigma_{t+1}^2 &= \omega + \alpha(r_t^2 - \sigma_t^2) + \beta \sigma_t^2. \end{aligned} \tag{2.16}$$

However, as we observed in the introduction of the thesis, the conditional asset returns show a non-gaussian behaviour. For this reason, Bollerslev (1987) assumes a Student's t density for the conditional asset distributions. Then the GARCH-t model reads

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \qquad \epsilon_t \sim t_\nu(0,1), \\ \sigma_{t+1}^2 &= \omega + \alpha(r_t^2 - \sigma_t^2) + \beta \sigma_t^2, \end{aligned} \tag{2.17}$$

where t_ν is a Student's t distribution with mean 0, variance 1 and tail parameter ν. The static parameters of the two models are estimated using maximum likelihood estimators. The update equations on models (2.16) and (2.17) are identical and intuitive. The update of the latent variable σ_{t+1}^2 has a 1-lag autoregressive component and depends on a zero mean observation driven component $r_t^2 - \sigma_t^2$.

We now investigate if the update rules given in (2.16) and (2.17) coincides with the update rule implied by the score-driven dynamic. In order to check whether it is the case or not, we denote with $f_t = \sigma_t^2$ and with $r_t \sim \mathcal{N}(0, \sigma_t^2)$ for the GARCH and with $r_t \sim t_\nu(0, \sigma_t^2)$ for the GARCH-t. In the Gaussian case, the log-likelihood function is

$$\log g(r_t | \sigma_t^2) = -\frac{1}{2}\left[\log(2\pi\sigma_t^2) - \frac{r_t^2}{\sigma_t^2}\right] \tag{2.18}$$

Evaluating the score w.r.t. σ_t^2 and the Information matrix from (2.18), we obtain

$$\begin{aligned} \nabla_t &= \frac{r_t^2}{2(\sigma_t^2)^2} - \frac{1}{2\sigma_t^2} \\ \mathcal{I}_{t|t-1} &= E_{t-1}\left[\nabla_t \nabla_t'\right] = \frac{1}{2(\sigma_t^2)^2} \\ s_t &= \nabla_t \mathcal{I}_{t|t-1}^{-1} = r_t^2 - \sigma_t^2. \end{aligned} \tag{2.19}$$

The update equation implied by the score-driven dynamic in (2.19) coincides with the update rule imposed by Engle (1982) in (2.16).

The second case is the conditionally t-distributed returns with a tail parameters ν. The density function for the Student's t distribution is given by

$$\begin{aligned}\log g(r_t|\sigma_t^2,\nu) =& \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu+1}{2}\right) - \frac{1}{2}\log(\pi) - \frac{1}{2}\log(\nu-2) \\ &- \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1+\frac{r_t^2}{(\nu-2)\sigma_t^2}\right),\end{aligned} \tag{2.20}$$

If we evaluate the score w.r.t. σ_t^2 and the Information matrix from (2.20) we get

$$\begin{aligned}\nabla_t &= \frac{(\nu+1)r_t^2}{2(\nu-2)\sigma_t^4}\left(1+\frac{r_t^2}{(\nu-2)\sigma_t^2}\right)^{-1} - \frac{1}{2\sigma_t^2} \\ \mathcal{I}_{t|t-1} &= -E_{t-1}\left[\frac{\partial^2 \log g}{\partial(\sigma_t^2)^2}\right] = \frac{\nu}{2(\sigma_t^2)^2(\nu+3)} \\ s_t &= \nabla_t \mathcal{I}_{t|t-1}^{-1} = \frac{\nu+3}{\nu}\left(\left(1+\frac{r_t^2}{(\nu-2)\sigma_t^2}\right)^{-1}\frac{(\nu+1)}{(\nu-2)}r_t^2 - \sigma_t^2\right).\end{aligned} \tag{2.21}$$

Firstly, equation (2.21) shows that the update from the score-driven specification is different from the update introduced in Bollerslev (1987). Secondly, if we define as $a_t = \left(1+\frac{r_t^2}{(\nu-2)\sigma_t^2}\right)^{-1}\frac{(\nu+1)}{(\nu-2)}$ the score-driven dynamic becomes

$$\sigma_t^2 = \omega + \alpha\frac{\nu+3}{\nu}(a_t r_t^2 - \sigma_t^2) + \beta\sigma_t^2 \tag{2.22}$$

where a_t is the weight that the current observation r_t^2 has on the update of the latent variable σ_t^2. The interpretation is quite simple. The smaller the value of ν, the higher the importance of the fat-tail and the smaller is the weight of a_t if an extreme observation occurs. Figure 2.1 shows the shape of the score for different values of the parameter ν. While the scores are similar for values of r_t close to 0, the values of the scores in the tails are

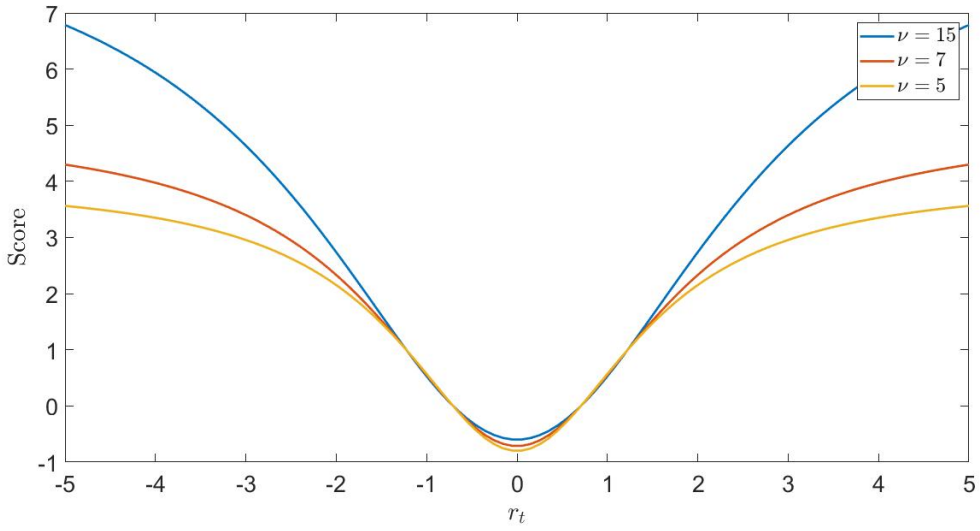

Figure 2.1: Score of GARCH-t for different values of the tail parameter ν. The definition of the score is reported in equation (2.21).

very different. For $\nu \to \infty$, the Student's t distribution collapses to the Gaussian distribution and the weight $a_t \to 1$. Creal et al. (2013) denotes this model as t-GAS.

Now, a legitimate question is whether the more complicated update rule introduced by the t-GAS model is better than the update rule of the GARCH-t, and according to what measure the two models should be compared. Blasques et al. (2015) shows that the score-driven update is optimal under a Kullback-Leibler (KL) divergence criterion. In particular, the authors show that if the data generating process has a density function $p(y_t|f_t)$ and the considered models have a misspecified density function $g(y_t|\tilde{f}_t;\theta)$, then the score-driven update reduces the KL divergence in expectation at every step. In addition, the score-driven updates are the only which can guarantee this property. In Blasques et al. (2015), the authors compare the GARCH-t and the t-GAS in equations (2.17) and (2.22) in a simulation exercise. The Data Generating Process is simulated using a fat-tail

stochastic volatility model and the two models are compared according to the KL-divergence measure and the Mean Square Error. The t-GAS performs better according to both measures.

In Chapter 4 of the thesis, two multivariate score-driven models are presented to forecast asset returns covariance matrices. The first score-driven model denoted as 2-step-W, assumes a Wishart conditional density function for the Realized covariance matrix. This distribution models the Realized covariance obtained from Gaussian distributed random vectors, which is in line with the normality assumption of conditional asset returns. The second score-driven model, denoted as 2-step-F, assumes a matrix-F conditional density function for the Realized covariance matrix. This distribution models the Realized covariance obtained from Student's t distributed random vectors, which is in line with the Student's t assumption of the conditional asset returns. However, the multivariate specification presents technical issues as the positive definiteness and the curse of dimensionality. Using the score-driven specification, we denote as X_t a measure of the covariance matrix of a K vector asset returns r_t. X_t may be the close to close covariance matrix, $X_t = r_t r_t' \in \mathbf{R}^{K\times K}$ obtained using daily prices, or it may be another realized measures of the covariance matrix obtained from intraday returns. The score-driven model reads

$$\begin{aligned} X_t|\mathcal{F}_{t-1} &\sim \mathcal{G}\left(V_t;\theta^X\right) \\ V_t &= \Omega + A' s_t A + B' V_t B. \end{aligned} \tag{2.23}$$

The model produces positive definite matrices V_t if the score matrix s_t is positive definite. However, the number of parameters of the matrices A, B and Ω increases as $\mathcal{O}(K^2)$. While the dimensionality of A and B can be controlled using parametric restrictions on the matrices, it is not easy to control the dimensionality of Ω as it captures the long-term average of

the covariance structure. The easiest way to solve this problem for high dimensions, $K \gg 1$, is to perform a two-step estimation procedure. In the first step, we ensure a reasonable value of the model-implied unconditional covariance matrix, which also helps to reduce the number of parameters in the maximization of the likelihood function, which will be performed in the second step.

However, in some cases where the matrix Ω has a specific structure, the targeting procedure cannot be applied. For instance, if the researcher is interested in modelling the correlation matrix, the targeting procedure cannot be directly applied. The Hypersphere Decomposition of Rebonato and Jäckel (2011) and the DCC decomposition of Engle (2002b) are two examples of methodologies to identify and estimate the matrix Ω. In Chapter 4 of this thesis, we use the DCC approach to target the matrix Ω and solve the problem of dimensionality in the likelihood-based estimation.

Chapter 3

Disentangling short-run and long-run components in sentiment dynamics

The contents of this chapter are the result of a joint work with Prof. Giacomo Bormetti and Prof. Fabrizio Lillo, also available as an online preprint in Vassallo et al. (2019). This paper is currently under revision in Quantitative Finance.

3.1 Introduction

Nowadays, as Ignacio Ramonet wrote in The Tyranny of Communication, "a single copy of the Sunday edition of the New York Times contains more information than an educated person in the eighteenth century would consume in a lifetime". This huge amount of information cannot be read by a single person. Recent developments in machine learning algorithms for sentiment analysis help us to categorise and extract signals from text data

and pave the way for a new area of research. The use of these new sources of textual data has become popular to analyse the relationship between sentiment and other economic variables using econometric techniques. Algaba et al. (2020) refer to this new strand of literature as *Sentometrics*. For instance, Groß-Klußman and Hautsch (2011) study the impact of unexpected news on the displayed quotes in a limit order book, Sun et al. (2016) show that intraday S&P 500 index returns are predictable using lagged half-hour investor sentiment, Antweiler and Frank (2004); Borovkova and Mahakena (2015); Allen et al. (2015); Smales (2015) study the impact of sentiment on volatility, Peterson (2016) investigates the trading strategies based on sentiment, Tetlock (2007); Garcia (2013) consider the Dow Jones Industrial Average (DJIA) index predictability using sentiment, Calomiris and Mamaysky (2019) show how the predictability can be exploited in different markets around the world, Ranco et al. (2015) analyse the impact of social media attention on market dynamics, Borovkova (2015) develops risk measures based on sentiment index, and Lillo et al. (2015) show that different types of investors react differently to news sentiment.

The approaches to sentiment analysis can be broadly classified into three categories. The first class is based on (mostly supervised) Machine Learning techniques. Three steps are typically considered. The first one is to collect textual data forming the training dataset. The second one is to select the text features for classification and to pre-process the data according to the selection. The final step is to apply a classification algorithm to the textual data. As an example, Pang et al. (2002) compare the performance of Naive Bayes, support vector machines, and maximum entropy algorithm to classify positive or negative movie reviews. The second category is the lexicon-based approach. It also typically consists of three steps. The first

step is the selection of a dictionary of N words which could be relevant for a specific topic (e.g. the word *great* is considered as a positive word to review a movie). The second one consists in tokenizing the textual data and, for each word in the dictionary, counting how many times it appears in the text. This process can be visualized with a vector of length N where the i-th element represents the number of times the i-th word of the dictionary is mentioned in the text. Finally, a measure takes the vector of length N as an input and gives a quantitative score as an output. One can refer to Loughran and McDonald (2011) for a relevant example in the financial literature. The third and last approach is a combination of methodologies coming from the first and second approaches. For an overview of textual data treatments and computational techniques, we refer to the review paper Vohra and Teraiya (2013) and the book Liu (2015).

However, as observed by Zygmunt Bauman in Consuming Life, as the amount of information increases also the amount of useless information increases, and the noise becomes predominant. Two different non-exclusive methods have been explored in the literature to remove or, at least, mitigate the impact of useless information. In the first case, a general-to-specific approach is used directly on the textual data. The amount of information can be reduced by selecting only verified news (i.e. eliminating fake news), considering only the words which are closely related to the topic of interest, considering the importance of any news (e.g. Da et al. 2011), selecting only news which appear for the first time (e.g. Thomson Reuters News Analytics engine uses the novelty variable, see Borovkova et al. 2017), or weighting news employing a measure of attention (e.g. with the number of clicks it receives when published in a news portal Ranco et al. 2016). Obviously, the selection of the relevant data is application-specific. For instance, fake news

may be irrelevant to forecast the GDP of a country but may be crucial to forecast the results of an election (e.g. Allcott and Gentzkow 2017).
In the second case, sentiment time series are directly considered, rather than the text source they are built from. The observed sentiment is noisy and various approaches have been proposed to filter it and to recover the latent signal. (Thorsrud, 2018) applies a 60-day moving average, (Peterson, 2016) uses the Moving Average Convergence-Divergence methodology proposed in (Appel, 2003) and (Borovkova and Mahakena, 2015; Audrino and Tetereva, 2019; Borovkova et al., 2017) introduce the Local News Sentiment Level model (LNSL), a univariate method which takes inspiration from the Local Level model of (Durbin and Koopman, 2012). In spite of its convenience from a practical perspective, the moving average approach is not statistically sound and the window length is usually chosen following rules of thumb, which have been tested empirically but lack a clear theoretical motivation. The methods based on the Kalman-Filter techniques present a natural and computationally simple choice to extract the informative signal. Unfortunately, when multiple assets are considered in the analysis, the LNSL model does not exploit the multivariate nature of the data. One goal of this chapter is to show that the covariance structure is very informative in sentiment time series analysis.

The first contribution of this chapter is to extend the existing time series methods in the latter stream of literature. We propose to model noisy sentiment disentangling two different sentiment signals. In our approach, the observed sentiment follows a linear Gaussian state-space model with three relevant components. The first component, named *long-term sentiment* is modeled as a random walk, the second component is termed *short-term sentiment* and follows a VAR(1) process, and the last component is an

i.i.d Gaussian *observation noise* process. We name the novel sentiment state-space model Multivariate Long Short Sentiment (MLSS). We empirically show that the decomposition provides a better insight on the nature of sentiment time series, linking the long-term sentiment to the long-term evolution of the market – proxied by the market factor – while the short-term sentiments reflect transient swing of the market mood and are more related to the market idiosyncratic components. Specifically, we find that i) the long-term sentiment cointegrates with the first market factor extracted via PCA; ii) the correlation structure of the short-term sentiment explains a significant and sizable fraction of correlation of return residuals of a CAPM model. Finally, we show that the multivariate local level model provides the best description of the data with respect to alternative models, such as the LNSL.

The second contribution of the chapter is to unravel the relation between news and market returns conditionally on quantile levels. We perform various quantile regressions showing that sentiment has good explanatory power of returns. When contemporaneous effects are considered, the result is expected and holds for all models at intermediate quantile levels. However, when the analysis is focused on abnormal days – i.e. days for which returns belong to the 1% and 99% quantiles – neither the noisy sentiment nor the filtered sentiment from an LNSL model explains the observed market returns. The only model achieving statistical significance is the MLSS. This result shows that it is essential to filter the noisy sentiment according to the MLSS, which exploits both the multivariate structure of the data and disentangles the long- and short-term components. Moreover, a test performed on the single components confirms the intuition that the short-term sentiment is the one responsible for the contemporaneous explanatory

power. The empirical evidence in favour of the MLSS becomes even more compelling when lagged relations are tested. When a single day lag is considered, i.e. one tests whether yesterday sentiment explains today returns, the significance of all models, but MLSS, drops to zero. This result holds across all quantile levels. Instead, for quantiles smaller than 10% and larger than 90%, the predictability of the returns for the MLSS model is highly significant. As before, the decomposition in two time scales is essential and the short-term component is the one responsible for the effect. The analysis extended including lagged sentiment – up to five days – confirms previous findings by (Garcia, 2013) that past sentiment contributes to predicting present returns. Interestingly, this is true for quantiles between 5% and 10%, both negative and positive, but neither in the median region nor for extreme days. In light of these findings, we finally investigated whether media and social news immediately digest market returns and whether this relation depends on the sign of returns. Our results provide a clear picture showing that i) the impact of market returns on sentiment is significant up to five days in the future when negative extreme returns – i.e. belonging to quantiles from 1% to 10% – are considered, ii) when positive returns are considered the impact rapidly fades out and is significant only for quantiles smaller than 5%, iii) previous findings become not significant if the MLSS sentiment is replaced by the observed noisy sentiment. Consistently with the intuition provided by these results, we test whether the predictability of the returns of the MLSS model can be exploited in a portfolio allocation exercise. We show that the portfolios generated with the MLSS sentiment series have a higher Sharpe ratio and lower risk than similar portfolios constructed with raw sentiment or sentiment filtered with the univariate LNSL model. Our model outperforms also the benchmark constituted by

the buy-and-hold equally weighted portfolio. This result remains true when transaction costs are included.

The rest of the chapter is organized as follows. In section 3.2, we develop the multivariate model for the sentiment and discuss the estimation technique. In section 3.3, we introduce the TRMI sentiment index and describe the data used in the analysis. In section 3.4, we report the empirical findings and discuss the advantages of the multivariate approach. In section 3.5, we compare the various techniques and report the performances of the long-short sentiment decomposition in explaining daily returns. Section 3.6 describes the portfolio allocation strategies using different filtering techniques and assesses the superiority of the MLSS filter among the others. Section 3.7 draws the relevant conclusions and sketch possible future research directions.

3.2 The Model

Consider K assets and the corresponding K observed daily sentiment series S_t^i where $i = 1, \ldots, K$. The observed daily sentiment S_t^i quantifies the opinions of investors and consumers about company i. In most cases, the observed sentiment is a continuous number in a compact set.

The Local News Sentiment Level model (LNSL), presented in (Borovkova and Mahakena, 2015) and subsequently used in (Audrino and Tetereva, 2019), reads as follows

$$\begin{aligned} S_t^i &= F_t^i + \epsilon_t, \qquad \epsilon_t \overset{d}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^i\right), \\ F_t^i &= F_{t-1}^i + v_t, \quad v_t \overset{d}{\sim} \mathcal{N}\left(0, \sigma_v^i\right). \end{aligned} \tag{3.1}$$

for every $i = 1, \ldots, K$. This model is a univariate specification of the Local Level model of (Durbin and Koopman, 2012). The latent sentiment series

F_t^i are considered as slowly changing components, modeled as independent random walks and the parameters σ_ϵ^i and σ_v^i are estimated via maximum likelihood (MLE).

Since the LNSL model does not consider the correlations of the innovations among the K assets, we can easily derive its multivariate version as

$$\begin{aligned} S_t &= F_t + \epsilon_t, \qquad \epsilon_t \overset{d}{\sim} \mathcal{N}(0, R), \\ F_t &= F_{t-1} + v_t, \quad v_t \overset{d}{\sim} \mathcal{N}(0, Q). \end{aligned} \tag{3.2}$$

where $S_t = \left[S_t^1, \ldots, S_t^K\right]'$ and $F_t = \left[F_t^1, \ldots, F_t^K\right]'$ are K dimensional vectors, Q is a $K \times K$ symmetric matrix and R is a $K \times K$ diagonal matrix. We refer to the multidimensional LNSL model as MLNSL. The synchronous correlation among the innovations of the latent sentiment is described by the covariance matrix Q, while the correlations among the observation noises are assumed to be 0. Clearly, the LNSL model is a special case of the MLNSL model when the matrix Q is diagonal. Since the number of parameters for this model scales as K^2, the MLE of the MLNSL model is computationally demanding. For this reason, we use the Kalman-EM approach described in (Corsi et al., 2015).

The idea of the LNSL and MLNSL models is that the latent sentiment is a slowly changing component with a Gaussian disturbance. In their empirical studies, (Audrino and Tetereva, 2019) observe that the signal to noise ratio $\frac{\sigma_v^2}{\sigma_\epsilon^2}$, obtained using the LNSL filter, is very small. This finding indicates that the majority of the daily changes in the sentiment series can be considered as noise. One possible explanation of this result is that the Local Level specification of these models is not sufficiently rich to capture all the signals from the observed sentiment. Indeed, in newspapers and social media, there is a consistent amount of articles and opinions which represent fast trends or rapidly changing consumer preferences. Following

the recent strand of literature on persuasion (Gerber et al., 2011; Hill et al., 2013), these fast trends have strong but short-lived effects on consumer preferences. Since the (M)LNSL model interprets the latent sentiment as an integrated series, these signals are considered as noise.

The main contribution of this chapter is to define a new model which disentangles the slowly changing sentiment from a rapidly changing sentiment, that we name short-term sentiment, and the observation noise. In addition, it is reasonable to think that the slowly changing components of a set of firms with common characteristics, for instance, belonging to the same sector, market, or country, should be affected by the same trends and shocks. For this reason, in our model we consider a number $q \leqslant K$ of common factors driving the slow component of the sentiment dynamics. We name these common factors as long-term sentiment. We do not fix the number q *a priori*, but we select it by means of an information criterion.

To provide a more quantitative intuition behind our modeling specification, let us consider the true, but unobserved, daily investor's mood M_t^i of asset i. We hypothesize today's daily mood can be written as

$$\text{Mood}_t^i = \text{Long-term Mood}_t^i + \text{Short-term Mood}_t^i. \tag{3.3}$$

The Long-term Mood is composed by yesterday's Long-term Mood plus a shock $s_t^{i,\,\text{long}}$, which is usually small but permanent, i.e.

$$\text{Long-term Mood}_t^i = \text{Long-term Mood}_{t-1}^i + s_t^{i,\,\text{long}}.$$

On the contrary, the Short-term Mood is short-lived, but with a strong and highly influential impact. In particular, the Short-term Mood is composed by a residual part of the yesterday Short-term Mood plus a shock $s^{i,\,\text{short}}$, i.e.

$$\text{Short-term Mood}_t^i = \phi^i \text{Short-term Mood}_{t-1}^i + s_t^{i,\,\text{short}}.$$

In this framework, the long-term shocks permanently change the investor's mood while the short-term shocks has an exponentially decaying persistence in the investor's mood. Equation (3.3) can be rewritten as

$$\text{Mood}_t^i = \text{Long-term Mood}_{t-1}^i + s_t^{i,\,\text{long}} + \phi^i \text{Short-term Mood}_{t-1}^i + s_t^{i,\,\text{short}}. \tag{3.4}$$

Considering the whole story and the dynamic of the two sentiments shocks, we can rewrite equation (3.4) as

$$\text{Mood}_t^i = \underbrace{\sum_{k=-\infty}^{t} (\phi_i)^{t-k} s_k^{i,\,\text{short}}}_{\text{Short-term Mood}_t^i} + \underbrace{\sum_{k=-\infty}^{t+1} s_k^{i,\,\text{long}}}_{\text{Long-term Mood}_t^i},$$

where we assumed $\text{Mood}_{-\infty}^i$ to be negligible and equal to zero. In full generality, the multivariate version of model (3.3) can be formulated as follows

$$\text{Mood}_t = A \times \text{Long-term Mood}_t + B \times \text{Short-term Mood}_t\,,$$

with A and B being $K \times K$ matrices. However, in light of the considerations in the previous paragraph, we restrict the matrix B to be the identity matrix. In this way, the Short-term Mood is purely company-specific. We replace $A\,\text{Long-term Mood}_t$ with the product between a factor loading matrix and a limited number of long-term and common factors, that is we rewrite the previous equation as

$$\text{Mood}_t = \Lambda\,\text{Long-term Factor Mood}_t + \text{Short-term Mood}_t, \tag{3.5}$$

where Λ belongs to $\mathbb{R}^{K\times q}$ with $q \leq K$. It is important to notice that the significance of Λ can be statistically tested and the selection of the number q of common factors can be performed by means of AIC and BIC criteria. Following Audrino and Tetereva (2019), we assume that the observed

sentiment S_t is a noisy observation of the investors Mood_t, and we formulate a state-space model for S_t consistent with the intuition provided by model (3.5). The Multivariate Long Short Sentiment model (MLSS) for the observed sentiment model, assuming a Gaussian specification for the short-term sentiment shock, long-term sentiment shock and the observation noise, reads

$$\begin{aligned} S_t &= \Lambda F_t + \Psi_t + \epsilon_t, \quad & \epsilon_t &\overset{d}{\sim} \mathcal{N}(0, R), \\ \Psi_t &= \Phi \Psi_{t-1} + u_t, & u_t &\overset{d}{\sim} \mathcal{N}(0, Q_{short}), \\ F_t &= F_{t-1} + v_t, & v_t &\overset{d}{\sim} \mathcal{N}(0, Q_{long}), \end{aligned} \tag{3.6}$$

where $R \in \mathbb{R}^{K \times K}$ is the diagonal covariance matrix of the observation noise ϵ_t, $\Phi \in \mathbb{R}^{K \times K}$ is the matrix of autoregressive coefficients, $Q_{short} \in \mathbb{R}^{K \times K}$ is the covariance matrix of the short-term sentiment innovations, and $Q_{long} \in \mathbb{R}^{q \times q}$ is the covariance matrix of the random walk innovations. In equation (3.6), F_t and Ψ_t are the latent processes which proxy the Long-term Factor Mood and Short-term Mood in (3.5), respectively. Please notice that the essential difference between equation (3.5) and equation (3.6) is that the observed sentiment, and its components, are noisy versions of the investors' mood and its long and short components. Finally, in this chapter, we force a diagonal structure on the matrix Φ, thus neglecting the possible lead-lag effects among sentiments. This restriction is introduced to limit the curse of dimensionality of the model.

It is worth noticing that, if the observed sentiment S_t lies in a compact set, the LNSL, MLNSL and MLSS models, in their current specification, do not consider the upper and lower bounds. This issue can be accounted for with a non-linear transformation of the data, e.g. by Fisher transforming the data, which maps them on the whole real line as commonly done for correlation time series. In this chapter, we do not apply any non-linear transformation since most of the daily sentiment observations are far from

the bounds and, as we verified, the Fisher transform mildly affects our analysis. In addition, the definition of unit root processes in presence of bounds is non standard but studied in the literature and it can be tested using ad-hoc unit root tests (Cavaliere and Xu, 2014).

The estimation of the unknown parameters is based on a combination of the Kalman filter with Expectation Maximization (Kalman, 1960; Shumway and Stoffer, 1982; Wu et al., 1996; Harvey, 1990; Banbura and Modugno, 2014; Jungbacker and Koopman, 2008). Given that model (3.2) is a special case of model (3.6), in the next session we only consider the estimation procedure of model (3.6).

Estimation procedure

The estimation of model (3.6) is performed using the Kalman filter (Kalman, 1960) and the Expectation Maximization (EM) method in (Dempster et al., 1977) and (Shumway and Stoffer, 1982) which was proposed to deal with incomplete or latent data and intractable likelihood. The EM algorithm is a two-step estimator. In the first step, we write the expectation of the likelihood considering the latent process as observed. In the second step, we re-estimate the static parameters maximizing the expectation obtained in the first step. This routine is repeated until some convergence criterion is satisfied. To cast (3.6) in a standard state-space representation, we use the same procedure of (Banbura and Modugno, 2014) and define the augmented states $\tilde{\Lambda}$, $\tilde{F}$, $\tilde{\Phi}$ and $\tilde{Q}$ s.t.

$$\begin{aligned} S_t &= \tilde{\Lambda}\tilde{F}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, R), \\ \tilde{F}_t &= \tilde{\Phi}\tilde{F}_{t-1} + v_t, \quad v_t \sim \mathcal{N}\left(0, \tilde{Q}\right), \end{aligned} \tag{3.7}$$

where

$$\tilde{\Lambda} = \begin{bmatrix} \Lambda & I_K \end{bmatrix} \in \mathbb{R}^{K \times (q+K)}, \quad \tilde{F}_t = \begin{bmatrix} F_t \\ \Psi_t \end{bmatrix} \in \mathbb{R}^{(q+K)\times 1}, \quad \tilde{\Phi} = \begin{bmatrix} I_q & 0 \\ 0 & \Phi \end{bmatrix} \in \mathbb{R}^{(q+K)\times(q+K)},$$

$$\tilde{Q} = \begin{bmatrix} Q_{long} & 0 \\ 0 & Q_{short} \end{bmatrix} \in \mathbb{R}^{(q+K)\times(q+K)}. \tag{3.8}$$

The EM renders the approach feasible in high dimensions. Indeed, while a direct numerical maximization of the likelihood is computationally demanding, the EM algorithm, thanks to the Kalman filtering and smoothing recursions, can be formulated using the closed-form equations reported in Section 2.1. In particular, it allows disentangling the long-term sentiment F_t and the short-term sentiment Ψ_t. To derive the EM steps we consider the log-likelihood $l\left(S_t, \tilde{F}_t, \theta\right)$ where θ denotes the set of static parameters $\tilde{\Lambda}$, $\tilde{\Phi}$, $\tilde{Q}$ and R. The EM proceeds in a sequence of steps:

1. E-step: it evaluates the expectation of the log-likelihood using the estimated parameters from the previous iteration $\theta(j)$:

   $$G\left(\tilde{\Lambda}(j), \tilde{\Phi}(j), \tilde{Q}(j), R(j)\right) = E\left[l\left(S_t, \tilde{F}_t, \theta(j)\right) | S_1, \ldots, S_T\right].$$

   The details of the E-step are explained in Chapter 2 Section 2.1.

2. M-step: the parameters are estimated again maximizing the expected log-likelihood with respect to θ:

   $$\theta(j+1) = \arg\max_{\theta} G\left(\tilde{\Lambda}(j), \tilde{\Phi}(j), \tilde{Q}(j), R(j)\right).$$

   The details of the M-step are explained in Chapter 2 Section 2.1.

We initialize the parameters $\theta(0)$ and repeat steps 1 and 2 until we reach the convergence criterion

$$\frac{|l\left(S_t, \tilde{F}_t, \theta(j)\right) - l\left(S_t, \tilde{F}_t, \theta(j-1)\right)|}{|l\left(S_t, \tilde{F}_t, \theta(j)\right) + l\left(S_t, \tilde{F}_t, \theta(j-1)\right)|} < \frac{\epsilon}{2}.$$

We set $\epsilon = 10^{-3}$. As observed in (Harvey, 1990), the dynamic factor model (3.7) is not identifiable. Indeed, if we consider a non singular invertible matrix M, then the parameters $\theta_1 = \{\Lambda, R, Q\}$ and $\theta_2 = \{\Lambda M^{-1}, R, MQM'\}$ are observationally equivalent, then starting from S_t we cannot distinguish θ_1 from θ_2. We solve this identification problem using the approach proposed by (Harvey, 1990), imposing the following restrictions

$$\tilde{Q} = \begin{bmatrix} I_q & 0 \\ 0 & Q_{short} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_{11} & 0 & 0 & \dots & 0 \\ \lambda_{21} & \lambda_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{K1} & \lambda_{K2} & \lambda_{K3} & \dots & \lambda_{Kq} \end{bmatrix} \tag{3.9}$$

where Λ is the $K \times q$ sub-matrix in (3.8).

The specifications of $\tilde{\Lambda}$, $\tilde{\Phi}$, $\tilde{Q}$ and R in (3.8), together with (3.9), impose several constraints to the estimation. The EM procedure allows us to impose restrictions on the parameters in a closed-form. The optimal restrictions are studied in (Wu et al., 1996) and (Bork, 2009). The details are reported in Section 2.1.

The final estimation scheme reads as follows:

1. Initialize $\tilde{\Lambda}(0)$, $\tilde{\Phi}(0)$, $\tilde{Q}(0)$ and $R(0)$
2. Perform the E-step using the estimations $\tilde{\Lambda}(j)$, $\tilde{\Phi}(j)$, $\tilde{Q}(j)$ and $R(j)$ and the Kalman Smoother (2.7).
3. Perform the M-step and evaluate the new estimators $\tilde{\Lambda}(j+1)$, $\tilde{\Phi}(j+1)$, $\tilde{Q}(j+1)$ and $R(j+1)$ using equations (2.12).
4. Use the unrestricted estimations and (2.13) and (2.14) to obtain the restricted ones.

5. Repeat 2, 3 and 4 above until the estimates and the log-likelihood reach convergence.

Finally, we select the number of long-term sentiment q using the AIC and BIC indicators.

3.3 Data

The TRMI sentiment index is constructed using over 700 primary sources, divided into news and social media, and collects more than two million articles per day. For every article, a "bag-of-words" technique is used to create a sentiment score, which lies between -1 and $+1$, a buzz variable[1], and one or more asset codes, which in our case refer to companies. The time resolution of the sentiment data is one minute.

For any asset a, minute s, and day t we denote as $S^a_{t,s}$ the sentiment score and as $\text{Buzz}^a_{t,s}$ the buzz variable. Since the following empirical analysis are performed using daily data, we need to aggregate the TRMI series on a daily basis. TRMI user guide suggests to use the following equation

$$S^a_t = \frac{\sum_{s=4:00\ \text{PM}^{t-1}}^{4:00\ \text{PM}^t} Buzz^a_{t,s} S^a_{t,s}}{\sum_{s=4:00\ \text{PM}^{t-1}}^{4:00\ \text{PM}^t} Buzz^a_{t,s}} \in [-1, 1]\,, \tag{3.10}$$

where S^a_t refers to the daily sentiment at day t, evaluated on a 24-hour window between the 4:00 PM of day $t-1$ (4:00 PM^{t-1}) and the 4:00 PM of day t (4:00 PM^t). Note that the TRMI server provides a daily frequency sentiment, where they use equation (3.10) with a 24-hours window from 3:30 PM

[1] "The buzz field represents a sum of entity-specific words and phrases used in TRMI computations. It can be non-integer when any of the words/phrases are described with a minimizer, which reduces the intensity of the primary word or phrase. For example, in the phrase "less concerned" the score of the word "concerned is" mitigated by "less". Additionally, common words such as "new" may have a minor but significant contribution to the Innovation TRMI. As a result, the scores of common words/phrases with minor TRMI contributions can be minimized." See TRMI user guide.

to 3:30 PM of the day after. However, since we want to relate the sentiment series with close to close returns, we construct the daily sentiment series aggregating the high-frequency sentiment according to the trading closing hour of the NYSE (4:00 PM). For more details, please refer to (Peterson, 2016).

For the empirical analysis, we consider the TRMI sentiment index of 27 out of 30 stocks[2] of the Dow Jones Industrial Average (DJIA) over the period 03/01/2006 – 29/12/2017. Since the TRMI index divides the news sentiment from the social sentiment, we have a total of 54 time series. A description of tickers and sectors is available in Appendix A.1. Finally, the MLSS model, in its current specification, does not manage missing values in data, while some of the sentiment time series present missing observations. The EM algorithm is naturally designed to handle missing observations. However, since the number of missing values is small[3], we fill them using the rolling mean over the last 5 days.

3.4 Empirical analysis

In this section, we present the results of the estimation of the MLSS model for the investigated stocks, providing an economic interpretation for the long- and short-term component of the sentiment. In the analyses, we consider separately the case of news and social sentiment indicators.

The first quantity to fix is the number q of long-term sentiment factors. Using the Bayesian information criteria (BIC) we select $q_{\text{news}} = 2$ and $q_{\text{social}} = 2$.

[2]We only consider 27 assets because one is missing in the Thomson Reuters dataset and two have a high ratio of missing values at the beginning of the sample.

[3]47 out of 54 sentiment series have less than 1% of missing observations. All the series have a percentage of missing which is smaller than 7.5%

Table 3.1 reports the values of Φ and Λ with the estimation errors[4]. Bold values indicate parameters that are significantly different from 0 with a p-value smaller than 0.05. We notice that most of the estimated parameters are statistically significant.

As an illustrative example, Figure 3.2 shows how the filter works for the Goldman Sachs news sentiment series. We observe that a high fraction of the sentiment daily variation is captured by the filter. In Appendix A.2 we quantify more in detail the signal-to-noise ratio of the proposed filter. We find that the MLSS model has a signal-to-noise ratio approximately twenty times larger than the MLNSL. Moreover, the noise in social media is generally higher than the noise in newspapers.

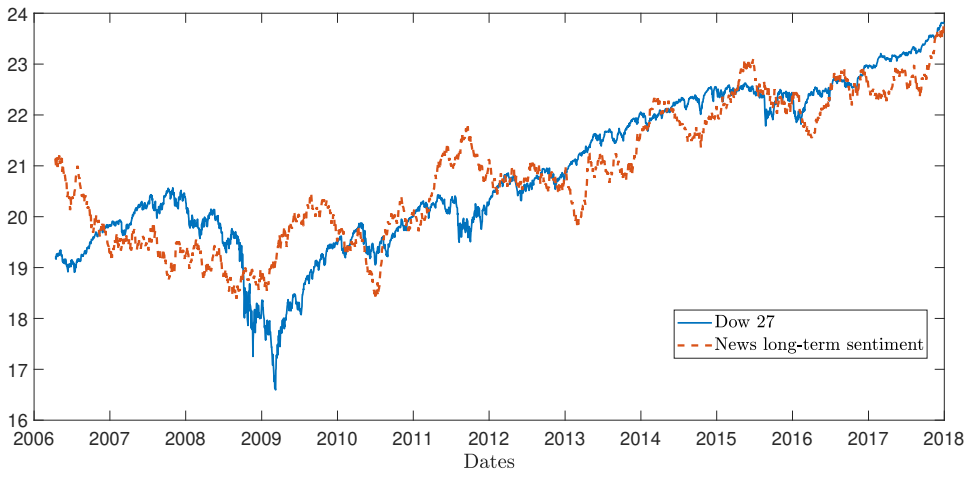

Figure 3.1: Co-integration between Dow 27, in blue, and the second factor of the news long-term sentiment, in orange. Time series are scaled. The Dow 27 is the first market factor built using log-prices.

The MLSS approach considers two new quantities extracted from the observed sentiment. The first novelty is the long-term sentiment which, by construction, represents the series of common trends in a particular basket

[4]Note that the Λ matrices, as discussed in the supplementary material, have the upper triangular submatrix equal to zero.

Tickers	$\Phi^{\mathbf{news}}$	$\Lambda^{\mathbf{news}}$		Signal to noise	
				MLSS	MLNSL
AXP	**0.464**	**1.177**		0.623	0.010
	(0.029)	(0.050)			
JPM	**0.732**	**−0.169**	**0.711**	0.326	0.023
	(0.016)	(0.035)	(0.058)		
VZ	**0.682**	**0.545**	−0.080	0.431	0.029
	(0.019)	(0.038)	(0.063)		
CVX	**0.545**	**0.103**	**0.894**	0.610	0.022
	(0.024)	(0.042)	(0.071)		
GS	**0.773**	**−0.239**	**0.718**	0.336	0.029
	(0.014)	(0.036)	(0.060)		
JNJ	**0.407**	**0.851**	**0.834**	0.788	0.010
	(0.030)	(0.039)	(0.065)		
MRK	**0.336**	**0.811**	**0.885**	0.832	0.008
	(0.033)	(0.036)	(0.059)		
PFE	**0.299**	**0.530**	**1.021**	1.185	0.007
	(0.029)	(0.031)	(0.052)		
UNH	**0.374**	**1.177**	**0.530**	0.574	0.009
	(0.037)	(0.056)	(0.093)		
BA	**0.585**	**0.376**	**0.742**	0.896	0.033
	(0.021)	(0.036)	(0.059)		
CAT	**0.633**	**0.309**	0.045	0.423	0.017
	(0.021)	(0.064)	(0.108)		
GE	**0.581**	**1.083**	**−0.196**	0.587	0.022
	(0.023)	(0.035)	(0.058)		
MMM	**0.295**	**0.958**	0.072	0.788	0.009
	(0.034)	(0.038)	(0.064)		
UTX	**0.331**	**0.422**	**−0.413**	0.690	0.011
	(0.035)	(0.057)	(0.094)		
XOM	**0.591**	−0.058	**1.025**	0.725	0.031
	(0.021)	(0.039)	(0.065)		
KO	**0.486**	**0.476**	**0.245**	0.620	0.015
	(0.028)	(0.033)	(0.055)		
PG	**0.337**	**0.838**	**−0.623**	0.929	0.008
	(0.031)	(0.041)	(0.068)		
AAPL	**0.593**	**0.221**	**0.160**	1.736	0.096
	(0.018)	(0.026)	(0.043)		
CSCO	**0.714**	**1.063**	**−1.094**	0.441	0.046
	(0.017)	(0.043)	(0.071)		
IBM	**0.603**	**0.754**	**−1.269**	0.853	0.040
	(0.020)	(0.038)	(0.063)		
INTC	**0.641**	**0.641**	**−0.299**	0.865	0.066
	(0.018)	(0.039)	(0.065)		
MSFT	**0.651**	**0.858**	−0.007	0.668	0.053
	(0.019)	(0.026)	(0.043)		
DIS	**0.439**	**0.454**	**−0.198**	1.074	0.013
	(0.025)	(0.028)	(0.046)		
HD	**0.611**	**1.137**	**0.232**	0.473	0.021
	(0.024)	(0.058)	(0.098)		
MCD	**0.404**	**−0.291**	0.020	1.401	0.013
	(0.024)	(0.034)	(0.057)		
NKE	**0.368**	**0.664**	**−0.285**	0.783	0.010
	(0.032)	(0.046)	(0.076)		
WMT	**0.516**	**0.147**	**0.619**	0.854	0.022
	(0.023)	(0.031)	(0.052)		

Table 3.1: Static parameters of model (3.6) for news sentiment. Values and standard errors of estimated Λ are multiplied by 10^3. In parenthesis we show the standard error of the estimated parameter. The last two columns show the signal to noise ratio for two competing models.

of sentiment time series. The second novelty is the multivariate structure of sentiment, extracted using the symmetric matrix Q_{short}. In the next sections, we separately analyse the relationship between these two quantities and the stock market prices. To this end, we extract the market factors from the stock prices of these assets. Denote as $r_t \in \mathbb{R}^{27}$ the vector of demeaned close-to-close log-returns and evaluate the unconditional covariance matrix Q_{ret} and the unconditional correlation matrix C_{ret}. We extract the factor loading matrix $\Lambda^{\text{mrk}} \in \mathbb{R}^{q_{\text{mrk}} \times 27}$ using the PCA on the matrix C_{ret} and define the return factors $R_t = \Lambda^{\text{mrk}} r_t \in \mathbb{R}^{q_{\text{mrk}}}$. We also define the market factors as $M_t^{\text{mrk}} = \Lambda^{\text{mrk}} p_t$, where $p_t \in \mathbb{R}^{27}$ is the vector of log-prices. In the following analysis, we consider $q_{\text{mrk}} = 1$ and name the first market factor Dow 27.

Long-term Sentiment

We first investigate the economic meaning of the long-term sentiment. Using the Engle-Granger test (Engle and Granger, 1987), we observe that one of the factors of the long-term sentiment is cointegrated with the Dow 27. Figure 3.1 shows the cointegration relation, pointing out that the main driver of the prices and the driver of the sentiment time series reflect the same common information. This result per se is not surprising. However, Figure 3.3 shows the standardized weights of the cointegrated factors. The weights of the market factor are very homogeneous across assets, as shown in the top panel, while the weights of the cointegrated factor of the long-term sentiment are very heterogeneous, as shown in the bottom panel. The values of the elements of the factor loading matrix Λ^{news} reported in Table 3.1 are either positive or negative [5]. Then, some firms' sentiment positively affects the common sentiment factors, while some other firms' sentiment

[5]The elements of the factor loading matrix Λ^{social} are available upon request.

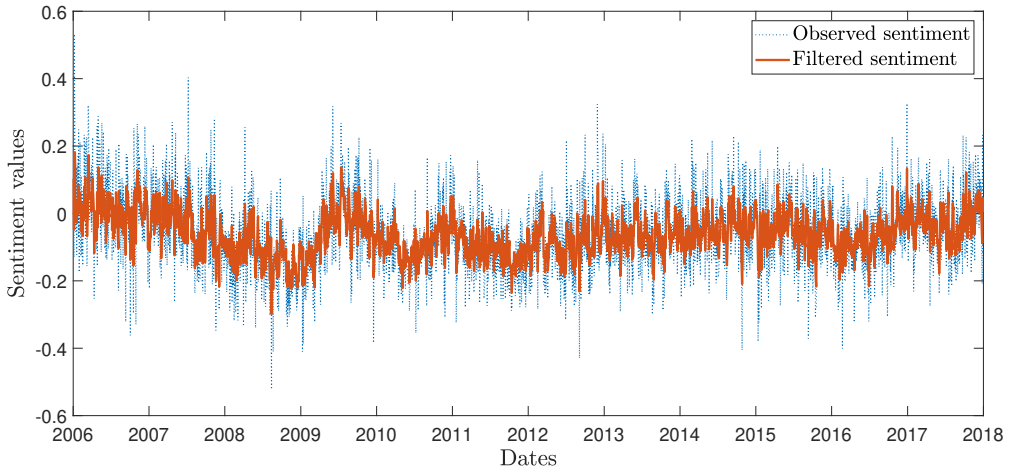

Figure 3.2: Goldman Sachs sentiment series. In blue the observed sentiment, in orange the filtered sentiment including both long-term and short-term component.

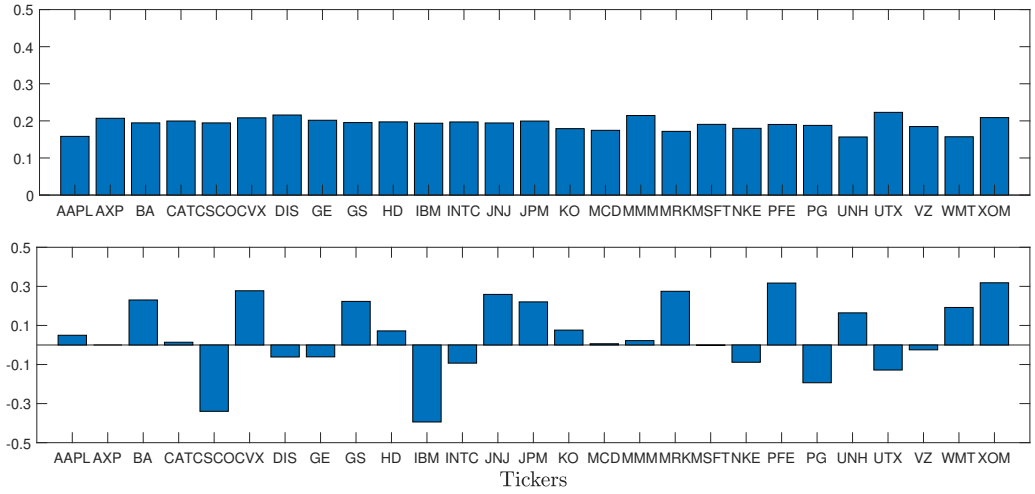

Figure 3.3: Values of the standardized factor loadings of the cointegrated series. Top panel: loadings of the Dow 27 index. Bottom panel: loadings of the second factor of the news long-term sentiment.

negatively affects them. We checked whether the heterogeneity of weights was related to the number of news of a given asset, or with the buzz index, but we found no significant evidence. Unravelling the origin of the detected heterogeneity is an interesting research question, that could be probably answered by looking at the contents of the articles from which the sentiment was computed. Unfortunately, we do not have access to this kind of information.

Short-term Sentiment

The second novelty of the MLSS model is the multivariate structure of the short-term sentiment series. The question we want to address in this section is whether the correlation structure of the short-term sentiment is (linearly) related to the correlation structure of the daily returns. In the previous section, we observed that one of the factors of the long-term sentiment is cointegrated with the first market factor. We, therefore, expect the short-term sentiment to capture asset-specific features, i.e. we expect a close relationship with the idiosyncratic dynamic of the returns[6]. To test this intuition for the correlation structure, we compare the results of the MLSS model with the results of the MLNSL model which, by construction, does not disentangle the factors from the sentiment series. If the intuition is correct, the correlation matrix of the sentiment extracted using the MLSS model should be linearly related with the return correlations and with the idiosyncratic return correlations. On the contrary, the correlation matrix of the sentiment extracted using the MLNSL model, which only captures the slowly changing dynamics of the sentiment series, and thus of the first

[6]We define idiosyncratic returns as the market returns where the first market factor is removed using the factor model (3.12)

market factor, should be linearly related with the returns correlation but mildly correlated with the idiosyncratic returns correlations. Finally, to test whether the filtering procedure is a crucial step in our approach, the correlation matrix of the observed sentiment is also considered.

We define C_{short} as the correlation matrix associated with the covariance matrix Q_{short}, C_{MLNSL} the correlation matrix associated with the covariance matrix Q of equation (3.2), $C_{\text{Obs}} = \text{Corr}\,(\Delta S_t)$ the unconditional correlation of the first difference of the observed sentiment, and C_{ret} the unconditional correlations matrix of the stock returns. We search for a linear element-wise relation between C_{ret} and C_{model}, where model is one of short, MLNSL, or Obs. The results are reported for the news case only, but the conclusions are similar for the social sentiment.

We perform a standard ordinary least squares estimation on the model

$$\mathsf{vechl}(C_{\text{ret}}) = \alpha + \beta^{\text{model}}\mathsf{vechl}(C_{\text{model}}), \tag{3.11}$$

where $\mathsf{vechl}(X)$ is the operator which collects the lower triangular elements of matrix X in a column vector. We compare the results obtained using the MLSS model ($C_{\text{model}} = C_{\text{short}}$), with the results obtained using the MLNSL model ($C_{\text{model}} = C_{\text{MLNSL}}$) and using the Observed sentiment ($C_{\text{model}} = C_{\text{Obs}}$). In addition, since the unconditional correlation between two assets is higher when they belong to the same sector, we separately consider two cases. In the first case, we estimate model (3.11) considering all the pairs of assets. In the second case, we estimate model (3.11) considering only the pairs of assets belonging to the same economic sector according to Table A.1.

The top left panel of Table 3.2 shows the results with all the correlation pairs. In the first column, we report the R^2 of the regression, in the second column we report the F-statistic and the relative p-value is reported in the

third column. The regressions with C_{short} and C_{MLNSL} have high and significant p-values, while the regression with C_{obs} is not statistically different from the model with the intercept only. This finding has two implications. The first one is that the sentiment innovations have a similar correlation structure as that of the returns innovations. In particular, if the returns of two assets are relatively highly correlated, then also the increment of the filtered sentiment of the news about these assets are relatively highly correlated. The second implication is that, if a filtering procedure is not applied to the observed sentiment data, the noise is too large to find significant results. In the top right panel of Table 3.2 we report the results of the model (3.11) applied to the pairs of assets belonging to the same sector. We observe that the R^2 increases for all models. This result is expected since it is well known that the return correlation is higher and more significant between two assets of the same sector. However, even if the R^2 increases, the number of pairs decreases. For this reason, the increment in the R^2 does not lead to an increment in the F-statistic, which fails to reject the null hypothesis for the C_{obs}. This result confirms that the C_{obs} matrix is not a significant regressor for C_{ret}.

Comparing the top panels of Table 3.2, we note that the increment in the R^2 is higher for the MLSS model rather than the MLNSL model. This evidence is consistent with the intuition that the short-term sentiment series, extracted using the MLSS model, are more related with the idiosyncratic returns. Indeed the correlation induced by the market factor is predominant in the first case, reported in the top left panel, where all the assets are considered, rather than the second case, reported in the top right panel, where the co-movements are not only driven by the first market factor, but they are also driven by sector-specific factors.

Models	All assets			Same sector		
	R^2	F-statistic	p-value	R^2	F-statistic	p-value
MLSS	13.77 %	55.713	0.0000	37.89 %	23.182	0.0000
MLNSL	15.63 %	64.669	0.0000	28.78 %	15.359	0.0004
Obs	0.95 %	3.330	0.0689	4.19 %	1.662	0.2052
MLSS	11.34 %	44.659	0.0000	30.91 %	17.001	0.0002
MLNSL	4.31 %	15.700	0.0001	7.50 %	3.081	0.0873
Obs	1.01 %	3.554	0.0602	4.88 %	1.950	0.1707

Table 3.2: Top rows: Results from the linear regression (3.11). Bottom rows: Results from the linear regression (3.13). Left columns: OLS estimates when all the assets are considered; right columns: OLS estimates when only the correlations between stocks belonging to the same sector are considered. Obs rows: estimation based on the observed sentiment.

Now we extract the Dow 27 return from the asset returns using a one-factor model. We repeat the analysis comparing the matrices C_{short}, C_{MLNSL} and C_{Obs} with the unconditional correlation of the idiosyncratic returns. We extract the market factor R_t from the returns using the factor model

$$r_t^i = \alpha^i + \beta^i R_t + z_t^i, \quad \forall i = 1, \dots, 27 \tag{3.12}$$

where $z_t^i \sim N(0, \tilde{Q}_{\text{ret}})$. We then compute the cross-correlation matrix $\tilde{C}_{\text{ret}}$ from the covariance matrix $\tilde{Q}_{\text{ret}}$ and estimate the following model

$$\mathsf{vechl}(\tilde{C}_{\text{ret}}) = \alpha + \beta^{\text{model}} \mathsf{vechl}(C_{\text{model}}). \tag{3.13}$$

The bottom panels of Table 3.2 report the results. In the bottom left panel, we show the results for the model (3.13) where all the correlation pairs are considered. The first evidence is that the MLNSL R^2 dramatically decreases, while the MLSS R^2 remains almost the same. This finding suggests that almost all the return correlations explained by the C_{MLNSL} matrix are associated with the market factor R_t, while the matrix C_{short}, which represents the fast trends on the sentiment data, also captures different dynamics.

In the bottom right panel, we show the results for the model (3.13) where

we consider only the correlation pairs for assets belonging to the same sector. In this case, the differences between the MLSS and MLNSL are more severe. Indeed, the MLSS model still has a high and highly significant R^2, while the F-statistic for the MLNSL model fails to reject the null that β^{MLNSL}, defined in equation (3.13), is equal to 0. Again, the model with the observed sentiment has no significant p-values.

As a last observation, we see the different behavior of the sectors in this regression exercise. Figure 3.4 reports the scatter plot of the elements of C_{short} versus the corresponding values of C_{ret} when the two stocks belong to the same economic sector, characterized by a specific marker. We also superimpose the regression line obtained from equation (3.11). Note that

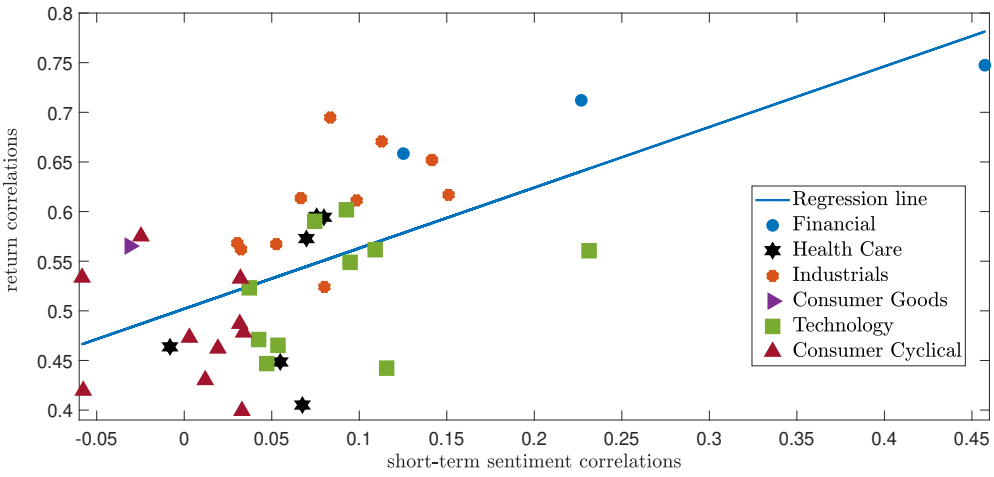

Figure 3.4: Scatter plot of the news short-term sentiment correlations and the return correlations for pairs of assets in the same sector. The line corresponds to the regression (3.11).

the behavior is different among sectors. The financial sector, marked with blue dots, is the one with the highest linear relation and the three assets belonging to this sector have all high returns and sentiment correlations. On the contrary, the consumer cyclical sector, marked with garnet-red triangles,

has a high dispersion among the correlations of the 5 assets.

In summary, Sections 3.4 and 3.4 support the intuition behind the MLSS model. Indeed, the slowly changing components of the sentiment are effectively captured by the long-term sentiment. We successfully confirmed this hypothesis in Section 3.4. At the same time, the short-term sentiment effectively describes the firm-specific behavior of the returns. Section 3.4 shows that the MLSS model can capture different features of the returns, while the MLNSL mainly captures the sentiment component associated with the market.

3.5 Contemporaneous and lagged relations

The goal of this section is to assess the explanatory power of the sentiment with respect to the market returns using the different filters presented in the previous sections. In particular, we show that both the extraction of long-term and short-term sentiment components and the multivariate specification of the model are crucial ingredients to capture the synchronous and lagged effects.

We consider the asset prices P_t^i of the 27 stocks of the Dow 30 and construct the equally weighted portfolio

$$M_t = \frac{1}{27} \sum_{i=1}^{27} P_t^i \tag{3.14}$$

as a representative portfolio and denote with r_t^m its log-returns. We consider a representative portfolio for two reasons. Firstly, (Beckers, 2018) shows that the returns predictability using sentiment indicators is higher when using market indexes rather than single stocks. Secondly, using a representative portfolio we can compare different filtering techniques which

do or do not consider the multivariate structure.

We define $\bar{S}_t^{\text{news}} = \frac{1}{27}\sum_{i=1}^{27} S_t^{i,news}$ and $\bar{S}_t^{\text{social}} = \frac{1}{27}\sum_{i=1}^{27} S_t^{i,social}$ as the sentiment associated to the representative portfolio. We consider five different filtering techniques defined as follow:

1. S_t^{MLSS} is the filtered signal obtained using the MLSS model in equation (3.6). The resulting filtered quantities are 4 long-term sentiment factors F_t^{MLSS}, 2 for the news and 2 for the social sentiment, and 54 short-term sentiment series Ψ_t^{MLSS}, 27 for the news and 27 for the social sentiment. We compute the cross-sectional average for the news short-term sentiment $\bar{\Psi}_t^{MLSS,news}$ and social short-term sentiment $\bar{\Psi}_t^{MLSS,social}$. As a final result, we define

$$S_t^{MLSS} = \left[\Delta F_t^{MLSS,news}, \Delta F_t^{MLSS,social}, \bar{\Psi}_t^{MLSS,news}, \bar{\Psi}_t^{MLSS,social}\right]' \in \mathbb{R}^6.$$

2. S_t^{LSS} is the filtered signal obtained applying the MLSS model directly to the univariate series $\bar{S}_t^{\text{news}}$ and $\bar{S}_t^{\text{social}}$. For identifiability reasons, the number of common factors is one. The motivation behind this model is to test whether a simple cross-sectional average of sentiment time series can be an effective proxy of the sentiment of the representative asset. This approach intentionally neglects the multivariate structure of the sentiment and treats it as a non-relevant feature. Similar reasoning has been used in (Borovkova et al., 2017). The resulting filtered quantities are 2 long-term sentiment factors F_t^{LSS}, one for the news and one for the social sentiment, and 2 short-term sentiment series $\bar{\Psi}_t^{LSS}$, one for the news and one for the social sentiment. The final model reads

$$S_t^{LSS} = \left[\Delta F_t^{LSS,news}, \Delta F_t^{LSS,social}, \bar{\Psi}_t^{LSS,news}, \bar{\Psi}_t^{LSS,social}\right]' \in \mathbb{R}^4.$$

3. S_t^{MLNSL} is the filtered signal obtained using the MLNSL model in equation (3.2) from the 54 observed sentiment time series. The resulting filtered quantities are 54 filtered sentiment series F_t^{MLNSL}, 27 for the news and 27 for the social sentiment. We compute the cross-sectional average for the news sentiment $\bar{F}_t^{MLNSL,news}$ and social sentiment $\bar{F}_t^{MLNSL,social}$. As a final result, we define

$$S_t^{MLNSL} = \left[\Delta \bar{F}_t^{MLNSL,news}, \Delta \bar{F}_t^{MLNSL,social}\right]' \in \mathbb{R}^2.$$

4. S_t^{LNSL} is the filtered signal obtained applying the LNSL model, introduced by (Borovkova and Mahakena, 2015) and presented in equation (3.1), to $\bar{S}_t^{\text{news}}$ and $\bar{S}_t^{\text{social}}$. As for the LSS model, the motivation behind this choice is to test whether the multivariate structure of sentiment is a relevant feature or not. We obtain two filtered sentiment series $\bar{F}_t^{LNSL}$, one for the news and one for the social sentiment. We then define

$$S_t^{LNSL} = \left[\Delta \bar{F}_t^{LNSL,news}, \Delta \bar{F}_t^{LNSL,social}\right]' \in \mathbb{R}^2.$$

5. S_t^{obs} only considers the observed sentiment $\bar{S}_t^{\text{news}}$ and $\bar{S}_t^{\text{social}}$

$$S_t^{Obs} = \left[\Delta \bar{S}_t^{\text{news}}, \Delta \bar{S}_t^{\text{social}}\right]' \in \mathbb{R}^2.$$

In summary, the five models allow us to separate the effect of the different components. The MLSS model exploits all the possible information from the multivariate time series and all the relevant common factors are considered. The average across assets is computed at a later stage on the short-term sentiment. For this reason, it does not affect the long-term components. The LSS model computes the cross-sectional average as a first step and does not exploit the multivariate structure. Then, both the short-term

and long-term components are different from the one of the MLSS model. The MLNSL and LNSL models differ only on the step of the aggregation. The first model applies the filter on the multivariate time series, while the second model applies the filter on the aggregated time series. Finally, the Obs model works as a benchmark.

Quantile regression

In this section, we investigate the lagged relation between sentiment and market returns. The recent literature for the DJIA (Garcia, 2013) and for the gold futures (Smales, 2014) found that the reaction to the news is more pronounced during recessions. For this reason, we use the quantile regression in place of a simple linear regression to obtain a more comprehensive analysis of the relationship between variables. In Appendix A.3 we report the investigation on the contemporaneous relation between sentiment and returns.

Lagged relations

We consider the following quantile regression

$$r^m(\tau) = \alpha(\tau) + \beta^{\text{model}}(\tau) S_{t-h}^{\text{model}},$$

where *model* denotes one of the five filtering models presented above. According to (Koenker and Machado, 1999), we can compare the explanatory power of a selected model according to the R^1 measure. In particular, if we consider the functional expression for the quantile regression

$$\hat{V}(\tau) = \min_{(\alpha,\beta)} \sum_{t=1}^{T} \rho_\tau \left(r_t^m - \alpha - \beta S_{t-h}\right), \tag{3.15}$$

where $\rho_\tau(u) = u(\tau - I_{u<0})$, we can define the quantile R^1 measure as

$$R^1(\tau) = 1 - \frac{\hat{V}(\tau)}{\tilde{V}(\tau)},$$

where $\tilde{V}(\tau)$ is evaluated restricting equation (3.15) with the intercept parameter only. In contrast with the R^2 measure of the linear models, $R^1(\tau)$ is a local measure of goodness of fit and only applies to a particular quantile. In addition, (Koenker and Machado, 1999) show that using $\hat{V}$ we can test the significance of the β^{model} parameters. However, the likelihood ratio test introduced in (Koenker and Machado, 1999) assumes that the residuals of the quantile regression are i.i.d. while the observed residuals show heteroskedasicity. For this reason, we perform a multivariate Wald test where the covariance structure of the estimators is estimated using a block bootstrap. For an introduction of confidence intervals estimations with bootstrap methods refer to MacKinnon (2006). To maintain the autocorrelation structure of the data, the block bootstrap length is set to 22 days (one month in financial time).

As a first step, we consider the lag $h = 1$. We evaluate the $R^1(\tau)$ statistic and test the significance using the χ^2 Wald-test. Table 3.3 reports the values and significance of the R^1 measure. A finding is common among all models: the values of R^1 are higher in the tails and lower close to the median. In addition, what we observe is extremely promising for the Long-Short modeling approach. The significance of the noisy sentiment is zero for almost all quantile levels. Filtering the time series is essential to recover predictability. However, filtering alone is not sufficient. Indeed, neither the predictability of the LSNL model nor of the multivariate extension MLSNL is statistically significant except for some cases. Significance is recovered only when the filtered sentiment is decomposed into the short-run

τ **quantiles**	$R^1(\tau)$ **measure**				
	MLSS	LSS	MLNSL	LNSL	Obs
0.01	12.7***%	4.5**%	0.3%	0.2%	0.1%
0.05	3.2**%	1.3%	0.1%	0.0%	0.1%
0.10	1.7**%	1.2**%	0.0%	0.0%	0.1%
0.33	0.2%	0.1%	0.0%	0.0%	0.0%
0.50	0.2**%	0.1**%	0.1**%	0.1**%	0.0%
0.66	0.4***%	0.2**%	0.1*%	0.1*%	0.0%
0.90	2.8***%	1.0**%	0.2***%	0.1%	0.1%
0.95	5.3***%	1.6**%	0.3*%	0.1%	0.2%
0.99	11.9***%	3.4**%	0.0%	0.5%	1.0**%

Table 3.3: The R^1 measure across the value τ for the one-lag quantile regression. We denote with *** the significance at 1%, ** the significance at 5% and * the significance at 10%

and long-run components. This is true for extreme returns, both positive and negative. The result is stronger when the LSS model is replaced by the MLSS, meaning that the cross-sectional dependence is an important ingredient to enhance predictability. It is worth noticing that the values of R^1 presented in Table 3.3 come from models with a different number of regressors, therefore an adjustment with respect to the number of regressors should be applied. However, given that the number of observations used to evaluate the measures is $T = 3020$, a correction in the spirit of the adjusted R^2 would be of order 0.2% and would not affect the results.

A further advantage of the long-short decomposition is that we can properly assess the relative contribution of the two components. In particular, we use a bootstrap Wald test to assess the significance of the parameters in the MLSS model. Considering the $S^{MLSS} = \left[\Delta F_t^{MLSS}, \bar{\Psi}_t^{MLSS}\right]$, the significance of the parameter $\beta^{LT} \in \mathbb{R}^4$ and $\beta^{ST} \in \mathbb{R}^2$ can be tested using

$$\tilde{V}^{LT}(\tau) = \min_{(\alpha,\beta^{LT})} \sum_{t=1}^{T} \rho_\tau \left(r_t^m - \alpha - \beta^{LT} \Delta F_{t-h}^{MLSS}\right)$$

τ **quantiles**	p**-values**	
	L^{ST}_{t-1}	L^{LT}_{t-1}
0.01	0.000***%	0.019***%
0.05	0.251***%	35.900%
0.10	0.170***%	5.700*%
0.33	3.300**%	28.500%
0.50	0.169***%	2.100**%
0.66	2.000**%	0.195***%
0.90	0.404***%	0.021***%
0.95	0.001***%	25.200%
0.99	0.003***%	20.800%

Table 3.4: p-values for the Wald test statistics. The statistics are asymptotically χ^2 with degrees of freedoms equal to the number of tested parameters.

and

$$\tilde{V}^{ST}(\tau) = \min_{(\alpha,\beta^{ST})} \sum_{t=1}^{T} \rho_\tau \left(r_t^m - \alpha - \beta^{ST} \bar{\Psi}_{t-h}^{MLSS} \right),$$

As before, we use a Wald test based on block bootstrap resampling to assess the significance of β^{ST} and β^{LT}.

We report the p-values of the Wald test statistics in Table 3.4. The contribution given by the short-term sentiment is strongly significant, in particular for extreme quantiles. On the contrary, the long-term sentiment is not significant in 6 out of 9 quantiles. The results support the intuition that, if today a very high or very low return appears, it can be partially explained by the yesterday's rapidly changing mood, while the permanent trend in the sentiment series have almost no impact.

The experiments performed in the contemporaneous (see Appendix A.3) and one-lag cases show that the MLSS model is the best model to capture the return variations. For this reason, for the multi-period analysis we will only consider the MLSS model.

Considering a general h, we wonder if extra lags can add explanatory power

τ	$h=2$	$h=3$	$h=4$	$h=5$
0.01	18.133%	29.136%	57.652%	72.784%
0.05	0.618%*	0.946%*	4.317%*	3.009%*
0.10	0.907%*	0.773%*	4.341%*	1.968%*
0.33	65.530%	47.389%	74.932%	74.071%
0.50	62.489%	70.078%	80.725%	90.581%
0.66	43.722%	53.518%	52.853%	74.962%
0.90	4.831%*	0.662%*	0.063%*	0.208%*
0.95	12.800%	2.504%*	0.628%*	2.468%*
0.99	38.580%	71.448%	81.945%	87.196%

Table 3.5: p-values for the Wald statistics for different values of h. We denote with * the values where β^2 is significantly different from zero.

to the regression exercise. Using the functional form

$$\hat{V}^{h,\mathrm{MLSS}}(\tau) = \min_{\Theta} \sum_{t=h+1}^{T} \rho_\tau \left(r_t^m - \alpha^0 - \alpha^1 r_{t-1}^m - \beta^1 S_{t-1}^{\mathrm{MLSS}} - \beta^2 \mathcal{L}_{h-1}(S_{t-1}^{\mathrm{MLSS}}) \right) ,$$

we separate the contributions given by the first and higher order lags. The vector $\Theta = (\alpha^0, \alpha^1, \beta^1, \beta^2)$. The bootstrap Wald test can be used to assess the null hypothesis that $\beta^2 = 0$.

Following (Tetlock, 2007; Garcia, 2013), we fix a maximum number of $h = 5$ and Table 3.5 reports the p-values for the different values of h. The h-lagged sentiment series are uninformative in the median region, where the one lag sentiment have less explanatory power too. However, in agreement with (Garcia, 2013), the lagged sentiment remains informative for a few days and, in our case, this is true for the 5%, 10%, 90%, and 95% quantile levels. It is worth noticing that the 1% and 99% quantiles are unaffected by higher-order lags. This shows that, in the case of very good or very bad days, the returns are strongly driven by very fresh news ($h = 1$) while the older news has no informative power.

3.6 Portfolio allocation with sentiment data

This section details an economic application of the MLSS model in portfolio selection and benchmarks the results against a buy-and-hold strategy. We consider the equally weighted portfolio in equation (3.14) and the five filtered signals S_t^{MLSS}, S_t^{LSS}, S_t^{MLNSL}, S_t^{LNSL} and S_t^{Obs} introduced in the previous section. It is worth noticing that (Beckers, 2018) and (Garcia, 2013) showed that the predictability power of the sentiment series declined after 2007. For this reason, we want to challenge the filtering techniques to predict the future daily returns in the time window 2007-2019.

In the first part of this section, we use the sentiment signals as exogenous variables to build a simple classifier and we introduce five trading strategies based on the five sentiment time series. Then, we test these strategies on the February 2007 - June 2017 window. This period offers a large series with different economic conditions. The sentiment models are estimated in the same time window. The estimation of multivariate models (MLSS and MLNLS) employs a backwards-looking technique based on smoothing recursions. Then, one may argue that for the multivariate case the estimation technique may introduce some sort of forward-looking bias. We test that this bias, if any, is not likely to be the dominant effect. We perform a robustness check where we use the parameter values from February 2007 - June 2017 period to filter the TRMI sentiment series from July 2017 to December 2019. In this way, the trading signals cannot be affected by any forward-looking bias. The results in the out-of-sample period confirm those from February 2007 - June 2017, showing that the trading strategies built on the MLSS model are the best performers. The details of the robustness check can be found in Appendix A.6.

Trading strategies

In the financial literature, several papers support the strong out-of-sample performance of the equally weighted portfolio (e.g. DeMiguel et al. 2009b). The $1/n$ portfolio without rebalancing is used as a baseline for our trading strategies and the long passive position in this portfolio is called *buy-and-hold* strategy. Given that the buy-and-hold portfolio offers good out-of-sample performance, we assume an investor who only deviates from the baseline strategy if a strong signal which predicts a negative return arrives from the sentiment series. For this reason, the criterion variable needs to capture the behavior of the left tail of returns distribution. We define the criterion binary variable as

$$Y_t = \begin{cases} 1, & \text{for } \tilde{r}_t^m < z_{1/3} \\ 0, & \text{otherwise} \end{cases}$$

where $z_{1/3}$ is the 1/3 Gaussian quantile and $\tilde{r}_t^m = r_t^m/\sqrt{RV_t}$ are the standardized market returns with the realized variance, RV_t, evaluated by means of 5-minute intraday returns. The standardization of the returns is crucial to eliminate possible effects due to the persistence of volatility. The choice of the 33% quantile is consistent with the findings of Section 3.5. Moreover, it is a balance between a more conservative choice – a smaller quantile only sensitive to more extreme and predictive events – and a larger quantile, which provides a larger number of selling signals but less predictive power.

Since the goal of this chapter is to show that the choice of the filtering procedure is essential, a simple classification technique is used. As a classifier, we consider the following conditional logit model

$$P\left(Y_{t+1} = 1 | X_t\right) = \text{logit}\left(X_t^{\text{mod}}\theta\right), \tag{3.16}$$

where $\text{logit}(X_t\theta) = \frac{e^{X_t\theta}}{1+e^{X_t\theta}}$ and $X_t^{\text{mod}} = \left[1, \tilde{r}_t^m, S_t^{\text{mod}}\right]$. We recall that S_t^{mod} is a vector whose dimension depends on the filtering model. For further details see the first part of Section 3.5.

The predicted binary value is defined as

$$\hat{Y}_{t+1}^{\text{mod}} = \begin{cases} 1, & \text{for } \text{logit}(X_t^{\text{mod}}\theta) > 0.5 \\ 0, & \text{otherwise} \ . \end{cases} \tag{3.17}$$

The main advantages of the conditional logit model are twofold. On one hand, the conditional logit model can be easily estimated using MLE. On the other hand, we can easily assess the fitness of the model on the data using the Mc Fadden's R^2 measure defined in (McFadden et al., 1974) as

$$R^2 = 1 - \frac{\log(L_m)}{\log(L_0)} \in [0, 1] \, .$$

L_m represents the maximum likelihood of the complete model (3.16) and L_0 is the maximum likelihood of the bare model based only on the intercept. The models are estimated using overlapping rolling windows of 6 months (126 observations). We verified that this choice is sufficient to capture the time-varying nature of the explanatory power of the sentiment series. Figure 3.5 shows the value of R^2 over the February 2007 - June 2017 period. The MLSS model has the highest R^2 w.r.t the other models, which typically translates in a higher predictive power. In addition, the MLSS R^2 has high variability, suggesting that the predictive power changes over time. This latter finding suggests that the sentiment signal can be a good returns predictor in certain periods and a poor predictor in others. This intuition will be exploited later to generate trading strategies based on the R^2.

The estimated $\bar{Y}_t^{\text{mod}}$ defined in (3.17) translates in the trading signal

$$s_{t+1}^{\text{mod}} = \begin{cases} 1, & \text{if } \hat{Y}_{t+1}^{\text{mod}} = 0 \\ -1, & \text{if } \hat{Y}_{t+1}^{\text{mod}} = 1 \end{cases} \tag{3.18}$$

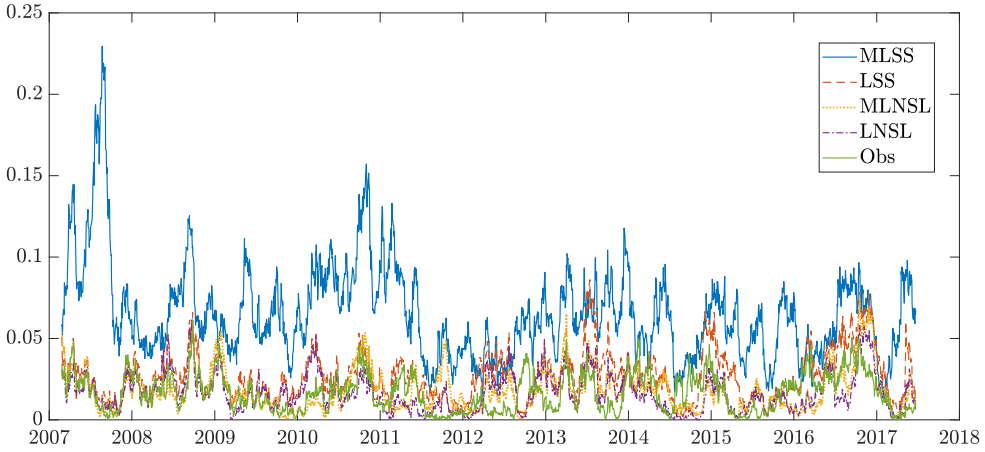

Figure 3.5: McFadden's R^2 for the different filtering methods using negative abnormal returns.

where $s_{t+1}^{\text{mod}} = 1$ ($s_{t+1}^{\text{mod}} = -1$) represents a buy (sell) signal in the equally weighted portfolio (3.14). At any day t, at the closing time of the trading day, the investor uses the sentiment signal S_t^{mod} and the standardized realized daily returns $\tilde{r}_t^m$ to forecast the binary variable $\hat{Y}_{t+1}^{\text{mod}}$ and the relative trading signal. Naming c_0 the number of shares bought or sold in any transaction, there are three possible scenarios

1. $s_t^{\text{mod}} = s_{t+1}^{\text{mod}}$: In this case the prediction on the future realization does not change and the investor does not re-balance the portfolio.

2. $s_t^{\text{mod}} = +1$ and $s_{t+1}^{\text{mod}} = -1$: The investor had a long position in the equally weighted portfolio at time t but the prediction changed. She sells the current position and short sells c_0 shares of the same portfolio.

3. $s_t^{\text{mod}} = -1$ and $s_{t+1}^{\text{mod}} = +1$: The investor had a short position in the equally weighted portfolio at time t but the prediction changed. She buys $2c_0$ shares of the portfolio.

Please notice that the only exception is for s_1^{mod} because we initialized $s_0^{\text{mod}} = 0$. In this case, the equally weighted portfolio is bought when $s_1^{\text{mod}} = 1$ and it is short sold when $s_1^{\text{mod}} = -1$. The investor's portfolio is then built as

$$\begin{cases} P_{t+1}^{\text{mod}} = s_{t+1}^{\text{mod}} c_0 M_{t+1} + \text{cash}_{t+1}, \\ \text{cash}_{t+1} = \text{cash}_t - (s_{t+1}^{\text{mod}} - s_t^{\text{mod}}) c_0 M_{t+1} - |s_{t+1}^{\text{mod}} - s_t^{\text{mod}}| c_0 M_{t+1} \frac{\text{cost}}{2}, \end{cases} \tag{3.19}$$

where *cost* is the percentage trading cost and M_t is defined in (3.14). The first equation in (3.19) shows that the value of the portfolio is composed by the value of the invested amount $s_{t+1}^{\text{mod}} c_0 M_{t+1}$ plus the cash position. The latter increases when $s_{t+1}^{\text{mod}} < s_t^{\text{mod}}$, meaning that the investor sells the portfolio and receives cash, and decreases when $s_{t+1}^{\text{mod}} > s_t^{\text{mod}}$, meaning that the investor buys and erodes the cash position. The second equation includes the impact of the transaction costs. Specifically, every time that a transaction happens, i.e. $s_{t+1}^{\text{mod}} \neq s_t^{\text{mod}}$, the investor pays an extra cost proportional to the current value of the equally weighted portfolio M_{t+1}.

We fix the starting point $s_0^{\text{mod}} = 0$, $\text{cash}_0 = 100,000\$$ and the parameter $c_0 = 100,000\$/M_0$. In the chapter we only report the results for the case with trading costs, while the results with zero trading costs are reported in Appendix A.5. From now on, we refer to *without trading costs* when the portfolio in equation (3.19) is evaluated with cost = 0 and to *with trading costs* when costs = 0.1% as in (Gilli and Schumann, 2009) and (Avellaneda and Lee, 2010). In the following sections, the number of transactions is evaluated as $Tr^{\text{mod}} = \sum_{i=0}^{T-1} |s_{i+1}^{\text{mod}} - s_i^{\text{mod}}|$ and the transaction costs are evaluated as $Tc^{\text{mod}} = \sum_{i=0}^{T-1} |s_{i+1}^{\text{mod}} - s_i^{\text{mod}}| c_0 M_{i+1} \frac{\text{cost}}{2}$. It is worth noticing that the change of signal effectively produces two transactions. For instance, if the signal moves from $s_t = 1$ to $s_{t+1} = -1$, the first transaction is the

liquidation of the long position and the second transaction is the short position on the asset. In addition, most of the time the selling signal appears for only one day and disappears the day after. Then, the typical path of a selling signal is given by $s_t = 1$, $s_{t+1} = -1$ and $s_{t+2} = 1$ producing a total of four transactions.

The transaction costs can strongly depress the overall performance of the portfolio. To partially mitigate this drawback, we can decrease the number of transactions using the McFadden's R^2 as a measure of the reliability of the signal $\hat{Y}_t^{\text{mod}}$. We compute the empirical quantile $z_\alpha^{1,t}(R^2)$ of the McFadden R^2 over the time window $(1, \cdots, t)$. The quantile $z_\alpha^{1,t}(R^2)$ is $\mathcal{F}_t$-measurable and does not introduce a forward looking bias. We can reduce the number of trades conditioning the selling signal at time t on the level of the McFadden's R^2 evaluated in the previous 6 months. The R^2 adjusted trading signal is then defined as follows

$$\bar{s}_t^{\text{mod}} = \begin{cases} -1, & \text{if } \hat{Y}_{t+1}^{\text{mod}} = 1 \text{ and } R_t^{2,\text{mod}} \geqslant z_\alpha^{1,t}\left(R^{2,\text{mod}}\right) \\ 1, & \text{otherwise .} \end{cases} \tag{3.20}$$

The value α determines the reduction in the number of trades. The higher α is, the smaller is the number of transactions. The parameter α should be considered as a subjective transaction cost of the investor. The five strategies, together with the buy-and-hold strategy itself, are evaluated according to six measures, the *annual return*, the *annual volatility*, the *annual negative volatility*, the *Sharpe ratio*, the *Sortino ratio*, and the *maximum drawdown* (MDD). In the next section, in the first step, the portfolios with the trading signals (3.18) with and without trading cost are analysed. Then, we assess the impact and the performance of the trade reduction strategy based on (3.20).

Empirical application: February 2007 - June 2017

The 2007-2009 crisis and the 2009-2017 bull market are good backtesting periods for the sentiment portfolios because we can test the return predictability during different market conditions.

Table 3.6 reports the performances of the five sentiment strategies together with the buy-and-hold portfolio with trading costs. The sentiment-based strategies have, excluding the LNSL and the Obs, smaller volatility and MDD than the buy-and-hold portfolio. In addition, the MLSS portfolio produces returns similar to the buy-and-hold strategy, lower negative volatility, and consequently higher Sharpe and Sortino ratios than all the other strategies. The lower performance for the annual returns is due to the higher transaction costs. Indeed in Appendix A.5 we show that, when the trading costs are not considered, the MLSS strategy produces higher annual returns than all the other strategies. In addition, when we compare *without trading costs* experiment with the *with trading costs* experiment, the excessive number of transactions for the MLSS strategy reduces the Sharpe ratio gain with respect to the buy-and-hold portfolio from 40% to 10% and the Sortino ratio gain from 48% to 16%. In Appendix B.1 we show that the selling signal generated by the MLSS sentiment series corresponds to statistically significant returns predictability. The transaction costs incurred by the MLSS portfolio throughout the nine years amount in total to 38% of the starting capital. For this reason, we employ the trading signal $\bar{s}^{\text{MLSS}}$ defined in equation (3.20), which penalizes signals with moderate McFadden's R^2. Table 3.7 reports the performances of the strategies based on the penalized signal for different values of α. As expected, the higher the value of α and the lower the number of transactions is. In addition, the R^2-based signal produces a higher quality signal and effectively increases the performance

Measures	BH	MLSS	LSS	MLNSL	LNSL	Obs
A. return (%)	**8.975**	7.891	6.977	8.882	8.143	6.986
A. volatility (%)	19.132	**15.209**	18.136	17.952	19.431	19.955
A. neg. volatility (%)	15.523	**11.767**	14.339	14.474	15.374	16.055
A. Sharpe ratio	0.469	**0.519**	0.385	0.495	0.419	0.35
A. Sortino ratio	0.578	**0.671**	0.487	0.614	0.53	0.435
MDD ($)	59377	57235	50335	**49773**	63595	62785
Number of trades	1	553	161	81	73	93
Transaction costs ($)	50	37974	14866	5565	4085	7544

Table 3.6: Performances of the six strategies with transaction cost for the period February 2007 - June 2017. In bold, the best performance per row. BH is the buy-and-hold portfolio, while MLSS, LSS, MLNSL, LSNSL, and Obs correspond to portfolios built from the corresponding model for the sentiment time series.

Measures	BH	$\alpha = 0\%$	$\alpha = 20\%$	$\alpha = 35\%$	$\alpha = 50\%$	$\alpha = 65\%$	$\alpha = 80\%$
A. return (%)	8.975	7.891	8.225	9.575	9.84	**10.248**	9.184
A. volatility (%)	19.132	15.209	15.679	14.083	13.601	**13.538**	17.201
A. neg. volatility (%)	13.601	10.888	11.196	9.901	9.511	**9.443**	12.216
A. Sharpe ratio	0.469	0.519	0.525	0.680	0.723	**0.757**	0.534
A. Sortino ratio	0.660	0.725	0.735	0.967	1.035	**1.085**	0.752
MDD ($)	59377	57235	63522	49160	**33264**	35600	59486
Number of trades	1	553	437	349	273	169	57
Transaction costs ($)	50	37974	30626	25283	20074	12007	4127

Table 3.7: Performances of the MLSS based strategies built from equation (3.20) for different values of α $\times 100$. BH is the buy-and-hold portfolio. In bold, the best performance per row.

of the portfolios. The number of transactions decreases almost linearly but the Sharpe and Sortino ratios strongly increase. They reach a maximum value when $\alpha = 0.65$. These findings further corroborate the intuition that the MLSS sentiment strongly anticipates future returns during the financial crisis, given that the R^2 values in figure 3.5 are higher than the unconditional average during the $2007 - 2009$ period. Again this feature is peculiar for the MLSS filter while no evidence of return predictability is reported for the other filtering techniques. Again, the statistical significance of these strategies is reported in Appendix B.1.

3.7 Conclusions

In this chapter, we presented a novel way to filter multivariate sentiment time series. The approach is very general and encompasses previous models discussed in the literature. Using a dynamic factor model, we were able to identify two different sentiment components. The first one, named long-term sentiment and modeled as a random walk, captures the common trends which drive the long-term dynamics. The second component, dubbed short-term sentiment and modeled as a VAR(1) process, captures short-term swings of market mood. An extensive empirical section investigates the different features of the two sentiment components. In a first analysis, we pointed out that one of the long-term sentiment factors cointegrates with the first principal component of the market. Quite surprisingly, the structure of the sentiment factor loadings does not mimic the typical uniform profile of the market factor. Some assets are over-expressed and contribute to the factor with a positive or negative sign, while others are under-expressed. Concerning the short-term sentiment, its multivariate dependence structure explains a sizable fraction of the residual covariance in a single factor market model. This result suggests that the short-term component captures transient and rapidly changing trends associated with the idiosyncratic components of the market. In a second analysis, based on quantile regression, we showed that the Multivariate Long-Short Sentiment model provides the highest explanatory power of lagged and contemporaneous returns. Essential to achieve statistical significance is the multivariate nature of the approach and the separation of the sentiment signal in a long and a short components. In particular, disentangling the short-term sentiment is crucial to capture the behavior of extreme returns. In a further analysis, we observed that newspapers and social media differently react

to negative and positive returns. Specifically, they can effectively explain abnormal returns from one to five days in advance, but they almost immediately digest the positive market realizations while they echo negative realizations for several days to come.

It is worth noting that (Tetlock, 2007) and (Garcia, 2013) reported results similar to ours for the unfiltered sentiment focusing on the period before 2007. Using the TRMI dataset, (Beckers, 2018) showed that the forecasting power on returns of the sentiment dropped dramatically after 2007. Our results suggest that the filtering procedures are more important nowadays than in the past. Consistently, in a final investigation, we performed an asset allocation exercise where the selling signals are based on the sentiment series. In line with results from the quantile regression, the portfolio based on the MLSS filter significantly outperforms the benchmark buy-and-hold strategy and the other strategies based on different filtering techniques.

Supplementary materials

The supplementary materials include additional analysis to support the evidences found in the chapter. Section A.1 provides an overview of the stocks used in the empirical analysis. Section A.2 compares the different signal-to-noise ratios filtered by the MLSS and MLNSL model. Section A.3 investigates the contemporaneous relation between the sentiment and return series. Appendices A.5, A.6, and B.1 report the *without trading costs* analysis, the robustness check, and the statistical significance of the portfolio allocation exercise, respectively.

Chapter 4

Modeling Realized Covariance with score-driven dynamics

The contents of this chapter are the result of a joint work with Giuseppe Buccheri and Prof. Fulvio Corsi, also available as published article in Vassallo et al. (2021).

4.1 Introduction

Covariance modeling and forecasting is a prominent topic in many financial applications. Since the seminal work of Andersen and Bollerslev (1998), the use of realized measures computed from intraday data has emerged as a preferential tool to forecast volatilities and correlations. Several univariate specifications for realized volatility have appeared in the econometric literature; see, among others, Engle and Gallo 2006, Corsi 2009, Shephard and Sheppard 2010, Hansen et al. 2012. More recently, a parallel interest arose towards multivariate models. Notably examples are given by Chiriac and Voev 2011, Bauer and Vorkink 2011, Bonato et al. 2012, Bauwens et al.

2012, Callot et al. 2017, Gorgi et al. 2018, Opschoor et al. 2017. The dynamic modeling of realized covariance poses a number of challenges that are absent in a univariate framework. The curse of dimensionality, i.e. the fact that the number of parameters rapidly increases with the cross-section dimension, might significantly affect the statistical efficiency of the estimation. Another complication is the need to guarantee positive-definite estimates, which is relevant in financial applications. Furthermore, estimation errors on realized measures play a crucial role in a multivariate framework, as they can lead to unstable solutions in portfolio optimization (Jagannathan and Ma 2003).

In this chapter, we introduce a methodology aimed at addressing the above mentioned difficulties. Inspired by the DCC model of Engle (2002a), we apply a two-step estimation procedure to a class of score-driven realized covariance models based on Wishart and matrix-F distributions. In the new models, the dynamics of *correlations* are driven by the score of the conditional density. Volatilities are instead estimated in a previous step through a sequence of univariate models. Such an approach inherits the main advantages of the DCC. In particular, it allows modeling separately volatilities and correlations, which possess different dynamics and leave more flexibility in the estimation of the model. Each univariate specification is estimated independently from the others, and it is therefore characterized by different parameters. In a similar fashion to the DCC, in the second step of the estimation, a set of restrictions on the parameter space can be imposed in such a way that the number of parameters scales well with the cross-section dimension. The new models are thus less affected by the typical curse of dimensionality problem of multivariate realized covariance models. Conditions guaranteeing positive-definite covariance estimates are easily derived.

Bauwens et al. (2012)[1] introduce a related DCC approach to realized covariance modeling based on the Wishart density. The proposed approach differs in the method used to update the time-varying parameters. In our framework, the update rule is determined by the score of the conditional density (Creal et al. 2013, Harvey 2013). In the case of the Wishart density, if the score is scaled through the inverse of the Fisher information matrix, our method leads to the same update rule of Bauwens et al. (2012). If a matrix-F density is instead used, we obtain a different update rule, which is robust to outliers. In our empirical application, we show that the model based on matrix-F density provides a better fit to the data and is superior in terms of out-of-sample forecast performance. It is worth noticing that the score of the conditional likelihood is by construction a martingale difference. For this reason, the score-driven update automatically provides a correction to the DCC model analogous to that introduced by Aielli (2013). In particular, it turns out that the specific correction of Aielli (2013) is recovered in the case of the Wishart density, whereas if the matrix-F density is used, the correction assumes a different form.

Score-driven models are a large class of observation-driven models (Cox 1981) where the time-varying parameters are driven by the score of the conditional density. They have been successfully applied in the recent econometric literature. Notably examples are given by Creal et al. (2011), Creal et al. (2014) and Oh and Patton (2018). Gorgi et al. (2018) and Opschoor et al. (2017) introduce a score-driven specification for realized covariances based on Wishart and matrix-F distributions, respectively. In both cases, variances and correlations are estimated in a single step. In contrast, we estimate such models using the above mentioned two-step procedure. We

[1]See also their extensions in Bauwens et al. (2016) and Bauwens et al. (2017).

show that this leads to significant forecast gains in common empirical applications as a result of the higher level of flexibility in parameter estimation. The choice of the score-driven approach as a starting point of our work is motivated by three main reasons. First, score-driven models provide a general methodology to update the time-varying parameters based on the full shape of the conditional density function. We thus obtain an update mechanism for volatilities and correlations which is consistent with the choice of the observation density. Second, this approach acknowledges the existence of measurement errors on realized covariance estimates. The latters are indeed modeled as noisy observations of the true latent covariances. The relevance of measurement errors on forecasting with realized measures has been underlined by Bollerslev et al. (2016), Bollerslev et al. (2018), Bekierman and Manner (2018) and Buccheri and Corsi (2019). In the empirical application, we compare our methodology to alternative methods that ignore estimation errors and show the resulting advantages in forecasting. Finally, in a score-driven framework, the likelihood can be written in closed form and can be optimized numerically. Estimation is thus feasible even when dealing with high dimensional matrices. General state-space models with latent covariances can only be estimated by simulation-based techniques, which become infeasible at high dimensions.

Our Monte-Carlo study has three main objectives. First, we investigate the finite sample properties of the maximum likelihood estimator that is employed in the second step of our procedure. We find that, as the number of observations increases, the estimator becomes unbiased and its distribution concentrates around the true parameters. Second, we simulate the covariances through a misspecified DGP and, according to the empirical evidence that realized covariances have fat-tails, we generate observations

from a matrix-F distribution. We thus misspecify the measurement density of the score-driven Wishart model. We find that both models can capture the misspecified dynamics of the covariances. However, the relative performance of the matrix-F model is superior, as it is robust to the outliers generated by the fat-tailed measurement density. As a final experiment, we compare the two-step estimated covariances with those resulting from joint estimation. To mimic a realistic setting, variances are generated based on realized measures computed from real data. We clearly find, both in-sample and out-of-sample, that the two-step procedure provides significantly lower loss measures.

We illustrate the advantages of the proposed method in an empirical study involving two different datasets. The first dataset includes 1-second resolution transactions of 100 NYSE stocks. The second includes 1-minute transaction data of 2767 stocks belonging to the Russell 3000 index. Score-driven realized covariance models are typically estimated by restricting the volatility persistences to be the same across assets. We first motivate the newly introduced class of models by performing several empirical tests which show that such assumption is too restrictive on real data. We then examine the in-sample and out-of-sample forecast performance of the two-step models and compare them to models based on joint estimation and to other benchmarks. We select groups of 5, 10, 25, 50, 100 assets. The proposed two-step procedure provides significant in-sample and out-of-sample forecast gains compared to models based on joint estimation. In particular, the Model Confidence Set of Hansen et al. (2011) is effective in selecting the two-step Wishart and matrix-F models and excluding the remaining alternatives. Notably, forecast gains over joint estimation based models become more significant as the cross-section dimension increases, given the higher

level of heterogeneity in the persistences of realized variances. Such results confirm that the additional flexibility provided by the two-step procedure translates into better forecasts. We also find that, in the vast majority of the scenarios examined in the analysis, the model based on the matrix-F density performs better than that based on the Wishart density. We thus confirm the result of Opschoor et al. (2017) that the matrix-F density is more suited to model time-series of realized covariances. Two additional empirical experiments are performed in order to assess the advantages of the two-step approach. In the first experiment, we show the behaviour of the methodology as estimation errors on realized covariance measures become more severe. We find that the model forecasts are significantly less affected by the noise compared to alternative methods which ignore estimation errors. As a second test, we assess the *economic* gains of switching from joint estimation based models to the proposed two-step models. By adopting a utility-based framework similar to that of Fleming et al. (2001), Fleming et al. (2003) and Bollerslev et al. (2018), we show that a risk-averse investor is willing to pay a positive annual amount to employ the forecasts of the two-step models in constructing her portfolio. Essentially, these economic gains are imputable to the lower ex-post risk featured by the portfolios constructed with the proposed methodology.

The rest of this chapter is organized as follows: Section (4.2) introduces the two-step estimation procedure and reports the main results; Section (4.3) shows the results obtained through the Monte-Carlo analysis; Section (4.4) describes the empirical application and reports the main results; Section (4.5) concludes.

4.2 Framework

Score-driven models for realized covariance

Let $\{X_t\}_{t=1}^T$ denote a sequence of $k \times k$ positive-definite realized covariance matrices and let $\mathcal{F}_t = \sigma(X_s : s \leqslant t)$ be the σ-field generated by past observations of X_t. We assume that the conditional distribution function of X_t is given by:

$$X_t|\mathcal{F}_{t-1} \sim \mathcal{G}_k(V_t) \tag{4.1}$$

where $\mathcal{G}_k$ is a matrix-variate distribution and $V_t \in \mathbb{R}^{k\times k}$ is a latent covariance matrix. Let us define $q = \frac{k(k+1)}{2}$ and let $f_t = \mathsf{vech}(V_t) \in \mathbb{R}^q$ collect the diagonal and upper diagonal elements of V_t. In the score-driven framework of Creal et al. (2013) and Harvey (2013), f_t is modeled as:

$$f_{t+1} = \omega + As_t + Bf_t, \tag{4.2}$$

where:

$$s_t = (\mathcal{I}_{t|t-1})^{-1}\nabla_t, \quad \nabla_t = \frac{\partial \log p_{\mathcal{G}_k}(X_t, f_t)}{\partial f_t}, \quad \mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t'] \tag{4.3}$$

Here, A and B are $q\times q$ matrices, $p_{\mathcal{G}_k}(X_t, f_t)$ denotes the conditional density function associated with $\mathcal{G}_k$, ∇_t is the score computed with respect to f_t and $\mathcal{I}_{t|t-1}$ is the Fisher information matrix. The score-driven framework provides two main advantages. First, the parameters in f_t evolve based on the full shape of the conditional density function. Different choices of $p_{\mathcal{G}_k}(X_t, f_t)$ thus determine different update rules. Second, the likelihood can be written in closed form and estimation can be performed by standard numerical optimization.

Two different specifications for the conditional density $\mathcal{G}_k$ have been proposed in the literature. Gorgi et al. (2018) set:

$$X_t|\mathcal{F}_{t-1} \sim W_k(V_t/\nu, \nu) \tag{4.4}$$

where $W_k\left(V_t/\nu,\nu\right)$ is the k-variate Wishart distribution with $\nu \geq k$ degrees of freedom. The corresponding density function is given by:

$$p_{W_k}\left(X_t;V_t/\nu,\nu\right)=\frac{|X_t|^{\frac{\nu-k-1}{2}}}{2^{\nu k/2}\nu^{-(\nu k)/2}|V_t|^{\nu/2}\Gamma_k\left(\frac{\nu}{2}\right)}\exp\left[-\frac{\nu}{2}\mathrm{tr}\left(V_t^{-1}X_t\right)\right], \quad (4.5)$$

where $\Gamma_k(\cdot)$ denotes the k-variate Gamma function. In this parameterization, the conditional mean is given by $\mathrm{E}_{t-1}[X_t]=V_t$. A similar density is employed by Golosnoy et al. (2012), Gourieroux et al. (2009) and Bonato et al. (2012), who introduce autoregressive processes for realized covariances based on the Wishart distribution.

The dynamics of realized covariances are often characterized by fat-tails and jumps. The Wishart distribution is not able to reproduce such features and, as a consequence, covariance estimates might be too sensitive to large movements. To obtain robust estimates, Opschoor et al. (2017) propose to model X_t through the matrix-F distribution:

$$X_t|\mathcal{F}_{t-1}\sim F_k\left(V_t,\nu_1,\nu_2\right), \quad (4.6)$$

where $\nu_1,\nu_2\geq k$ are degrees of freedom. The density function of the matrix-F distribution is given by:

$$p_{F_k}(X_t;V_t,\nu_1,\nu_2)=K(\nu_1,\nu_2)\frac{\left|\frac{\nu_1}{\nu_2-k-1}V_t^{-1}\right|^{\nu_1/2}|X_t|^{\frac{\nu_1-k-1}{2}}}{\left|\mathbb{I}_k+\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\right|^{\frac{\nu_1+\nu_2}{2}}} \quad (4.7)$$

where $\mathbb{I}_k$ denotes the $k\times k$ identity matrix and:

$$K(\nu_1,\nu_2)=\frac{\Gamma_k\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma_k\left(\frac{\nu_1}{2}\right)\Gamma_k\left(\frac{\nu_2}{2}\right)} \quad (4.8)$$

The $k\times k$ positive-definite matrix V_t turns out to be the conditional mean of the matrix-F distribution. The Wishart distribution can be obtained as a limit from the matrix-F when ν_2 goes to infinity. If ν_2 is finite, the matrix-F

distribution exhibits fat-tail behavior and it is thus more suited to describe the dynamics of realized measures. Indeed, the score computed from the matrix-F density exhibits a nonlinear structure which underweights the impact of outliers on the dynamics of the covariances (see also discussions in Opschoor et al. 2017).

The main shortcoming of such modeling approach is that the dimension of f_t grows quadratically with k. Estimation is thus feasible only by imposing strong restrictions on the structure of the matrices A, B. For instance, Gorgi et al. (2018) and Opschoor et al. (2017) impose a scalar structure, namely they set $A = \alpha \mathbb{I}_q$ and $B = \beta \mathbb{I}_q$, with α, $\beta \in \mathbb{R}$. Another drawback is that long-memory effects are not directly taken into account. To this end, Opschoor et al. (2017) impose a HAR specification in the dynamic in Eq. (4.2). However, the additional parameters are estimated under the same scalar restrictions. Our objective is to relax such restrictions while maintaining the model computationally simple to estimate.

DCC-type score-driven models for realized covariance

Inspired by Engle (2002a), we write the covariance matrix V_t as:

$$V_t = D_t R_t D_t \tag{4.9}$$

where D_t is a diagonal matrix of standard deviations and R_t is a correlation matrix. We propose to estimate D_t and R_t in two steps. In the first step, the individual standard deviations are separately estimated by a sequence of k realized volatility models. We allow for complete flexibility in the choice of the univariate specification. In our applications, we choose the univariate distribution as the marginal density function for the diagonal elements in Eq. (4.5) and (4.7). In the case of the Wishart density, we obtain, for

$i = 1, \dots, k$:

$$x_t^{(i)} | \mathcal{F}_{t-1} \sim W_1 \left(\frac{v_t^{(i)}}{\nu^{(i)}}, \nu^{(i)} \right) \sim \frac{v_t^{(i)}}{\nu^{(i)}} \chi^2_{\nu^{(i)}} \tag{4.10}$$

where $x_t^{(i)}$ and $v_t^{(i)}$ denote the i-th diagonal element of X_t and V_t, respectively, and $\chi^2_{\nu^{(i)}}$ is a chi-squared distribution with $\nu^{(i)}$ degrees of freedom. In the case of the matrix-F density, we obtain, for $i = 1, \dots, k$:

$$x_t^{(i)} | \mathcal{F}_{t-1} \sim F_1 \left(v_t^{(i)}, \nu_1^{(i)}, \nu_2^{(i)} \right) \sim \frac{\nu_2^{(i)} - 2}{\nu_2^{(i)}} v_t^{(i)} F_{\nu_1^{(i)}, \nu_2^{(i)}} \tag{4.11}$$

where $F_{\nu_1^{(i)}, \nu_2^{(i)}}$ denotes the scalar F density with degrees of freedom $\nu_1^{(i)}$ and $\nu_2^{(i)}$. To guarantee positive variance estimates, we set:

$$v_t^{(i)} = e^{\lambda_t^{(i)}} \tag{4.12}$$

and model the log-variance $\lambda_t^{(i)}$ through the score of the conditional likelihood. Specifically, the dynamic of $\lambda_t^{(i)}$ is given by:

$$\lambda_{t+1}^{(i)} = \omega_i + \alpha_i s_t^{(i)} + \beta_i^{(d)} \lambda_t^{(i)} + \beta_i^{(w)} \lambda_{t-1|t-5}^{(i)} + \beta_i^{(m)} \lambda_{t-6|t-22}^{(i)} \tag{4.13}$$

where $\lambda_{t-m|t-n}^{(i)} = \frac{1}{n-m+1} \sum_{j=t-n}^{t-m} \lambda_j^{(i)}$ and $s_t^{(i)}$ denotes the scaled score of the i-th univariate density. The expression of $s_t^{(i)}$ depends on the choice of the probability density function and is computed in Appendix B.2. The HAR-like structure allows to parsimoniously capture the strong persistence observed on realized variance. After estimating the k univariate models, we compute the diagonal elements of the matrix D_t as:

$$D_t^{(i)} = \exp \left[\frac{\lambda_t^{(i)}}{2} \right] \tag{4.14}$$

Thus, in the first step, k univariate score-driven models are independently estimated and their forecasts are used to build the matrix D_t of

standard deviations. This operation is extremely simple from a computational viewpoint, and at the same time allows to have different parameters $\omega_i, \beta_i^{(d)}, \beta_i^{(w)}, \beta_i^{(m)}, \nu^{(i)}, \nu_1^{(i)}, \nu_2^{(i)}$ for different assets.

To model the correlations, as in Engle (2002a), we introduce the matrix Q_t such that:

$$R_t = \Delta_t^{-1} Q_t \Delta_t^{-1} \tag{4.15}$$

where $\Delta_t = \mathsf{diag}\left(Q_t\right)^{1/2}$. We then model dynamically $f_t = \mathsf{vech}(Q_t)$ based on the score of the conditional density:

$$f_{t+1} = \omega + A s_t + B f_t \tag{4.16}$$

where the scaled score s_t is computed by assuming D_t known and given by Eq. (4.14). In the next two subsections, we compute the expression of s_t for both the Wishart and the matrix-F densities.

Wishart

The log-density function associated with the density in Eq. (4.5) is:

$$\log p_{W_k}\left(X_t; V_t, \nu\right) = \frac{1}{2} d_X\left(k, \nu\right) + \frac{\nu - k - 1}{2} \log\left|X_t\right| - \frac{\nu}{2} \log\left|V_t\right| - \frac{\nu}{2} \mathsf{tr}\left(V_t^{-1} X_t\right) \tag{4.17}$$

where $d_X\left(k, \nu\right) = \nu k \log\left(\nu/2\right) - 2 \log \Gamma_k\left(\nu/2\right)$. In Appendix B.3, B.4 we prove the following two results:

Proposition 1. *For the density in Eq. (4.17), the score* $\nabla_t^W = \frac{\partial \log p_{W_k}(X_t; f_t, \nu)}{\partial f_t}$ *is given by:*

$$\nabla_t^W = \frac{\nu}{2} \mathcal{D}_k' \Psi_t' \left(D_t^{-1} \Delta_t Q_t^{-1} \otimes D_t^{-1} \Delta_t Q_t^{-1}\right) \left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \tag{4.18}$$

where $\mathcal{D}_k$ *denotes the duplication*[2] *matrix,* $\Psi_t = \mathbb{I}_{k^2} - \left(\Delta_t^{-1} Q_t \otimes \mathbb{I}_k + \mathbb{I}_k \otimes \Delta_t^{-1} Q_t\right) W_Q$ *and* W_Q *is a sparse* $k^2 \times k^2$ *diagonal matrix defined in Appendix B.3.*

[2]The duplication matrix $\mathcal{D}_k$ is the unique $k^2 \times \frac{k(k+1)}{2}$ matrix such that, for any symmetric matrix A, $\mathcal{D}_k \mathsf{vech}\left(A\right) = \mathsf{vec}\left(A\right)$

Proposition 2. *For the density in Eq. (4.17), the Fisher information matrix* $\mathcal{I}^W_{t|t-1} = \mathrm{E}_{t-1}[\nabla^W_t \nabla^{W\prime}_t]$ *is given by:*

$$\mathcal{I}^W_{t|t-1} = \frac{\nu}{2}\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^+\Psi_t\mathcal{D}_k \tag{4.19}$$

where $\mathcal{D}_k^+$ *denotes the elimination*[3] *matrix and* $H_t = D_t\Delta_t^{-1}$.

Taking the inverse of $\mathcal{I}^W_{t|t-1}$, as required by Eq. (4.2), poses two problems. First, $\mathcal{I}^W_{t|t-1}$ is singular, as f_t includes $k(k+1)/2$ time-varying parameters whereas the number of time-varying parameters in R_t is $k(k-1)/2$. Second, for $k \gg 1$, matrix pseudo-inversion is computationally cumbersome and can lead to numerical instabilities. We solve both problems by setting $\mathcal{I}^W_t = \frac{\nu}{2}\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k$. This is equivalent to approximate the second Ψ_t factor appearing in Eq. (4.19) as $\Psi_t \approx \mathbb{I}_{k^2}$ (note that $\mathcal{D}_k^+\mathcal{D}_k = \mathbb{I}_k$), thus neglecting the term $\left(\Delta_t^{-1}Q_t \otimes \mathbb{I}_k + \mathbb{I}_k \otimes \Delta_t^{-1}Q_t\right)W_Q$. The latter has indeed a tiny influence on Ψ_t, as W_Q is a very sparse $k^2 \times k^2$ matrix with only k nonzero elements on the main diagonal. By assuming[4] such expression for $\mathcal{I}^W_{t|t-1}$, not only the latter becomes non-singular, but one can also compute in closed form the product $(\mathcal{I}^W_{t|t-1})^{-1}\nabla^W_t$ appearing in Eq. (4.2). In particular, in Appendix B.5 we prove the following:

Proposition 3. *For the density in Eq. (4.17), the scaled score vector* $s^W_t = (\mathcal{I}^W_{t|t-1})^{-1}\nabla^W_t$ *is given by:*

$$s^W_t = \mathsf{vech}(H_t^{-1}X_tH_t^{-1}) - \mathsf{vech}\,(Q_t) \tag{4.20}$$

[3]The elimination matrix $\mathcal{D}_k^+$ is s $\frac{k(k+1)}{2} \times k^2$ matrix such that, for any matrix A, $\mathcal{D}_k^+\mathsf{vec}\,(A) = \mathsf{vech}\,(A)$

[4]Such approximation is harmless from a dynamic modeling perspective. Indeed, the scaling matrix in score-driven models is typically related to the inverse of the Fisher information matrix but is not necessarily equal to the latter. For instance, different powers of the inverse can be chosen. In this case, our approximating expression turns out to be very close to the full Information matrix and very similar results are obtained by explicitly computing the pseudo-inverse. However, the latter operation is computationally unstable and becomes infeasible in large dimensions.

This means that, in order to update the time-varying parameter f_t, one only needs to compute the scaled score s_t, which is a q-dimensional vector having the simple expression given by Eq. (4.20). Correlations are obtained by updating the time-varying parameter f_t using Eq. (4.16) and then constructing the matrix R_t through Eq. (4.15). Note that imposing a scalar structure on parameters A, B, namely $A = \alpha \mathbb{I}_q$, $B = \beta \mathbb{I}_q$ is now less restrictive, as variances are estimated independently and have different persistences. The correlation matrix R_t is positive-definite if and only if Q_t is positive-definite. If a scalar structure is imposed, one has:

$$\begin{aligned} \mathsf{vech}\left(Q_{t+1}\right) =& \omega + \beta \mathsf{vech}\left(Q_t\right) + \alpha \mathsf{vech}(H_t^{-1} X_t H_t^{-1}) - \alpha \mathsf{vech}\left(Q_t\right) && (4.21) \\ =& \omega + (\beta - \alpha)\, \mathsf{vech}\left(Q_t\right) + \alpha \mathsf{vech}(H_t^{-1} X_t H_t^{-1}), && (4.22) \end{aligned}$$

which is positive-definite for all t's if ω is positive-definite and if $\alpha \geqslant 0$, $\beta - \alpha \geqslant 0$. Note that the update rule in Eq. (4.22) coincides with the RecDCC specification of Bauwens et al. (2012) which adopt the Aielli (2013) correction in the context of a DCC model with realized covariance.

Matrix-F

The log-density function in Eq. (4.7) is given by:

$$\log p_{F_k}\left(X_t; V_t, \nu_1, \nu_2\right) = d(\nu_1, \nu_2) + \frac{\nu_1 - k - 1}{2} \log |X_t| - \frac{\nu_1}{2} \log |V_t| - \frac{\nu_1 + \nu_2}{2} \log |\tilde{W}_t| \tag{4.23}$$

where:

$$\tilde{W}_t = \mathbb{I}_k + \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} X_t \tag{4.24}$$

$$d(\nu_1, \nu_2) = \frac{\nu_1}{2} \log\left(\frac{\nu_1}{\nu_2 - k - 1}\right) + \log \Gamma_k\left(\frac{\nu_1 + \nu_2}{2}\right) - \log \Gamma_k\left(\frac{\nu_1}{2}\right) - \log \Gamma_k\left(\frac{\nu_2}{2}\right) \tag{4.25}$$

For more details on the Matrix-F distribution see Gupta and Nagar (1999). In Appendix B.6, we prove the following:

Proposition 4. *For the density in Eq. (4.23), the score* $\nabla_t^F = \frac{\partial \log p_{F_k}(X_t; f_t, \nu_1, \nu_2)}{\partial f_t}$ *is given by*

$$\nabla_t^F = \frac{\nu_1}{2} \mathcal{D}_k' \Psi_t' \left(H_t^{-1} Q_t^{-1} \otimes H_t^{-1} Q_t^{-1} \right) \left[\frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vec} \left(X_t \tilde{W}_t^{-1} \right) - \mathsf{vec} \left(V_t \right) \right] \tag{4.26}$$

As in Opschoor et al. (2017), to take into account the curvature of the log-density, we scale the score through the inverse of the Fisher information matrix of the Wishart density. This has two main advantages. First, it avoids the numerical computation of the Fisher information matrix of the matrix-F distribution, which is not available in closed form. Second, as in the Wishart density, we can compute in closed form the product $(\mathcal{I}_{t|t-1}^W)^{-1} \nabla_t^F$. Indeed, in Appendix B.7 we prove the following:

Proposition 5. *The scaled score vector* $s_t^F = (\mathcal{I}_{t|t-1}^W)^{-1} \nabla_t^F$ *is given by:*

$$s_t^F = \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vech} \left(H_t^{-1} X_t \tilde{W}_t^{-1} H_t^{-1} \right) - \mathsf{vech} \left(Q_t \right) \tag{4.27}$$

As in the Wishart density, the dynamic modeling of f_t solely requires the computation of the scaled score s_t, which is a q-dimensional vector given by Eq. (4.27). By imposing a scalar structure on parameters A, B, we have:

$$\mathsf{vech} \left(Q_{t+1} \right) = \omega + \beta \mathsf{vech} \left(Q_t \right) + \alpha s_t \tag{4.28}$$

$$= \omega + \beta \mathsf{vech} \left(Q_t \right) + \alpha \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vech} \left(H_t^{-1} X_t \tilde{W}_t^{-1} H_t^{-1} \right) - \alpha \mathsf{vech} \left(Q_t \right) \tag{4.29}$$

$$= \omega + (\beta - \alpha) \mathsf{vech} \left(Q_t \right) + \alpha \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vech} \left(H_t^{-1} X_t \tilde{W}_t^{-1} H_t^{-1} \right), \tag{4.30}$$

which is positive-definite for all t's if ω is positive-definite and if $\alpha \geqslant 0$, $\beta - \alpha \geqslant 0$. As in the previous case, the score-driven update mechanism includes by construction a correction based on the matrix H_t. Note that, with this choice for the scaling matrix, the scaled score of the matrix-F density converges to the scaled score of the Wishart density as $\nu_2 \to \infty$:

Proposition 6. *When $\nu_2 \to \infty$, we get $s_t^F = s_t^W$*

The proof is in Appendix B.8.

In principle, a HAR structure similar to that in Eq. (4.13) might be imposed even in the update rule for the correlations. As it will be extensively discussed in the empirical application in Section (4.4), a similar structure does not lead to any out-of-sample forecast gain. We thus maintain the current specification with a HAR structure in the variances and a simple AR(1) score-driven update for the correlations.

We conclude this section by noticing that both Gorgi et al. (2018) and Opschoor et al. (2017) consider a larger filtration also including daily returns. Under the assumption of conditional independence between realized covariances and daily returns, this translates into an additional term in the score of the two models. Our DCC-type approach is readily generalizable to include the effect of daily returns. However, as the latter has less explanatory power, and given that the aim of the present work is to highlight the advantages of the two-step approach, we neglect them for simplicity.

Estimation

The two-step procedure can be summarized as follows:

1. Maximum likelihood estimation of the k univariate models described in Section (4.2) (or alternative univariate specifications), from which

the matrices D_t are built

2. Maximum likelihood estimation of the Wishart log-density in Eq. (4.17) or the matrix-F log-density in Eq. (4.23) assuming D_t known and f_t evolving according to Eq. (4.16)

In a similar fashion to the DCC model, the static parameters A, B governing the dynamics of correlations are estimated under the scalar restrictions $A = \alpha \mathbb{I}_q$, $B = \beta \mathbb{I}_q$. In contrast, ω is a q-dimensional vector with different elements. We estimate it through (co)variance targeting, i.e. we set:

$$\omega = (\mathbb{I}_q - B)\,\mathsf{vech}(\overline{R}) \tag{4.31}$$

where $\overline{R}$ is the correlation matrix computed from the sample mean of realized covariances, namely $\overline{Q} = \frac{1}{T}\sum_{t=1}^{T} X_t$.

4.3 Monte-Carlo analysis

Finite sample properties

In this section, we study the finite sample properties of the two-step estimator via Monte-Carlo simulations. We consider both the two-step Wishart and the two-step matrix-F models. For ease of notation, we indicate the two models by "2-step-W" and "2-step-F". We set $D_t = \mathbb{I}_k$ for every t, i.e. we assume constant variances. Similar results are recovered when employing dynamic variances. We choose different values of k and T to test the properties of the estimator on a variety of different scenarios. Specifically, we set $k = \{2, 5, 10\}$ and $T = \{250, 500, 1000\}$. For each scenario, the model is simulated and estimated 1000 times. The two-step procedure is applied as described in Section (4.2). The initial parameter f_1 is set equal

to the correlation matrix computed from the sample mean of the simulated covariances.

We consider a specification for the Wishart model given by:

$$\alpha = 0.3, \quad \beta = 0.97, \quad \omega = (1-\beta)\mathsf{vech}\left[C(\rho)\right], \quad \nu = k + 10.$$

where $C(\rho)$ denotes the equi-correlation matrix with correlation parameter equal to ρ. We consider two different values for the correlation, namely $\rho = 0.5$ and $\rho = 0.9$. Figures (4.1), (4.2) show kernel density estimates of the probability density function of parameters in each scenario. The distribution concentrates around the true values as the time-series length T and/or the cross-section dimension k increase. This is due to the scalar specification adopted for the correlations. The distribution of α and ν is symmetric, whereas the distribution of β is slightly skewed at $T = 250$ and exhibits a bias. However, note that for $T = 500$ and $T = 1000$ all parameters are estimated accurately.

We perform the same analysis for the two-step model based on matrix-F distribution. Parameters are set as follows:

$$\alpha = 0.3, \quad \beta = 0.97, \quad \omega = (1-\beta)\mathsf{vech}\left[C(\rho)\right], \quad \nu_1 = 22, \quad \nu_2 = 35.$$

where $\rho = \{0.5, 0.9\}$. Figures (4.3), (4.4) show the results. We find again that parameters are accurately estimated in small samples and tend to concentrate around the true values as the time-series length T and/or the cross-section dimension k increase.

Analysis based on a misspecified DGP

We test the performance of the two-step procedure under a misspecified data generating process for volatilities and correlations. For simplicity, we

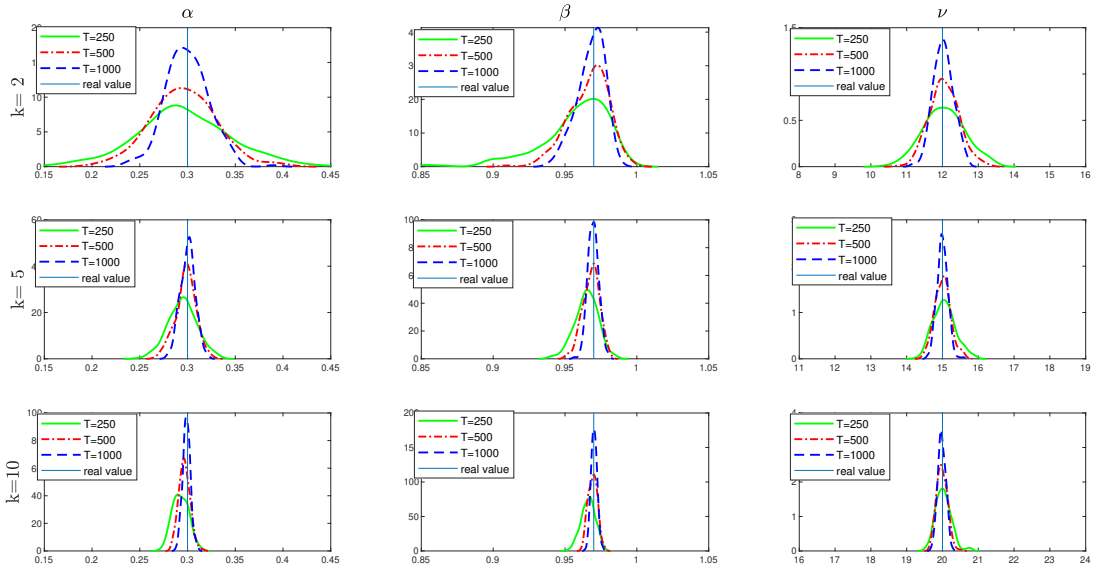

Figure 4.1: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step Wishart model with equi-correlation parameter $\rho = 0.5$.

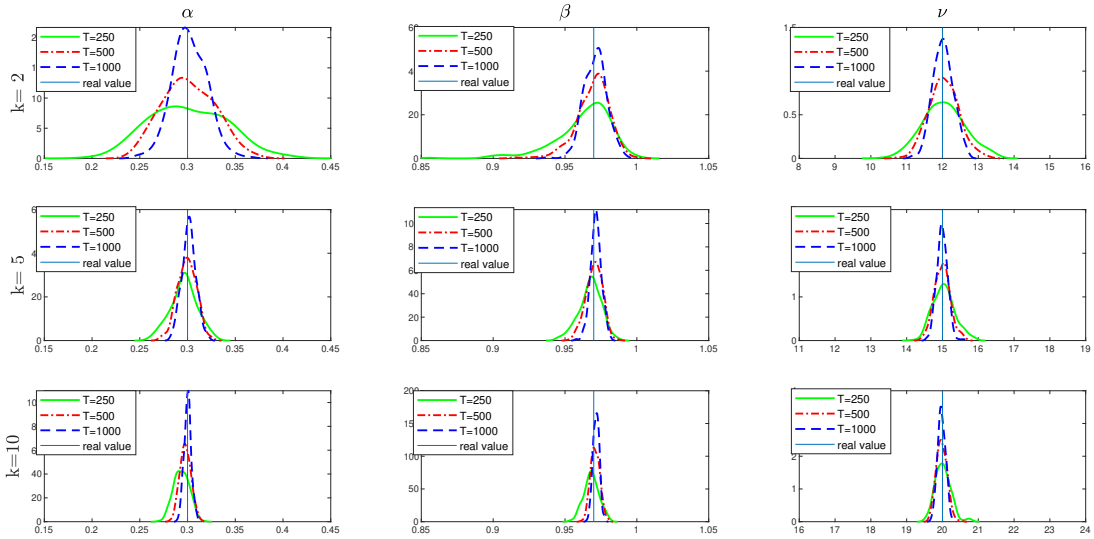

Figure 4.2: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step Wishart model with equi-correlation parameter $\rho = 0.9$.

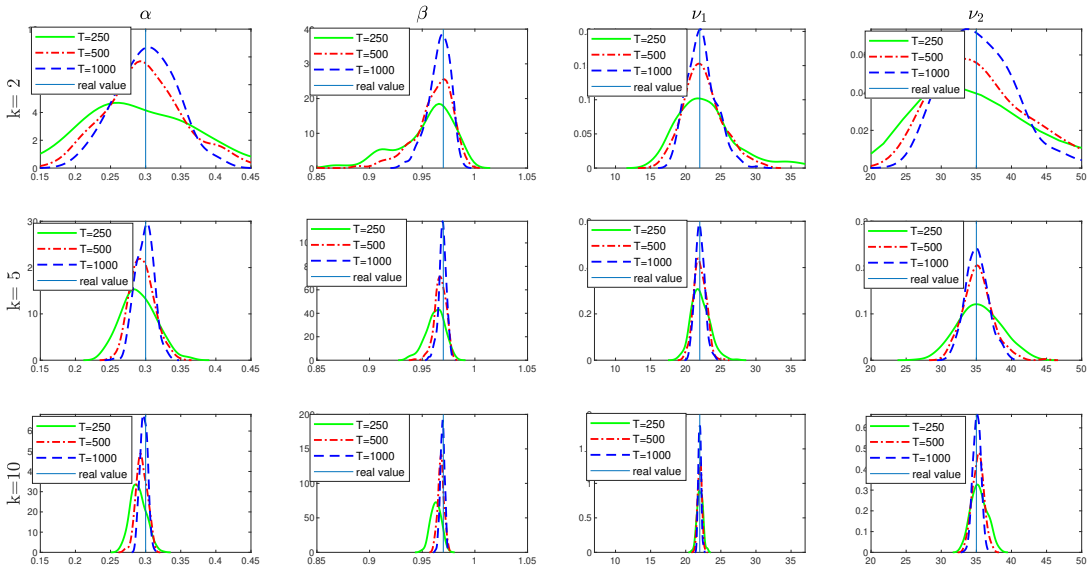

Figure 4.3: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step matrix-F model with equi-correlation parameter $\rho = 0.5$.

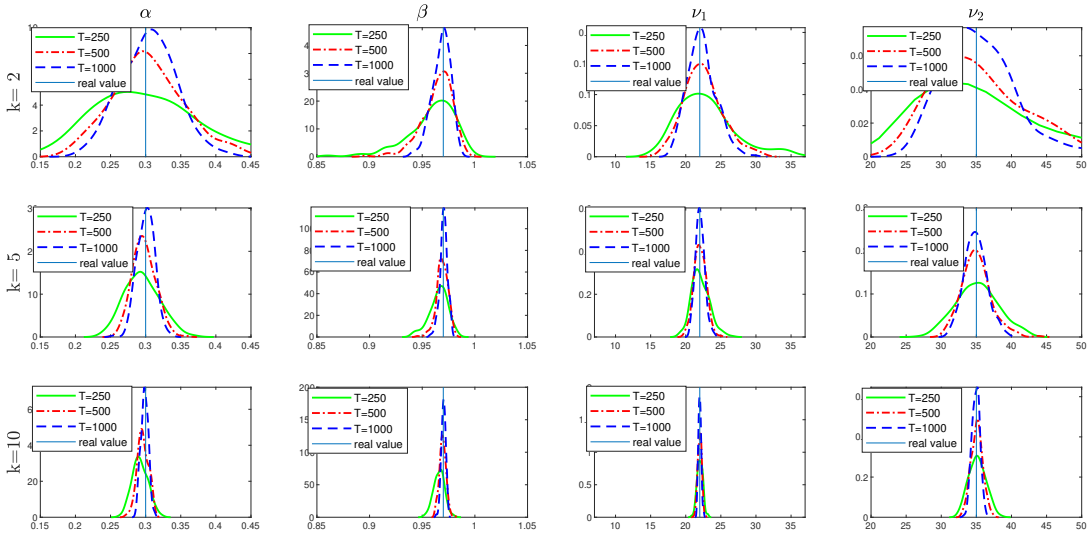

Figure 4.4: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step matrix-F model with equi-correlation parameter $\rho = 0.9$.

consider a bivariate model, in a similar fashion to Engle (2002a) and Creal et al. (2011). In Section (4.4), we show that the matrix-F distribution provides a better fit to empirical data. We, therefore, employ it as a conditional density to generate observations:

$$X_t|\mathcal{F}_{t-1} \sim F_2(V_t, \nu_1, \nu_2) \tag{4.32}$$

where we set $\nu_1 = 20$ and $\nu_2 = 8$. The latent covariance matrix V_t is written as:

$$V_t = \begin{pmatrix} c_1^2 & c_1 c_2 \rho_t \\ c_1 c_2 \rho_t & c_2^2 \end{pmatrix}$$

The misspecified volatilities c_1 and c_2 are modeled as:

$$c_{i,t} = \exp[0.97 \log(c_{i,t-1}) + \epsilon_t^{(i)}]$$

where $\epsilon_t^{(i)} \sim \mathcal{N}(0, 0.1)$, $i = 1, 2$. The correlation ρ_t evolves over time based on different dynamic patterns. They are given by:

$$\text{Sine:} \qquad \rho_t^{(1)} = 0.5 + 0.4 \sin(2\pi t/500) \tag{4.33}$$

$$\text{Fast sine:} \qquad \rho_t^{(2)} = 0.5 + 0.4 \sin(2\pi t/125) \tag{4.34}$$

$$\text{Step:} \qquad \rho_t^{(3)} = 0.9 - 0.4H\left(t - \frac{T}{4}\right) + 0.4H\left(t - \frac{T}{2}\right) - 0.4H\left(t - \frac{3T}{4}\right) \tag{4.35}$$

$$\text{Ramp:} \qquad \rho_t^{(4)} = \sum_{i=0}^{3} \left(\frac{4 \times 0.9}{T} t - 0.9i\right) \left[H\left(t - \frac{iT}{4}\right) - H\left(t - \frac{(i+1)\,T}{4}\right)\right] \tag{4.36}$$

where $T = 2000$ and H is the Heaviside step function. For each dynamic pattern, we generate $N = 1000$ time-series of realized covariances through the observation density in Eq. (4.32). We estimate both the 2-step-W and 2-step-F on the subsample comprising the first 1000 observations. In-sample

loss measures are computed in this subsample. We then use the recovered parameter estimates to filter the covariances in the subsample comprising the last 1000 observations. Figures (B.1)-(B.4), reported in Appendix B.9, show the simulated patterns and the average filtered correlations of 2-step-W and 2-step-F. Both models are able to capture the pattern of correlations accurately. Note that the confidence bands of 2-step-W filtered estimates are slightly larger than those of the 2-step-F model. The motivation is that the 2-step-W is not robust to the outliers generated by the matrix-F density.

Pattern	In-sample		Out-of-sample	
	2-step-W	2-step-F	2-step-W	2-step-F
Sine	1.000	0.736	1.000	0.718
Fast sine	1.000	0.846	1.000	0.807
Step	1.000	0.933	1.000	0.891
Ramp	1.000	0.937	1.000	0.937

Table 4.1: Relative in-sample and out-of-sample RMSE of two-step-W and two-step-F for the all the correlation patterns.

Table (4.1) reports relative in-sample and out-of-sample average Frobenius norms, defined as:

$$\text{Frob}\left(\hat{V}_t, V_t\right) = \sqrt{\mathsf{Tr}[(V_t - \hat{V}_t)(V_t - \hat{V}_t)']} \tag{4.37}$$

Hereafter, we refer to the Frobenius norm as RMSE, as it coincides with the square root of the sum of the MSE's of the entries of $\hat{V}_t$. The 2-step-F provides much lower in-sample and out-of-sample RMSE in patterns "sine" and "fast sine". In patterns "step" and "ramp", the 2-step-F performs better, though with a lower relative difference. This is due to the discontinuity in the correlation pattern, which is treated as an outlier by the 2-step-F

model. As a consequence, it adapts more slowly to the abrupt change of correlation.

Comparison with joint estimation

We compare now the performance of the two-step procedure to that of standard joint estimation. In order to generate variance paths that have the same dynamic behavior of real data, we construct the matrices D_t based on a dataset of 100 realized variance time-series computed from stocks traded on the NYSE. The dataset is described in more detail in Section (4.4). In each simulation, the matrices D_t are generated by randomly selecting k realized variance series from the dataset and replacing each diagonal element[5] $D_t^{(i)}$ with the true realized volatility $(X_t^{(i)})^{1/2}$:

$$D_t = \begin{pmatrix} (X_t^{(1)})^{1/2} & & 0 \\ & \ddots & \\ 0 & & (X_t^{(k)})^{1/2} \end{pmatrix}$$

To simulate the correlation matrix R_t, we use the equicorrelation model of Engle and Kelly (2012). It has the form:

$$R_t = \rho_t I_k + [1 - \rho_t] J_{k,k} \tag{4.38}$$

where $J_{k,k} \in \mathbb{R}^{k \times k}$ is a matrix of ones. The correlation parameter ρ_t evolves based on the following process:

$$\rho_t = 1 - \frac{1}{k-1} [1 - \tanh(\theta_t)] \tag{4.39}$$

$$\theta_{t+1} = \phi\theta_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}\left(0, \sigma^2\right) \tag{4.40}$$

[5] As D_t is the matrix of latent standard deviations, we apply an EWMA smoothing scheme to clean the real time-series from measurement error effects.

where we set $\sigma^2 = 0.25$. We then compute the covariance matrix as $V_t = D_t R_t D_t$ and generate realized covariances through both the Wishart and the matrix-F density:

$$X_t^W \sim W_k\left(V_t/\nu, \nu\right) \tag{4.41}$$

$$X_t^F \sim F_k\left(V_t, \nu_1, \nu_2\right) \tag{4.42}$$

with $\nu = k + 10$, $\nu_1 = 140$ and $\nu_2 = 120$. These values are similar to those estimated on real data.

We set $k = 10$ and perform $N = 500$ simulations of realized covariance series $\{X_t^W\}$, $\{X_t^F\}$ of $T = 2000$ observations. As in the previous simulation study, in-sample loss measures are computed in the subsample comprising the first 1000 observations. Out-of-sample loss-measures are instead computed in the remaining subsample of 1000 observations based on the parameters estimated in the first subsample. As loss measures, we consider the RMSE and the Quasi-likelihood (Qlike). The latter is defined as:

$$\text{Qlike}\left(\hat{V}_t, V_t\right) = -\log|\hat{V}_t^{-1}V_t| + \mathsf{Tr}\left(\hat{V}_t^{-1}V_t\right) - k \tag{4.43}$$

Table (4.2) compares in-sample and out-of-sample loss measures of two-step and joint estimation for both Wishart and matrix-F models. The joint estimation approach coincides with the models of Gorgi et al. (2018) and Opschoor et al. (2017). In both in-sample and out-of-sample, the two-step estimator performs significantly better than the joint estimator, according to both loss measures. To test the significance of our results, we perform the Diebold-Mariano test (Diebold and Mariano 2002) at 5% significance level for each simulation. We report in the table the number of times each loss measure is judged as significantly smaller by the Diebold-Mariano test. It clearly emerges that, as a result of the higher flexibility provided by the

Loss	Model	Wishart		matrix-F	
		In-sample	Out-of-sample	In-sample	Out-of-sample
RMSE	two-step	1.000	1.000	1.000	1.000
		(420)	(395)	(475)	(500)
	joint	1.118	1.092	1.154	1.189
		(15)	(20)	(0)	(0)
Qlike	two-step	1.000	1.000	1.000	1.000
		(460)	(430)	(400)	(400)
	joint	1.096	1.086	1.213	1.141
		(15)	(20)	(0)	(0)

Table 4.2: Relative in-sample and out-of-sample RMSE and Qlike of two-step models compared with joint estimation models. We report in parenthesis the number of times the corresponding loss measure is judged as significantly smaller by the Diebold-Mariano test.

two-step models, the latter provide significantly smaller loss measures in most of the $N = 500$ simulations.

4.4 Empirical evidence

Dataset

The empirical analysis is performed on two different datasets. The first dataset consists of transactions of 100 assets traded on the NYSE. The time resolution is 1-second. Data are available from 03-01-2006 to 31-12-2014, corresponding to 2265 business days. The second dataset consists of transactions of 2767 assets belonging to the Russell 3000 index. The time resolution is 1-minute and data are available from 03-01-2006 to 27-09-2013, corresponding to 1948 business days. As a realized covariance estimator, we use the multivariate Realized Kernel of Barndorff-Nielsen et al. (2011).

Before computing it, we perform the standard cleaning procedures described by Barndorff-Nielsen et al. (2009).

Preliminary analysis

The main advantage of the proposed models is the higher flexibility in the estimation of the static parameters. Gorgi et al. (2018) and Opschoor et al. (2017) use the scalar restrictions $A = \alpha \mathbb{I}_q$ and $B = \beta \mathbb{I}_q$, i.e. they assume the same persistence for the volatilities and correlations of all the assets. In contrast, we find empirical evidence supporting the hypothesis of a high level of heterogeneity among the persistences of volatilities. Such a hypothesis is tested by means of two different experiments.

The first experiment is performed using the dataset of 2767 assets of the Russell 3000 index. The results, presented in Figures (4.5), (4.6), (4.7), provide a first evidence that the persistences of volatilities can vary a lot among assets. The blue lines represent the (sorted) 2767 maximum likelihood estimates of the coefficients $\beta^{(d)}$, $\beta^{(w)}$ and $\beta^{(m)}$ appearing in Eq. (4.13). The red lines represent 2767 (sorted) realizations of $\beta^{(d)}$, $\beta^{(w)}$ and $\beta^{(m)}$ given the null hypothesis that the different assets have the same HAR coefficients. They are built using the following method: we simulate 2767×1000 random variables distributed as $\mathcal{N}\left(\bar{\beta}^{(j)}, \bar{\sigma^2}^{(j)}\right)$, $j = d, w, m$, where $\bar{\beta}^{(j)}$ represents the sample mean of the estimated $\beta^{(j)}$ coefficient across the 2767 assets, and $\bar{\sigma^2}^{(j)}$ represents the sample mean of the estimation error variances of the $\beta^{(j)}$ coefficient (computed as the inverse of the hessian at the maximum). We compute the blue line by averaging the sorted 2767 realizations over the 1000 replications. The 95% confidence bands are computed as empirical quantiles of the simulated distribution. They show that the real estimated coefficients are incompatible with the null assumption that

volatility persistences are equal.

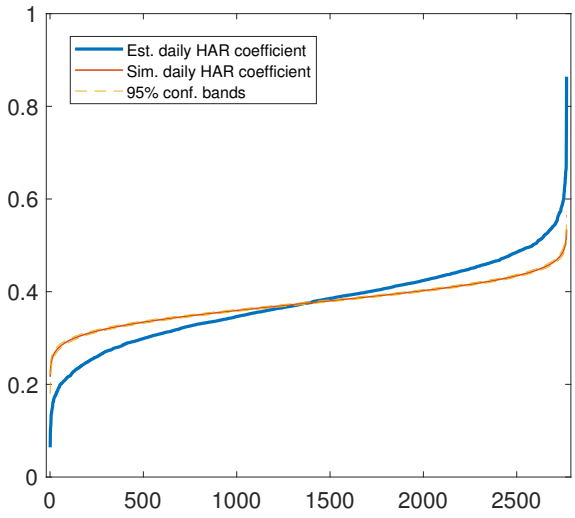

Figure 4.5: We report in blue the sorted maximum likelihood estimates of the $\beta^{(d)}$ coefficient corresponding to the 2767 assets in the Russell 3000 dataset. The red line and the 95% confidence bands are computed under the null hypothesis that the 2767 assets have the same $\beta^{(d)}$ coefficient.

In the second experiment, we formally test on real data the scalar restriction assumed by Gorgi et al. (2018). We consider 100 portfolios each composed by k randomly selected assets from the NYSE dataset. We estimate the Wishart model of Gorgi et al. (2018) by imposing two different restrictions on the matrix B. In the first case, we assume a scalar restriction, whereas in the second case we assume a diagonal restriction (i.e. the diagonal elements are not required to be equal among each other). In both cases, the matrix A is assumed to have a scalar structure. Under the diagonal restriction, the number of parameters scales as k^2. We thus set $k = 5$ in order to avoid the curse of dimensionality.

We compare the fitting ability of the two different restrictions using the

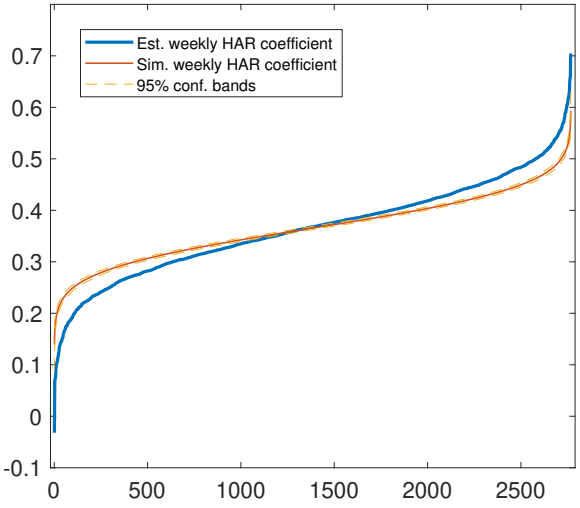

Figure 4.6: We report in blue the sorted maximum likelihood estimates of the $\beta^{(w)}$ coefficient corresponding to the 2767 assets in the Russell 3000 dataset. The red line and the 95% confidence bands are computed under the null hypothesis that the 2767 assets have the same $\beta^{(w)}$ coefficient.

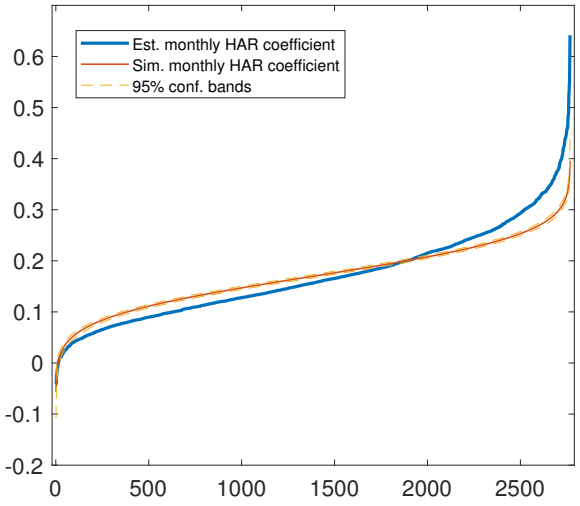

Figure 4.7: We report in blue the sorted maximum likelihood estimates of the $\beta^{(m)}$ coefficient corresponding to the 2767 assets in the Russell 3000 dataset. The red line and the 95% confidence bands are computed under the null hypothesis that the 2767 assets have the same $\beta^{(m)}$ coefficient.

likelihood ratio test. In particular, we test the null hypothesis

$$H_0 : B = \beta I \tag{4.44}$$

against the alternative hypothesis

$$H_1 : B = \begin{pmatrix} \beta_1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \beta_{15} \end{pmatrix}, \quad \text{where } \beta_1, \dots, \beta_{15} \text{ are not all equal.} \tag{4.45}$$

To compute the p-values, we use a χ^2 distribution for the test statistics, since the model with a scalar restriction is nested into the model with a diagonal restriction. The test rejects the null hypothesis with p-values smaller than 0.01 for all the 100 portfolios. We obtain the same results when testing the model based on matrix-F density.

The results of these two tests suggest that the scalar assumption of Gorgi et al. (2018) and Opschoor et al. (2017) might be too restrictive and warrant for a more flexible specification. In the next sections, we show the advantages of the proposed two-step estimation procedure over standard joint estimation.

Analysis settings and benchmark models

We perform the analysis at different cross-section dimensions and with different portfolios. In the NYSE dataset, which contains 1-second data, we set $k = 5, 10, 25, 50, 100$. The assets belonging to each group are randomly selected, except for the group of $k = 100$ stocks, which includes all the assets in the dataset. In the Russell 3000 dataset, the time resolution is 1-minute and the number of time-stamps per day is much smaller compared to the NYSE dataset. To avoid ill-conditioned realized covariance estimates, we

limit the maximum cross-section to 50 assets, i.e. we set $k = 5, 10, 25, 50$. Even in this case, the assets in each portfolio are randomly selected.

In the in-sample analysis, we estimate the models in the entire NYSE (Russell 3000) dataset of 2265 (1948) business days. In the out-of-sample analysis, the models are estimated in the subsample comprising the first 1000 business days. The estimated parameters are then used to forecast the covariances of the last 1265 (948) observations, from 22-12-2009 to 31-12-2014 (27-09-2013). As a loss measure, we follow Patton (2011) and use the RMSE and the Qlike. We assess whether loss differences are significant through the Model Confidence Set (MCS) of Hansen et al. (2011).

The 2-step-W and 2-step-F are compared with the following benchmark models:

1. RWG: the Realized Wishart GARCH model of Gorgi et al. (2018) based on the Wishart density in Eq. (4.5)

2. RWG-HAR: the RWG model equipped with a HAR-like structure in the dynamic equation for the covariances:

$$f_{t+1} = \omega + B_1 f_t + B_2 f_{t-1|t-5} + B_3 f_{t-6|t-22} + A s_t$$

   where $f_t = \mathsf{vech}(V_t)$.

3. GAS-F: the model of Opschoor et al. (2017) based on the matrix-F density in Eq. (4.7)

4. GAS-F-HAR: the GAS-F model is equipped with a HAR-like structure on static parameters similar to that of the RWG-HAR

5. HAR-DRD: the model of Oh and Patton (2018), combining univariate HAR models with a panel HAR specification for the correlations.

Similarly to our approach, the estimation is made in two steps. In the first step, the volatilities are estimated using k distinct HAR models. In the second step, a multivariate HAR-like specification is fitted on the time-series of realized correlations. Covariance forecasts are then computed by combining volatility and correlation forecasts through the DRD decomposition in Eq. (4.9).

6. DCC: the standard Dynamic Conditional Correlation model Engle (2002a) for daily returns estimated through the two-step procedure.

We include the RWG-HAR and the GAS-F-HAR in the set of benchmark models in order to exclude that forecast gains provided by our two-step procedure are merely due to long memory effects captured by the univariate models, which have a similar HAR-like structure, see Eq. (4.13). In principle, we might impose the same structure in our correlation model. We verified that such an assumption does not provide significant forecast gains. We thus maintain a parsimonious specification with a HAR-like structure in the univariate models and a standard AR(1) score-driven specification for the correlations.

All the realized covariance models are estimated by assuming a scalar specification for the matrix parameters appearing in the score-driven update equation. These include the two matrices A, B in the RWG and GAS-F models, which are constrained as $A = \alpha \mathbb{I}_q$, $B = \beta \mathbb{I}_q$, and the matrices B_1, B_2, B_3 in the RWG-HAR and GAS-F-HAR, which are constrained as $B_1 = \beta_1 \mathbb{I}_q$, $B_2 = \beta_2 \mathbb{I}_q$, $B_3 = \beta_3 \mathbb{I}_q$.

In-Sample results

Table (4.3) reports the maximum likelihood estimates, obtained in the NYSE dataset, of the parameters of the models described in the previous section. They are estimated in the entire dataset of 2265 business days and for each cross-section dimension $k = 5, 10, 25, 50, 100$. Note that the parameter ν in the RWG model and the parameters ν_1, ν_2 in the GAS-F model is significantly different from the corresponding estimates in the 2-step-W and 2-step-F models. In particular, ν is lower in the RWG model, ν_1 is generally lower in the GAS-F model and ν_2 is lower in the 2-step-F model for $k = 5, 10, 25$ and larger for $k = 50, 100$. Such differences indicate that the fat-tail behavior of covariances is significantly different from that of correlations, and it is a further motivation for the introduction of the proposed two-step approach.

Table (4.4) shows average in-sample loss measures computed in both datasets. We note that score-driven models based on the two-step procedure (2-step-W and 2-step-F) fit the data better than those based on joint estimation (RWG, RWG-HAR, GAS-F, GAS-F-HAR). The relative performance between the two classes of models tends to increase with the cross-section dimension, as can be seen from the RMSE and Qlike gains at $k = 50, 100$, which are both large and highly significant. This result is due to the fact that large portfolios are more likely to include assets with different volatility persistences, which are better captured by the two-step models. We also see that the 2-step-F often provides lower loss measures compared to the 2-step-W, as a result of being robust to the outliers that are typically observed on real-time series. In the vast majority of the cases, the 2-step-F has significantly lower loss measures, whereas in a few cases the performances of the two models are statistically indistinguishable. The

RWG-HAR and GAS-F-HAR tend to perform better than the RWG and GAS-F because of the long-memory captured by the HAR specification. We finally note that the HAR-DRD often provides a lower RMSE. This is natural when looking at in-sample results, as the HAR-DRD is estimated based on the same RMSE criterion.

We report in Table (4.5) the AIC and BIC statistics computed in the NYSE dataset. As can be seen, the 2-step-F has the lowest AIC and BIC for all the cross-section dimensions. We also note that the 2-step-W has lower AIC and BIC compared to both the RWG and the RWG-HAR. In order to further examine the in-sample performance of the proposed methodology, we carry out the likelihood based model selection test of Rivers and Vuong (2002), which is applicable to non-nested, possibly misspecified nonlinear dynamic models. In particular, we test 2-step and joint estimation based models (2-step-W vs RWG-HAR and 2-step-F vs GAS-F-HAR) on the one hand, and matrix-F and Wishart based models (2-step-F vs 2-step-W and GAS-F-HAR vs RWG-HAR) on the other hand. The first test compares the likelihood of 2-step models to that of joint estimation based models, whereas the second test compares the likelihood of matrix-F models to that of Wishart models. Table (4.6) shows the test statistics of Rivers and Vuong (2002), computed as the difference between the log-likelihoods of the two models indicated in the first line, divided by the square root of its asymptotic variance. The latter is computed through the Newey-West estimator. Under the null that the likelihoods of the two models are equal, the test statistics are asymptotically distributed as a standard normal. We see that the null is strongly rejected in all the cases, implying that 2-step models outperform joint estimation based models and matrix-F models outperform Wishart models.

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
				k = 5					**k = 10**						**k = 25**						**k = 50**			
2-step-W	0.293	0.980			33.2		0.221	0.999			57.7		0.119	0.990			98.8		0.064	0.992			133.3	
	(0.012)	(0.000)			(0.242)		(0.004)	(0.000)			(0.218)		(0.001)	(0.000)			(0.149)		(0.000)	(0.000)			(0.096)	
2-step-W(1)	0.542	0.805	0.130	0.047	50.5		0.528	0.799	0.141	0.040	51.9		0.480	0.781	0.161	0.035	47.1		0.466	0.769	0.170	0.039	46.0	
	(0.020)	(0.029)	(0.033)	(0.015)	(1.376)		(0.020)	(0.029)	(0.033)	(0.014)	(1.393)		(0.017)	(0.029)	(0.033)	(0.015)	(1.152)		(0.020)	(0.032)	(0.036)	(0.016)	(1.177)	
2-step-F	0.900	0.986			151.1	55.3	0.832	0.999			245.5	97.7	0.282	0.999			190.4	152.9	0.056	0.999			162.7	180.5
	(0.060)	(0.003)			(9.018)	(0.754)	(0.023)	(0.001)			(6.964)	(0.707)	(0.004)	(0.000)			(2.890)	(0.565)	(0.001)	(0.000)			(0.381)	(0.692)
2-step-F(1)	2.247	0.820	0.117	0.049	201.6	78.6	2.040	0.814	0.127	0.043	195.7	82.8	2.422	0.807	0.134	0.041	229.1	68.1	2.429	0.789	0.152	0.042	234.8	68.2
	(0.087)	(0.024)	(0.031)	(0.016)	(7.841)	(1.750)	(0.084)	(0.028)	(0.032)	(0.016)	(2.774)	(1.609)	(0.087)	(0.030)	(0.035)	(0.016)	(2.576)	(1.413)	(0.085)	(0.030)	(0.035)	(0.016)	(3.295)	(1.573)
RWG	0.568	0.992			33.1		0.481	0.992			56.8		0.346	0.996			95.2		0.242	0.995			127.5	
	(0.007)	(0.000)			(0.253)		(0.003)	(0.000)			(0.222)		(0.001)	(0.000)			(0.139)		(0.000)	(0.000)			(0.088)	
GAS-F	0.905	0.995			75.4	72.3	0.978	0.995			144.9	117.4	0.824	0.999			161.2	182.0	0.479	0.999			186.3	157.0
	(0.015)	(0.012)			(1.438)	(1.326)	(0.008)	(0.008)			(1.532)	(1.033)	(0.000)	(0.000)			(0.000)	(0.000)	(0.001)	(0.000)			(0.351)	(0.653)
RWG-HAR	0.615	0.832	0.111	0.050	33.8		0.546	0.813	0.131	0.049	58.0		0.432	0.785	0.148	0.063	97.2		0.320	0.777	0.138	0.080	129.4	
	(0.007)	(0.007)	(0.007)	(0.004)	(0.272)		(0.004)	(0.005)	(0.005)	(0.002)	(0.210)		(0.001)	(0.002)	(0.002)	(0.001)	(0.140)		(0.001)	(0.001)	(0.002)	(0.001)	(0.093)	
GAS-F-HAR	0.987	0.870	0.080	0.046	72.7	77.4	0.999	0.860	0.088	0.047	139.6	125.1	0.9169	0.8646	0.0807	0.0534	284.9	182.1	0.3812	0.9500	0.0000	0.0485	205.9	237.7
	(0.016)	(0.007)	(0.008)	(0.004)	(1.484)	(1.357)	(0.009)	(0.004)	(0.005)	(0.002)	(1.523)	(1.100)	(0.003)	(0.002)	(0.002)	(0.001)	(1.027)	(0.921)	(0.001)	(0.001)	(0.002)	(0.001)	(0.350)	(0.699)
DCC	0.016	0.962					0.011	0.970					0.005	0.977					0.002	0.978				
	(0.002)	(0.005)					(0.001)	(0.004)					(0.000)	(0.007)					(0.000)	(0.001)				
HAR-DRD		0.399	0.437	0.107				0.563	0.328	0.070				0.357	0.406	0.154				0.339	0.398	0.176		
		(0.063)	(0.087)	(0.075)				(0.028)	(0.039)	(0.036)				(0.010)	(0.014)	(0.012)				(0.005)	(0.007)	(0.006)		
HAR-DRD(1)		0.609	0.268	0.086				0.563	0.328	0.070				0.531	0.328	0.097				0.512	0.343	0.103		
		(0.019)	(0.025)	(0.019)				(0.020)	(0.025)	(0.020)				(0.020)	(0.026)	(0.021)				(0.020)	(0.027)	(0.022)		

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
				k = 100		
2-step-W	0.031	0.993			147.4	
	(0.000)	(0.000)			(0.044)	
2-step-W(1)	0.449	0.765	0.168	0.045	42.8	
	(0.004)	(0.037)	(0.033)	(0.019)	(0.459)	
2-step-F	0.059	0.993			193.1	507.4
	(0.000)	(0.000)			(0.091)	(0.635)
2-step-F(1)	2.312	0.790	0.149	0.045	219.1	66.0
	(0.078)	(0.031)	(0.036)	(0.017)	(2.635)	(1.132)
RWG	0.147	0.994			142.6	
	(0.000)	(0.000)			(0.042)	
GAS-F	0.256	0.992			140.7	202.5
	(0.000)	(0.000)			(0.000)	(0.001)
RWG-HAR	0.217	0.763	0.135	0.096	143.7	
	(0.000)	(0.001)	(0.001)	(0.000)	(0.042)	
GAS-F-HAR	0.3095	0.9062	0.0860	0.0000	107.9	222.0
	(0.000)	(0.001)	(0.001)	(0.000)	(0.094)	(0.538)
DCC	0.001	0.959				
	(0.000)	(0.006)				
HAR-DRD		0.300	0.392	0.206		
		(0.002)	(0.003)	(0.003)		
HAR-DRD(1)		0.508	0.337	0.113		
		(0.020)	(0.027)	(0.022)		

Table 4.3: Maximum likelihood estimates of the parameters of 2-step-W, 2-step-F and of benchmark models obtained in the NYSE dataset. Standard errors are reported in parenthesis. For the k univariate models, denoted by 2-step-W(1) and 2-step-F(1), we report the average of the maximum likelihood estimates and of the standard errors. Similarly, we report the averages of the parameters of the univariate models in the HAR-DRD.

Model	MSE					Qlike				
	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$
NYSE										
2-step-W	0.9859*	1.0143	1.0255	1.0308	1.0478	**0.9995***	0.9964	**0.9869***	**0.9962***	1.0051
2-step-F	1.0000	1.0000*	1.0000	1.0000	1.0000*	1.0000*	1.0000	1.0000	1.0000*	**1.0000***
RWG	1.0320	1.0285	1.0521	1.0494	1.0818	1.0476	1.0470	1.0458	1.0575	1.0548
RWG-HAR	1.0184	1.0240	1.0457	1.0407	1.0913	1.0203	1.0126	1.0137	1.0320	1.0356
GAS-F	1.0352	1.0254	1.0396	1.1230	1.3608	1.0522	1.0417	1.0393	1.1114	1.1925
GAS-F-HAR	1.0151	1.0169	1.0162	1.0620	1.3223	1.0287	1.0135	1.0189	1.0530	1.1504
DCC	1.8458	1.7460	1.5017	1.4146	1.3819	1.5974	1.6370	1.7858	1.6352	1.4153
HAR-DRD	**0.9858***	**0.9975***	**0.9946***	**0.9812***	**0.9957***	1.0024*	**0.9898***	0.9922	1.0250	1.0562
Russell										
2-step-W	1.0133	1.0208	1.0256	1.0469	-	1.0029*	1.0145	1.0081	1.0055	-
2-step-F	**1.0000***	**1.0000***	1.0000*	1.0000*	-	1.0000*	**1.0000***	**1.0000***	**1.0000***	-
RWG	1.0683	1.0573	1.0735	1.0941	-	1.0322	1.0490	1.0524	1.0583	-
RWG-HAR	1.0525	1.0528	1.0587	1.0878	-	1.0100	1.0190	1.0276	1.0347	-
GAS-F	1.0408	1.0201	1.0210	1.0730	-	1.0329	1.0346	1.0340	1.0744	-
GAS-F-HAR	1.0281	1.0116	1.0131	1.0633	-	1.0073*	1.0144	1.0278	1.0597	-
DCC	1.4863	1.5754	1.5065	1.8216	-	1.5392	1.6012	1.5702	1.4774	-
HAR-DRD	1.0035*	1.0140	**0.9988***	**0.9907***	-	**0.9976***	1.0107	1.0196	1.0365	-

Table 4.4: Relative in-sample average loss measures. All measures are reported with respect to the 2-step-F model. A value lower than one indicates that the corresponding model outperforms the 2-step-F model. Bold numbers denote the models having lowest loss measures. Asterisks denote the models that are part of the 90% model confidence set (MCS).

Model	AIC				
	$k = 5$	$k = 10$	$k = 25$	$k = 50$	$k = 100$
	$(\times 10^5)$	$(\times 10^6)$	$(\times 10^7)$	$(\times 10^7)$	$(\times 10^8)$
2-step-W	−6.8400	−2.4546	−1.4677	−5.7651	−2.3129
2-step-F	**−7.1643**	**−2.6124**	**−1.5617**	**−6.0493**	**−2.4197**
RWG	−5.8542	−2.2686	−1.4190	−5.6213	−2.2535
GAS-F	−6.1604	−2.4064	−1.5091	−5.9459	−2.3749
RWG-HAR	−5.7970	−2.2475	−1.4059	−5.5688	−2.2318
GAS-F-HAR	−6.0981	−2.3833	−0.0028	−5.8736	−2.3432
Model	BIC				
	$k = 5$	$k = 10$	$k = 25$	$k = 50$	$k = 100$
	$(\times 10^5)$	$(\times 10^6)$	$(\times 10^7)$	$(\times 10^7)$	$(\times 10^8)$
2-step-W	−6.8381	−2.4542	−1.4676	−5.7649	−2.3129
2-step-F	**−7.1621**	**−2.6119**	**−1.5616**	**−6.0491**	**−2.4196**
RWG	−5.8541	−2.2686	−1.4190	−5.6213	−2.2535
GAS-F	−6.1602	−2.4064	−1.5091	−5.9459	−2.3749
RWG-HAR	−5.7967	−2.2475	−1.4059	−5.5688	−2.2318
GAS-F-HAR	−6.0978	−2.3832	−0.0028	−5.8736	−2.3432

Table 4.5: AIC and BIC statistics of all the considered models computed in the NYSE dataset.

	2-step vs joint		Matrix-F vs Wishart	
Cross-section	2-step-W vs RWG-HAR	2-step-F vs GAS-F-HAR	2-step-F vs 2-step-W	GAS-F-HAR vs RWG-HAR
$k = 5$	69.79	85.25	23.14	31.64
$k = 10$	67.18	94.15	28.24	30.73
$k = 25$	49.34	98.24	28.27	33.29
$k = 50$	37.49	57.35	39.69	43.84
$k = 100$	16.31	16.56	42.51	47.88

Table 4.6: Test statistics of the Rivers and Vuong (2002) test. In each column, we report the results of the test for the two models indicated in the first line. The test statistics are computed as the difference between the log-likelihoods of the two models divided by its asymptotic standard deviation. The latter is estimated through the Newey-West estimator.

Out-of-Sample results

We now check whether the in-sample gains found in the previous analysis translate into out-of-sample forecast gains. In Table (4.7), we report the average out-of-sample loss measures of all the models considered above. Forecasts are computed in the NYSE (Russell 3000) subsample including the last 1265 (948) business days and are based on parameter estimates obtained in the subsample comprising the first 1000 business days. We immediately notice that, with only a few exceptions, the 2-step-F model is the best performing model and it is included in the model confidence set. Similarly to the in-sample analysis, the relative performance of the 2-step-W and 2-step-F over joint estimation based models increases as the cross-section dimension increases. This confirms the intuition that the flexibility provided by the two-step procedure leads to better covariance forecasts. The results in the table also confirm that the matrix-F density is more suited than the Wishart density to capture the fat-tails of covariances and correlations.

The performance of the HAR-DRD is comparable to that of the 2-step-F model when we look at the RMSE loss. When we instead consider the Qlike loss, the relative performance of the HAR-DRD rapidly deteriorates when k increases. The comparison of our methodology with the HAR-DRD is examined further in the next section, where we study the behavior of both models in relation to measurement errors on covariance estimates.

One may argue that additional forecast gains might be recovered by imposing a HAR structure on the score-driven dynamic equation for the correlations in the 2-step-W and 2-step-F models. As outlined in Section (4.4), this is not the case, as we verified by performing similar experiments. Most of the benefits provided by modeling long-memory come from the uni-

variate HAR models, whereas a similar HAR structure on correlations does not lead to significant out-of-sample improvements. This is confirmed by the fact that the RWG-HAR and GAS-F-HAR, which have a HAR structure in both volatilities and correlations, have a performance which in some cases is sub-optimal with respect to the RWG and the GAS-F. In this sense, our proposed two-step approach provides a further advantage, as it easily allows for different modeling strategies for volatilities and correlations.

Robustness to noise

As underlined in the introduction, one of the advantages of the proposed approach is that it allows taking into account the unavoidable estimation error that affects realized covariance estimates. The latter is indeed modeled as noisy observations of a latent matrix process representing the true integrated covariance of the intraday log-prices. In this sense, our methodology is in sharp contrast with the HAR-DRD, where realized covariances are modeled as if they were the true covariances. We refer the reader to Bollerslev et al. (2016), Bollerslev et al. (2018) and Buccheri and Corsi (2019) for a more detailed discussion on the impact of estimation errors on forecasting with realized measures.

In this section, we assess the effect of different levels of noise on covariance forecasts by comparing our methodology, which accounts for noise, to the HAR-DRD, which does not possess such ability. To increase the level of noise, we sample the log-prices at the frequency of 5-minutes. The number of sampled log-prices at this frequency is considerably smaller than that at 1-second (NYSE dataset) and 1-minute (Russell 3000 dataset). As a consequence, the multivariate realized kernel estimator of Barndorff-Nielsen et al.

Model	MSE					Qlike				
	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$
NYSE										
2-step-W	0.9882*	1.0087*	1.0413	1.0814	1.0927	**0.9987***	1.0044*	1.0107	1.0141	1.0121
2-step-F	1.0000	**1.0000***	**1.0000***	**1.0000***	**1.0000***	1.0000*	**1.0000***	**1.0000***	**1.0000***	**1.0000***
RWG	1.0248	1.0278	1.0479	1.0523	1.0531	1.0261	1.0293	1.0659	1.0740	1.0588
RWG-HAR	1.0422	1.0471	1.0530	1.0602	1.0683	1.0507	1.0279	1.0555	1.0767	1.0708
GAS-F	1.0209	1.0122	1.0312	1.0778	1.0444	1.0248	1.0130	1.0731	1.0644	1.1180
GAS-F-HAR	1.0294	1.1061	1.0232	1.1489	1.1649	1.0360	1.0768	1.0467	1.1099	1.1275
DCC	2.0354	1.8998	1.6256	1.4739	1.3265	5.2720	3.5086	1.9520	1.4058	1.3014
HAR-DRD	**0.9840***	1.0007*	1.0051*	1.0091	1.0105	1.0091*	1.0196	1.0271	1.0433	1.0591
Russell										
2-step-W	1.0341	1.0315	1.0483	1.0284	-	1.0246	1.0266	1.0141	1.0107	-
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	-	**1.0000***	**1.0000***	**1.0000***	**1.0000***	-
RWG	1.0538	1.0542	1.0711	1.0435	-	1.0434	1.0586	1.0612	1.0554	-
RWG-HAR	1.0605	1.0513	1.0686	1.0561	-	1.1015	1.0751	1.0697	1.0699	-
GAS-F	1.0199	1.0172	1.0148	1.0496	-	1.0206	1.0169	1.0311	1.1345	-
GAS-F-HAR	1.0276	1.0190	1.0284	1.0231	-	1.0961	1.0554	1.0605	1.0877	-
DCC	1.1081	1.2586	1.2739	1.2929	-	1.2654	1.5282	1.4649	1.4237	-
HAR-DRD	1.0084	1.0099*	1.0083	1.0027*	-	1.0020*	1.0183	1.0263	1.0277	-

Table 4.7: Relative out-of-sample average loss measures. All measures are reported with respect to the 2-step-F model. A value lower than one indicates that the corresponding model outperforms the 2-step-F model. Bold numbers denote the models having lowest loss measures. Asterisks denote the models that are part of the 90% model confidence set (MCS).

(2011) used in our analysis will be subject to larger estimation errors.

Table (4.8) shows the out-of-sample average loss measures of the HAR-DRD with respect to the 2-step-F model at different sampling frequencies. Specifically, on the first line of each box, we show the result of Table (4.7) corresponding to the original frequencies Δ = 1-second (NYSE dataset) and Δ = 1-minute (Russell 3000 dataset). On the second line, we report the results obtained by sampling the log-prices at Δ = 5-minutes. With 5-minutes data, the maximum number of time-stamps per day is 78, and thus for $k = 100$ the realized covariance estimator is ill-conditioned. We thus report the results for $k = 5, 10, 25, 50$. We immediately note that the forecast gains of the 2-step-F are larger and more significant for Δ = 5-minutes. Of course, such effect is more pronounced in the NYSE dataset, since the new sampling at 5-minutes implies huge data reduction compared to the original 1-second sampling frequency. We also see that the statistical significance of the forecast gains improves considerably at 5-minutes, as shown by the model confidence set test, which always excludes the HAR-DRD from the 90% MCS.

The above results are a consequence of the score-driven filtering mechanism, which smooths out the noise and guarantees more robust estimates compared to the HAR-DRD. The fact that we can successfully test for such effect in the absence of observations of the true latent covariance is due to the use of loss measures that are robust to estimation errors (see Patton and Sheppard (2009) and Patton 2011). Finally, note that the use of lower frequencies is common in financial applications, as it may be dictated by liquidity or data source restrictions.

Model	MSE				Qlike			
NYSE	$k=5$	$k=10$	$k=25$	$k=50$	$k=5$	$k=10$	$k=25$	$k=50$
				$\Delta=1$-second				
2-step-F	1.0000	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	**0.9840***	1.0007*	1.0051*	1.0091	1.0091*	1.0196	1.0271	1.0433
				$\Delta=5$-minutes				
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	1.0833	1.0807	1.0915	1.0963	1.2760	1.2196	1.1189	1.0875
Russell	$k=5$	$k=10$	$k=25$	$k=50$	$k=5$	$k=10$	$k=25$	$k=50$
				$\Delta=1$-minute				
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	1.0084	1.0099	1.0083	1.0027*	1.0020*	1.0183	1.0263	1.0277
				$\Delta=5$-minutes				
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	1.0247	1.0205	1.0131	1.0103	1.0192	1.0208	1.0256	1.0251

Table 4.8: Relative Out-of-sample average loss measures of HAR-DRD with respect to the 2-step-F model at different sampling frequencies. Bold numbers denote the models having lowest loss measures. Asterisks denote the models that are part of the 90% model confidence set (MCS).

Portfolio analysis

In this section, we use a framework similar to that of Fleming et al. (2001), Fleming et al. (2003) and Bollerslev et al. (2018) to provide a quantitative assessment of the *economic* gains of switching from joint estimation based models (RWG, RWG-HAR, GAS-F, GAS-F-HAR) to the proposed two-step estimation based models (2-step-W, 2-step-F). Let us consider an investor who allocates her funds into k risky assets by pursuing a volatility timing strategy in the period of time from day 1 to day T. On day $t-1$, $1 < t \leq T$, the investor solves the Global Minimum Variance (GMV) problem:

$$\begin{aligned} &\min \omega_t' \Sigma_{t|t-1} \omega_t, \\ &\text{subject to } \omega_t' \mathbf{1} = 1 \end{aligned} \tag{4.46}$$

where ω_t is a $k \times 1$ vector of portfolio weights and $\Sigma_{t|t-1}$ is the covariance matrix of asset returns at time t computed based on information available up to time $t-1$. We assume that the investor faces transaction costs proportional to the portfolio turnover. The latter is defined as in DeMiguel et al. (2014), namely:

$$TO_{t-1} = \sum_{i=1}^{k} \left| \omega_t^{(i)} - \omega_{t-1}^{(i)} \frac{1 + r_{t-1}^{(i)}}{1 + \omega_{t-1}' r_{t-1}} \right| \tag{4.47}$$

where $\omega_t^{(i)}$, $r_t^{(i)}$ denote the i-th component of the vector of weights ω_t and the returns vector r_t, respectively. The portfolio return in excess of transaction costs is computed as:

$$r_{p,t} = \omega_t' r_t - cTO_t \tag{4.48}$$

where c is constant. We finally assume that the investor has a quadratic utility function:

$$U(r_{p,t}, \gamma) = (1 + r_{p,t}) - \frac{\gamma}{2(1+\gamma)}(1 + r_{p,t})^2 \tag{4.49}$$

where $\gamma > 0$ is the coefficient of risk aversion.

The economic gain of switching from a covariance model "s" to another covariance model "l" is defined as the quantity Δ_γ such that:

$$\sum_{t=1}^{T} U(r_{p,t}^{(s)}, \gamma) = \sum_{t=1}^{T} U(r_{p,t}^{(l)} - \Delta_\gamma, \gamma) \tag{4.50}$$

where $r_{p,t}^{(s)}$ and $r_{p,t}^{(l)}$ are the returns of the portfolios constructed with the covariance forecasts of model s and l, respectively. In other words, Δ_γ represents the return an investor with risk aversion γ would be willing to sacrifice in order to use the forecasts of model l in place of those of

model s in solving the GMV problem[6]. Table (4.9) reports the values of Δ_γ computed for $\gamma = 1$ and $\gamma = 10$. For sake of clarity, we report in the table the results obtained with the group of $k = 25$ assets belonging to the NYSE dataset. However, similar results are obtained when considering other portfolio dimensions. The setting of the analysis is the same as in the out-of-sample test of Section (4.4): the models are estimated in the sub-sample comprising the first 1000 business days, and the out-of-sample portfolios are computed for the remaining 1265 business days.

The values of Δ_γ reported under the columns entitled RWG and RWG-HAR represent the economic gains in annual basis points of switching from the latter two models to the 2-step-W. Similarly, those reported under the columns entitled GAS-F and GAS-F-HAR represent the economic gains of switching from the latter two models to the 2-step-F. We also show in the first column the economic gains of switching from 2-step-W to 2-step-F. Significance levels on Δ_γ are assessed through the Reality Check of White (2000), using the stationary bootstrap of Politis and Romano (1994). In addition, we report in the table the turnover, computed as in Eq. (4.47), the ex-post portfolio volatility, the average portfolio return and the Sharpe ratio. The constant c multiplying the turnover is set as $c = 0\%, 1\%, 2\%$, as in Fleming et al. (2003) and Bollerslev et al. (2018).

We first note that all the economic gains are positive, meaning that the investor would be willing to pay a positive annual amount in order to switch to the proposed two-step estimation based models. In particular,

[6]In principle, the investor might solve a general mean-variance problem instead of the GMV problem considered here. As discussed by Jagannathan and Ma (2003) and DeMiguel et al. (2009a), mean-variance problems rely on forecasts of expected returns, which are notoriously subject to large estimation errors and tend to distort the optimal portfolio solution. As we are interested in examining covariance forecasts, we focus on the GMV problem, in a similar fashion to Bollerslev et al. (2018).

	2-step-W	2-step-F	RWG	RWG-HAR	GAS-F	GAS-F-HAR
Turnover	0.239	0.263	0.205	0.241	0.246	0.323
Average return (%)	9.847	10.427	9.371	9.273	9.644	8.473
Ex-post volatility (%)	7.887	**7.875***	7.887	7.912	7.893	8.202
				$c = 0\%$		
Sharpe ratio	1.138	**1.209**	1.084	1.069	1.116	0.941
Δ_1	59.3		47.6	57.4	78.6*	198.7*
Δ_{10}	70.9		46.9	58.0	81.3*	228.1*
				$c = 1\%$		
Sharpe ratio	1.069	**1.132**	1.024	0.999	1.045	0.851
Δ_1	53.3		39.0	57.9	74.2*	213.8*
Δ_{10}	65.0		38.4	58.4	76.8*	243.2*
				$c = 2\%$		
Sharpe ratio	0.999	**1.055**	0.964	0.929	0.973	0.760
Δ_1	47.2		30.5	58.3	69.8*	228.9*
Δ_{10}	59.2		29.9	58.9	72.3*	258.3*

Table 4.9: For each covariance model, we report the portfolio transaction costs computed as in Eq. (4.47), the average portfolio return, the ex-post portfolio volatility and the Sharpe ratio. All quantities are annualized. We also report, for $c = 0\%, 1\%, 2\%$, the economic gains Δ_γ in annual basis points of switching from joint estimation based models with Wishart (matrix-F) density to the 2-step-W (2-step-F), for $\gamma = 1, 10$. Similarly, we report the economic gains of switching from the 2-step-W to the 2-step-F. The asterisks in portfolio volatilities indicate the models that belong to the 90% MCS of lowest ex-post volatilities. Significance levels on economic gains are instead computed based on the Reality Check of White (2000). The asterisks indicate the economic gains that are significantly different from zero at the 5% confidence level. Bold quantities denote the best performing model according to the measure specified in the first column.

the economic gains of switching from GAS-F and GAS-F-HAR to 2-step-F are highly significant. These gains are mainly imputable to the variance reduction featured by the 2-step-F model. On the one side, the flexibility provided by the proposed approach leads to better out-of-sample forecasts, which translate into lower ex-post portfolio volatility (see Engle and Colacito 2006 and Patton and Sheppard 2009). On the other side, the two-step procedure naturally leads to a loss of statistical efficiency, which in turn determines more erratic portfolio weights and a larger turnover. Our results

clearly indicate that the first effect, namely the superior forecasting ability, significantly dominates the loss of statistical efficiency in the context of a risk-averse investor maximizing a quadratic utility function. In the case of models based on the Wishart density, the ex-post volatility of the 2-step-W portfolio is statistically indistinguishable from that of the RWG portfolio. Accordingly, we find that the corresponding economic gains are not statistically significant. However, they are still positive due to the larger average return featured by the 2-step-W portfolio. We finally note that the economic gains of switching from the 2-step-W to the 2-step-F are also positive, confirming the result that the matrix-F density provides a better description of realized covariance time-series.

4.5 Conclusions

We have introduced a two-step estimation procedure for score-driven realized covariance models based on Wishart and matrix-F densities. By employing the score, which is a martingale difference by construction, the proposed models automatically include a correction in the spirit of Aielli (2013). In the case of the Wishart density, our class of models reduces to the model proposed by Bauwens et al. (2012). In the case of the matrix-F density, we obtain a new realized covariance model that can be estimated in two steps and that accounts for the fat-tails of realized covariance time-series. More specifically, the main advantage of the method is that it has a higher degree of flexibility in the estimation of volatilities. In the first step, the latter are indeed separately estimated by univariate realized volatility models with different parameters. The model is therefore easy to estimate in large dimensions and, compared to joint estimation, is less affected by

the curse of dimensionality.

Through a Monte-Carlo study, we have examined the finite sample properties of the two-step estimator and shown that it recovers accurate estimates of the parameters. We have also examined the ability of the model to capture misspecified correlation dynamics. The comparison between the two-step procedure and standard joint estimation has been studied extensively through both Monte-Carlo simulations and empirical data. We have found that the additional flexibility of the two-step procedure translates into in-sample and out-of-sample forecast gains. The latters are found to be more significant in portfolios characterized by different levels of heterogeneity in the persistence of the volatilities. In particular, we have found statistically significant evidence of superior forecast ability in portfolios of high-dimensions, where volatilities are more likely to exhibit different persistences.

We have also examined the performance of the methodology under very noisy estimates of the underlying matrix-variate process. It is obtained that the score-driven filtering mechanism leads to increasingly better forecasts compared to methods not accounting for estimation errors. As a final experiment, we have assessed the economic gains of switching from standard joint estimation to the proposed two-step approach. It is found that a risk-averse investor would be willing to pay a positive amount to adopt the forecasts of the two-step models in constructing her portfolio.

Chapter 5

Asset allocation and Portfolio Instability in the Low for Long environment

The contents of this chapter are the result of a joint work with Thomas Kostka and Lieven Hermans at the European Central Bank. This paper is currently under revision to the ECB Working Paper Series.

5.1 Introduction

Monetary policy has desired and undesired effects on financial assets' risk premia and financial intermediaries' balance sheets and ultimately on financial stability. The *risk-taking channel (RTC) of monetary policy* established this link, originally for the banking sector (Borio and Zhu, 2012), and later extended to non-banks (Adrian and Song Shin, 2010): Low interest rates prolonged for an extended period (Low for Long in financial jargon) invite expanding the leverage on financial institutions balance sheets and increased

investment in riskier financial assets, thereby lowering their risk premia and reducing their volatility (Drechsler et al., 2018). But, as investors reach for yield during spells of low rates, their portfolios are increasingly prone to shocks as volatility spikes. As investors de-leverage and reduce their riskier positions - that is, as they run for the exit - they exacerbate the asset price deflation and inflict even higher volatility on their portfolios.

This chapter provides evidence that a simple Markowitz portfolio (Markowitz, 1952) is able to replicate this very mechanism. In doing so, it proposes a simple indicator of portfolio instability - measuring the variability of optimal portfolio shares in response to external shocks - which turns out to be higher in low rate environments. We show empirically that low rates boost excess returns but also reduce asset price volatility, thereby raising optimal leverage ratios and risky asset shares in the portfolio. As a result, these portfolios are noticeably more sensitive to changes in the model parameters (expected mean return and variance). Moreover, we provide a real-world example of our model prediction showing how portfolio shares were affected by the Covid shock as opposed to a counterfactual scenario in which the initial interest rate level was higher: Initial levels of portfolio leverage was higher and thus the decline in risky asset portfolio shares more rapid as relative to a higher rate environment. Hence, the low rate environment has arguably amplified the financial market turmoil in spring 2020. Figure 5.1 documents this noticeable drop in riskier positions held by euro area based investors. Finally, we extend our analysis to *Risk-Parity portfolios* introduced in Maillard et al. (2010), an increasingly popular approach to risk diversification in the portfolio selection. This approach is more robust to estimation errors in multi-asset portfolios (Cesarone et al., 2020). The results suggest that the interest rate level impact on portfolio instability

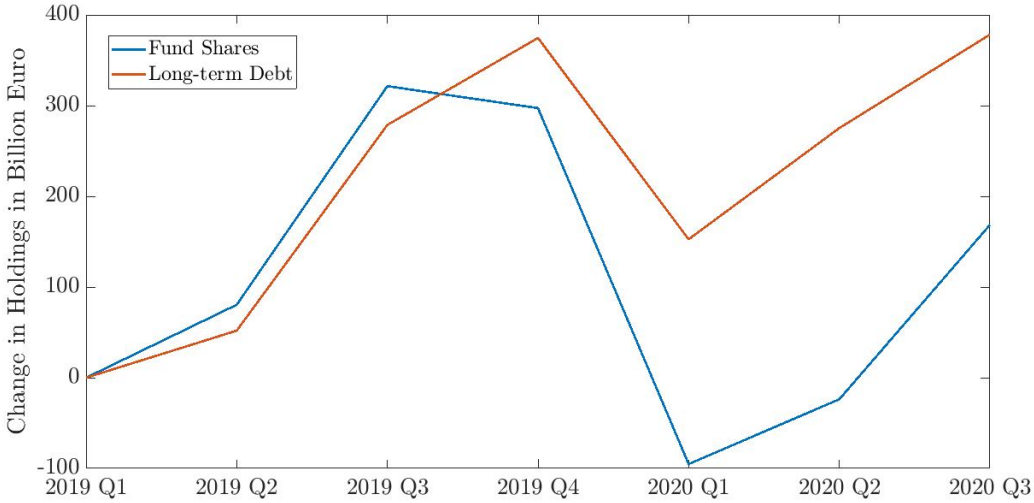

Figure 5.1: Quarterly aggregate evolution of Fund shares and Long-term debt in the Euro Area investors' portfolios. Amounts are in billions of euros. The data are taken from Security and Holdings Statistics Sector (SHSS) database.

is not specific for the mean-variance portfolio, but it is important also for investors who engage in volatility targeting. In the appendix, we show that the results are robust across different regions.

5.2 Framework

We consider a universe of K assets and an investor who allocates his/her portfolio solving, at any time t, the asset allocation problem $\omega_t^* = f(\theta_t)$. θ_t is the conditioning information set available to the investor at time t and ω_t^* portfolio weights of K asset portfolio. From now on, where possible, we use θ instead of θ_t for simplicity of notation. In the seminal paper of Markowitz (1952), the mean variance investors is introduced under the assumption

that returns are normally distributed,

$$\omega^* = \underset{\omega \in \mathbf{R}^K}{\arg\min} \quad \frac{1}{2}\omega' \Sigma \omega - \gamma \left[\omega' \tilde{\mu} + (1 - \omega' \mathbf{1}) r^f\right] \tag{5.1}$$

where $\mathbf{1}$ is the K-vector of all ones, r_t^f is the risk-free rate, $\tilde{\mu}$ is the K vector of expected returns and Σ is the $K \times K$ matrix of expected covariance matrix. γ is known as risk-aversion parameter. The higher γ, the higher is the investor's risk tolerance and the higher is the risk of the optimal portfolio. The optimization problem (5.1) has a closed form solution given by

$$\omega_t^* = \gamma \Sigma^{-1} \mu \tag{5.2}$$

where $\mu = \tilde{\mu} - r^f$ is the vector of *expected excess return*. The portfolio in equation (5.2) is known as tangency portfolio. The mutual fund theorem of Merton (1972b) shows that the optimal portfolio is a linear combination of the tangency portfolio (5.2) and the risk-free asset.

In the Markowitz portfolio, the vector of relevant information needed to build the optimal portfolio is $\theta = (\mu, \text{vech}(\Sigma_t), \gamma)'$[1].

Portfolio Instability

Subsequently, we study the instability of investors portfolios with respect to changes in the input variables θ. We define the *Portfolio Instability* (PI) as a measure to evaluate the amount needed to re-balance the portfolio in the face of a shock. When expectations change, the investor needs to re-balance his/her portfolio and may thereby also amplify market turbulence or liquidity shortages. Let us consider θ^{pre} as the vector of relevant information before the advent of a shock and θ^{post} as the vector of relevant information

[1]The half-vectorization operator vech converts a matrix into a column vector. Specifically, the half-vectorization of a symmetric $K \times K$ matrix A is the $K(K+1)/2 \times 1$ column vector obtained by vectorizing only the lower triangular part of A

after the shock. Hence, by definition the external shock to the information vector is given by $v = \theta^{\text{post}} - \theta^{\text{pre}}$. Portfolio instability is then defined as the directional derivative of the optimal portfolio $\omega^*(\theta^{\text{pre}})$ with respect to the shock v;

$$\text{PI}_v(\theta^{\text{pre}}) = D_v\omega(\theta^{\text{pre}}) = \lim_{h\to 0} \frac{\omega(\theta^{\text{pre}} + hv) - \omega(\theta^{\text{pre}})}{h} =^* D\omega(\theta^{\text{pre}}) \cdot v \in \mathbf{R}^K \tag{5.3}$$

where $D\omega(\theta)$ is the Jacobian and * holds when the solution $\omega(\theta)$ is differentiable. It's worth noticing that $D_v\omega(\theta^{\text{pre}})$ is a K dimensional vector where the entry i represents the volume of the portfolio reallocation in the asset i. The higher the absolute value of the element i, the higher is the instability of the portfolio weight ω_i^* to the shock v. The element of v bears a positive (negative) sign for buy (sell) transactions. We define the *Total portfolio Instability* (TPI) as the ℓ_2-norm of the vector $\text{PI}_v(\theta)$

$$\text{TPI}_v(\theta) = ||\text{PI}_v(\theta)||_2 = \sqrt{\sum_{i=1}^{K} (\text{PI}_v(\theta)(i))^2}. \tag{5.4}$$

The higher the value of $\text{TPI}_v(\theta)$, the higher the turnover induced by the shock v and the higher is the possible instability that the shock v will cause in the investor's portfolio.

Example of portfolio instability with two risky assets.

The following example provides some economic intuition behind the framework. Consider a market with two risky assets A and B where the expected annual variance of asset A is 4%, the expected annual variance of asset B is 1% and the expected correlation between the two assets is 0.2, which leads to an expected covariance of 0.4%. Let us consider that the initial expectation for the annual returns is initially 6.5% for asset A and 4% for

asset B. An investor allocates his/her portfolio by solving the Markowitz problem (5.1) with $\gamma = 0.2$. The investor finds that the optimal portfolio is composed by investing 26% of the portfolio in asset A and the 74% in asset B. After an external shock, the investor revises down the expected returns to 4.5% for asset A and 3% for asset B. Overall, the initial value of θ was

$$\theta^{\text{pre-shock}} = \left[\mu, \sigma_A^2, \sigma_B^2, \sigma_{A,B}, \gamma\right] = [6.5\%, 4\%, 4\%, 0.4\%, 1\%, 0.2]$$

and after a shock in the expectations with

$$v^{\text{shock}} = [-0.02, -0.01, 0, 0, 0, 0]$$

the investor finds itself in a new state of the world which is characterised by

$$\theta^{\text{post-shock}} = \theta^{\text{pre-shock}} + v^{\text{shock}} = [4.5\%, 3\%, 4\%, 0.4\%, 1\%, 0.2]\,.$$

In the new state of the world $\theta^{\text{post-shock}}$, the optimal asset allocation is given by investing the 21% of the portfolio into asset A and the 79% in asset B. That corresponds to a reduction of 5 percentage points in asset A which are being reinvested in asset B. Figure 5.2 shows the optimal allocation surface which represents the investment in asset A for different values of expected returns. The red dot shows the optimal allocation in the $\theta^{\text{pre-shock}}$ environment while the blue dot shows the new optimal allocation in the state of the world $\theta^{\text{post-shock}}$. The blue line which connects the two dots is the line generated by the directional derivative. The total portfolio instability is the length of the segment between the two dots. The length of the segment depends on the shape of the optimal allocation surface. In this example, $\text{TPI}_{v^{\text{shock}}}(\theta^{\text{pre-shock}}) = 0.067$.

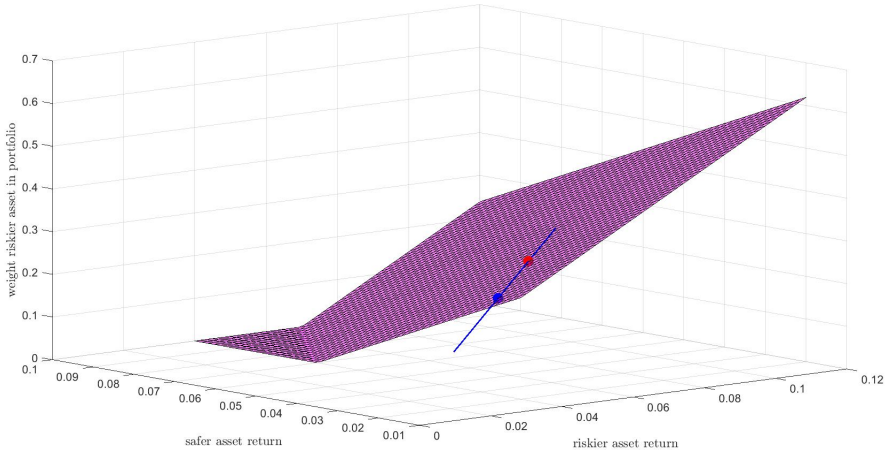

Figure 5.2: Directional derivative. In the z-axis we report the weights of the riskier asset A according to the expected returns of the two assets. The x-axis and y-axis show the expected returns of asset B and asset A respectively.

5.3 The role of monetary policy in the portfolio instability

In this section, we consider the Markowitz portfolio in equation (5.1) and we study the sensitivity of portfolio weights and the Total Portfolio Instability metric with respect to changes in the risk-free short rate. Thanks to the mutual fund theorem, we may restrict our analysis to two assets; one risky asset and one risk-free asset. We denote with σ_t^2 and r_t the variance and the expected returns of the risky asset and with r_t^f the risk-free rate with $r_t > r_t^f$. In this framework, the effect of the interest rate change can be transmitted using three channels: i) the direct impact of the risk-free rate on the volatility, ii) the direct impact of the risk-free rate on the excess returns and iii) the indirect impact of the risk-free rate on the excess returns through the volatility. We denote the key relations, with assumptions on

the signs, as follow

$$\begin{aligned} \frac{\partial \sigma_t^2}{\partial r_t^f} &= \lambda_t > 0 \\ \frac{\partial (r_t - r_t^f)}{\partial r_t^f} &= \psi_t. \\ \frac{\partial (r_t - r_t^f)}{\partial \sigma_t^2} &= \mu_M > 0. \end{aligned} \tag{5.5}$$

The first line of equation (5.5) denotes the relation between the expected conditional variance of the risky asset and the risk-free rate. According to Campbell (1987), Glosten et al. (1993) and Scruggs (1998) the variance is positively related with the risk-free rate. The third line refers to the risk premium. Even if hard to evaluate, the risk premium is strictly positive for any risk-averse investor. The second line is the more interesting. Mechanically, a lower risk-free rate augments the excess return of the risky asset ($\psi_t < 0$). However, a lower risk-free rate may also lower the excess return ($\psi_t > 0$) through its dampening effect on volatility, or the quantity of risk, ($\lambda_t > 0$) which in turn lowers the risk premium ($\mu_M > 0$). Given that both the alternative are plausible, we do not make any assumption on the sign of ψ_t.

The investor optimal portfolio (5.2) with only one risky asset reduces to

$$\begin{aligned} \omega_t^1 &= \frac{\gamma (r_t - r_t^f)}{\sigma_t^2} \\ \omega_t^2 &= 1 - \omega_t^1, \end{aligned} \tag{5.6}$$

where ω_t^1 and ω_t^2 are the portfolio share of the risky asset and the risk-free asset, respectively. The time index t will henceforth be omitted for simplicity of notation.

The first quantity of interest is the *Risk-taking channel*, i.e. the sensitivity of the risky asset's portfolio weight (ω^1) with respect to the risk-free

rate level. In principle, a lower interest rate level should increase the attractiveness and hence the portfolio weight of the risky asset.

Proposition 1. *The Risk-taking channel of Monetary Policy exists in the Markowitz framework if and only if the relative weight of the risky asset is above a certain threshold. In other words*

$$\frac{d\omega^1}{dr^f} < 0 \iff \omega^1 > \frac{\gamma\psi}{\lambda}. \tag{5.7}$$

The proof is in Appendix C.1. A sufficient condition for (5.7) to hold is $\lambda > 0$ and $\psi < 0$, that is that risky asset volatility declines and their excess return rises as the risk-free is lowered, thus making the risky asset more attractive from both a risk and return perspective. Central banks exploit this mechanism in the context of the *Risk-taking channel of monetary policy* Borio and Zhu (2012).

The second quantity we are interested in is the sensitivity of the portfolio instability with respect to the risk-free rate level. For simplicity we separate the analysis for exogenous changes in the different parameters of $\theta = \left[r - r^f, \sigma^2, \gamma\right]$.

Proposition 2. *The sensitivities of the Total Portfolio Instability measures* (5.4) *for the different type of exogenous shock with respect to the interest rate are*

$$\frac{dTPI_{[v^{ret},0,0]}}{dr^f} = -\sqrt{2}\frac{\gamma\lambda|v^{ret}|}{(\sigma^2)^2}. \tag{5.8}$$

$$\frac{dTPI_{[0,v^{var},0]}}{dr^f} = \begin{cases} \sqrt{2}\frac{\frac{d\omega^1}{dr^f}\sigma^2 - \lambda(\omega^1 - \gamma\mu_M)}{(\sigma^2)^2}|v^{var}|, & \text{if } \omega^1 > \gamma\mu_M, \\ \sqrt{2}\frac{-\frac{d\omega^1}{dr^f}\sigma^2 - \lambda(\gamma\mu_M - \omega^1)}{(\sigma^2)^2}|v^{var}|, & \text{if } \omega^1 < \gamma\mu_M, \end{cases} \tag{5.9}$$

$$\frac{dTPI_{[0,0,v^{\gamma}]}}{dr^f} = \sqrt{2}\frac{\frac{d\omega^1}{dr^f}}{\gamma}|v^{\gamma}|. \tag{5.10}$$

The proof of the proposition can be found in Appendix C.1. The signs of derivatives (5.8), (5.9) and (5.10) denotes the type of relation between the Total Portfolio Fragility and the interest rate. If the sign of the derivatives are positive, then a low level of interest rate generates, ceteris paribus, a lower selling or buying pressure following a shock and the optimal portfolios are more robust. If the signs are negative, then the optimal portfolio in a low rate environment is, ceteris paribus, more fragile to exogenous shocks. In the latter scenario, stability risks may arise as an unexpected event generates bigger rebalancing pressures and may exacerbate the liquidity and financial conditions.

The sign of the derivative (5.8) is negative under the first assumption in (5.5) supported by Campbell (1987), Glosten et al. (1993) and Scruggs (1998). The sign of the third derivative strictly depends on the sign of the right-hand side of equation (5.7), i.e. if the Risk-taking channel exists then the sensitivity Total Portfolio Instability sensitivity to the interest rate is negative. The second equation is the more complicated however, in the applications the first case ($\omega^1 > \gamma\mu_M$) holds as we show in 5.3. In this scenario, again the sign of the Total Portfolio Fragility depends on the risk-taking channel; if the risk-taking channel exists, then a shock in the variance creates bigger market movements in the low rate environment. These findings posit that when the central bank decreases the interest rate, investors react by increasing the weight in the risky asset and decreasing the portfolio weight in the risk-free asset. While this may be intended by the central bank, the increasing weight in the risky asset creates instability in the investor's portfolio with respect to exogenous shocks in any of the model parameters. It's worth noticing that these results do not depend on the model used to evaluate the excess returns and volatility but they are

intrinsic in the mean-variance optimization. In the next section we test the assumptions in (5.5) using parametric models inspired by Scruggs (1998).

Econometric model

We deploy the econometric model developed in Glosten et al. (1993) and Scruggs (1998) to test the hypotheses put forth in the previous section. In particular, we explore different model specifications in which we estimate the parameters from equations (5.7), (5.8), (5.9) and (5.10).

Tests are based on weekly European market data spanning from January 1999 to October 2020. For the risk-free asset, we use the 1 week EONIA rate and the market portfolio is proxied by the broad EuroStoxx 600. The in-sample period spans from January 1999 to December 2019. The Covid-19 crisis period serves as the out-of-sample period. In Appendix C.2 we present a similar study for the US and a longer time span (January 1952 to December 2019), showing that our main findings are robust across time and regions.

We start from the conditional one-factor model derived which takes the theoretical ground from the ICAPM model (henceforth *model 1.A*). This model assumes the existence of a risk-averse representative agent with the wealth function $J(W(t), X(t), t)$, where $W(t)$ is wealth and $X(t)$ is the state variable. The expected excess return of the one risky asset is given by

$$E_{t-1}[r_t - r_t^f] = \left[\frac{-J_{WW}W}{J_W}\right]\sigma_t^2 + \left[\frac{-J_{WX}}{J_W}\right]\sigma_{MX,t}. \tag{5.11}$$

where $\sigma_{MF,t}$ is the expected covariance of the risky asset with the state variable X, given the filtration at time $t-1$. Since $J_W > 0$ and $J_{WW} < 0$, $\frac{-J_{WW}W}{J_W}$ is greater than zero while we assume, for simplicity, that the marginal utility of wealth is state-independent ($J_{MX} = 0$), thereby following

Merton (1980), Campbell (1987) and Glosten et al. (1993). The expected excess return is therefore given by

$$\begin{aligned} r_t - r_t^f &= \mu_0 + \mu_M \sigma_t^2 + \epsilon_t, \\ \sigma_t^2 &= Var_{t-1}(\epsilon_t) \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned} \tag{5.12}$$

Model in equation (5.12) is the baseline model where the risk-free rate does neither affect the market excess return nor the conditional variance. Empirically, the sign of μ_M is not clear in the financial literature. On the first hand, several results show significant positive risk-return relation (e.g. Scruggs 1998; Bali and Peng 2006; Chiang et al. 2015). On the other hand, negative risk-return relation is also found in the literature (e.g. Campbell 1987; Brandt and Kang 2004; Ang et al. 2009). Other papers find contradictory results. For instance, Glosten et al. (1993) finds a positive sign of μ_M which becomes negative when the federal rate is included in the GARCH equation. Glosten et al. (1993) introduce the empirical setting to test the risk-return relationship with a GARCH-M model with the interest rate as an exogenous variable (henceforth *model 1.B*).

$$\begin{aligned} r_t - r_t^f &= \mu_0 + \mu_M \sigma_t^2 + \epsilon_t, \\ \sigma_t^2 &= Var_{t-1}(\epsilon_t) \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \tilde{\lambda} r_t^f. \end{aligned} \tag{5.13}$$

In (5.13) the risk-free rate indirectly affects excess return via its effect on the conditional variance ($\mu_M \tilde{\lambda}$.) We find a positive relationship between the risk-free rate and the conditional variance ($\tilde{\lambda} > 0$) [2]. According to the

[2]Glosten et al. (1993) finds an intuitive explanation behind this relationship: *"The use of nominal interest rates in conditional variance models has some intuitive appeal. It has been well known since Fischer et al. (1981) that the variance of inflation increases with its level. To the extent that short-term nominal interest rates embody expectations about inflation, they could be a good predictor for future volatility in excess returns."*

ICAPM of Merton (1973) there exists a positive relationship between the conditional variance and the excess return reflecting investors' risk aversion ($\mu_M > 0$). These two findings imply that model $1.B$ exhibits a positive relationship between the risk-free rate and the excess return[3].

Finally, we are aiming for more clarity about the role of the risk-free rate by allowing it to directly affect excess returns in *model 1.C*.

$$\begin{aligned} r_t - r_t^f &= \mu_0 + \mu_M \sigma_t^2 + \mu_f r_t^f + \epsilon_t, \\ \sigma_t^2 &= Var_{t-1}(\epsilon_t) \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \tilde{\lambda} r_t^f. \end{aligned} \tag{5.14}$$

To summarise, models 1.A, B and C relate to assumptions (5.5) as follow:

$$\begin{aligned} \text{model 1.A} \quad & \frac{\partial \sigma_t^2}{\partial r_t^f} = \lambda = 0 \\ & \frac{\partial (r - r_t^f)}{\partial r_t^f} = \psi = 0 \\ & \frac{\partial (r_t - r_t^f)}{\partial \sigma_t^2} = \mu_M \\ \text{.model 1.B} \quad & \frac{\partial \sigma_t^2}{\partial r_t^f} = \lambda = \tilde{\lambda} \\ & \frac{\partial (r - r_t^f)}{\partial r_t^f} = \psi = \mu_M \tilde{\lambda} \\ & \frac{\partial (r_t - r_t^f)}{\partial \sigma_t^2} = \mu_M \\ \text{.model 1.C} \quad & \frac{\partial \sigma_t^2}{\partial r_t^f} = \lambda = \tilde{\lambda} \\ & \frac{\partial (r - r_t^f)}{\partial r_t^f} = \psi = \mu_M \tilde{\lambda} + \mu_f \\ & \frac{\partial (r_t - r_t^f)}{\partial \sigma_t^2} = \mu_M \end{aligned} \tag{5.15}$$

.

[3]In turn,Laine (2020) finds an opposite sign for Europe during the Covid-19 crisis

where $\tilde{\lambda}$ is estimated in model 1.A, B and C, μ_M is estimated in model 1.B and C and μ_f is estimated in model 1.C. Table 5.1 reports the corresponding econometric estimates of the parameters of models 1.A, 1.B and 1.C. We refer to *slightly significant* when the p-value of the two-tailed t-test is between 0.05 and 0.1, to *significant* when the p-value of the two-tailed t-test is smaller than 0.05. In model 1.A the relationship between the conditional variance and the excess return is positive and slightly significant ($\mu_M = 0.045$, t-stat = 1.696). In model 1.B, the risk-free rate is found to have a positive and slightly significant impact on the conditional variance ($\tilde{\lambda} = 1.770$ t-stat 1.668) as the impact of the conditional variance on the excess return becomes non-significant. As we explained, this behaviour may be due to the indirect impact of the interest rate in the excess return and it was previously observed in Glosten et al. (1993) and Scruggs (1998). In model 1.C where we separately consider the impact of the risk-free rate in the conditional variance and the expected return, $\tilde{\lambda}$ remains positive and marginally significant. The direct impact of the risk-free rate on the excess return is negative and non-significant ($\mu_f = -3.756$ and t-stat -1.484) but the relationship between the conditional variance and the risk-free rate becomes significant ($\mu_M = 0.059$ and t-stat 1.992).

Considering the lower part of the table, model 1.B and 1.C are in line with our assumption on the negative sign of the quantities (5.7), (5.8), (5.9) and (5.10). In model 1.C the sufficient condition for a negative relationship between the risk-free rate on the risky asset portfolio weight ($\psi < 0$) is satisfied since ψ is negative and $\tilde{\lambda}$ is positive. Model 1.B has a positive ψ and the sufficient condition is not satisfied. Yet, when evaluating (5.7), (5.8), (5.9) and (5.10) at each point in time, it is found that total portfolio instability tends to rise with declining levels in the risk-free rate.

In the last row of the table we show the AIC (Aikake Information Criterion). Model 1.A is the worst to capture the dynamic of the data, model 1.B ranks second while model 1.C is the best performing.

The difference in terms of fit from model 1.B and model 1.C is very small. If we consider the AIC, model 1.C is preferred over model 1.B, however, the Likelihood Ratio Test does not reject the null of model 1.B with a p-value of 0.06. For this reason, the results of both model 1.B and model 1.C are examined. We evaluate the Total Portfolio Instability of investor (5.1) over the horizon from January 1999-December 2019. We assume an investor with a risk aversion parameter of $\gamma = 1\%$, and set the exogenous shocks to $v^{\text{ret}} = -5\%$ (i.e. a decrease in the excess return by 5 percentage points per annum), $v^{\text{var}} = 30\%$ (i.e. an increase in the annualised volatility by 30 percentage points) and $v^{\gamma} = -0.2\%$ (i.e. a decline in the risk aversion parameter by 0.2 percentage points which, in our case, means that the investors become more risk-averse) . Figure 5.3 shows the evolution of the Total Portfolio Instability with respect to the two models. The results are filtered with a year rolling window average. Using both models, we record a steady increase in portfolio instability in recent years. At the face of the results of the previous section, the secular decline in both the risk-free rates and in market volatility have set the conditions for total portfolio instability to rise to the highest level ever recorded.

Non-linear behaviour of the Total Portfolio Instability

In the previous two sections, we showed mathematically and empirically that portfolio shifts in reactions to shocks are more pronounced in the presence of low (vs. high) rates. Indeed, owing to the non-linear properties of the TPI function, this relationship is convex with disproportionately high

Parameters	model 1.A	model 1.B	model 1.C
μ_0	0.049	0.056	0.086
	(1.106)	(1.281)	(1.939)
μ_M	0.045	0.039	0.059
	(1.696)	(1.282)	(1.992)
μ_f			-3.756
			(-1.484)
ω	0.072	0.048	0.051
	(2.592)	(1.960)	(1.983)
α	0.193	0.182	0.183
	(2.842)	(2.936)	(3.071)
β	0.779	0.773	0.768
	(14.868)	(14.040)	(14.515)
$\tilde{\lambda}$		1.770	1.886
		(1.668)	(1.714)
ψ		0.068	-3.645
		(0.969)	(-1.454)
$\frac{\mathrm{d}\omega^1}{\mathrm{d}r^f}$		< 0	< 0
$\frac{\mathrm{d}TPI_{[v^{\mathrm{ret}},0,0]}}{\mathrm{d}r^f}$		< 0	< 0
$\frac{\mathrm{d}TPI_{[0,v^{\mathrm{var}},0]}}{\mathrm{d}r^f}$		< 0	< 0
$\frac{\mathrm{d}TPI_{[0,0,v^{\gamma}]}}{\mathrm{d}r^f}$		< 0	< 0
AIC	3403.3	3397.3	**3395.9**

Table 5.1: Estimation of models 1.A, 1.B and 1.C in equations (5.13) and (5.14) and model implied sensitivities to the interest rate. In the first section we report the estimates and the Robust t ratios in parenthesis. The estimates ψ is evaluated according to (5.5) and the Robust t-statistic is evaluated with the delta method. In the second section we report the different quantities in equations (5.7), (5.8), (5.9) and (5.10). When the sign is < 0 the partial derivative is negative in the whole sample, when the sign is > 0 the partial derivative is positive and when the sign is $\pm$ the partial derivative can be both positive and negative. In the last row we show the AIC and the best performing model according to this measure is bolded.

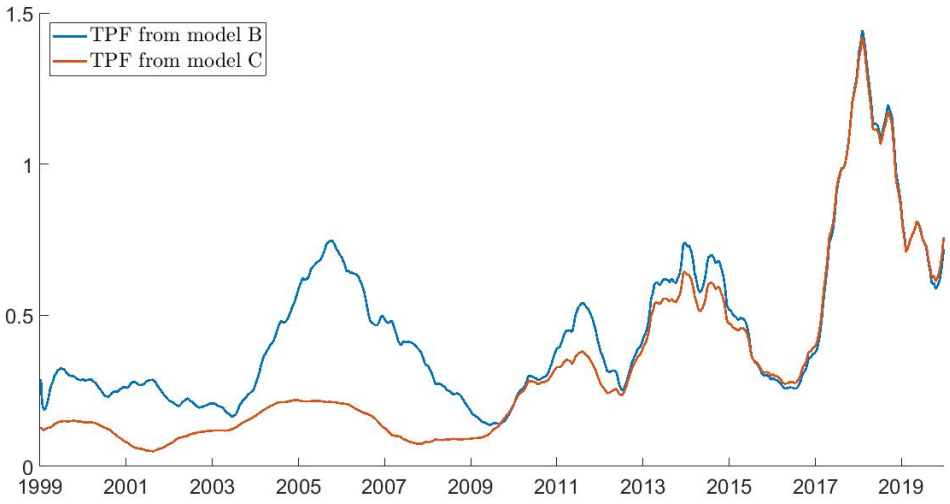

Figure 5.3: Total Portfolio Instability evaluated from model 1.B (5.13) and model 1.C (5.14). The measures are filtered using a 1-year moving window and assuming an impact of 30% in the variance and -5% in the expected excess return.

levels of portfolio instability for rates around or even below zero. Figure 5.4 depicts this non-linearity for the estimated parameters in table 5.1 for models 1.B and 1.C.

This property is particularly warning for policymakers as a decrease in the interest rate has non-linear effects. Cutting the interest rate when the levels are high may have a small effect on portfolio instability. However, this effect becomes more and more relevant for low levels of interest rate. In the next section, we provide a quantification of the implications of this property.

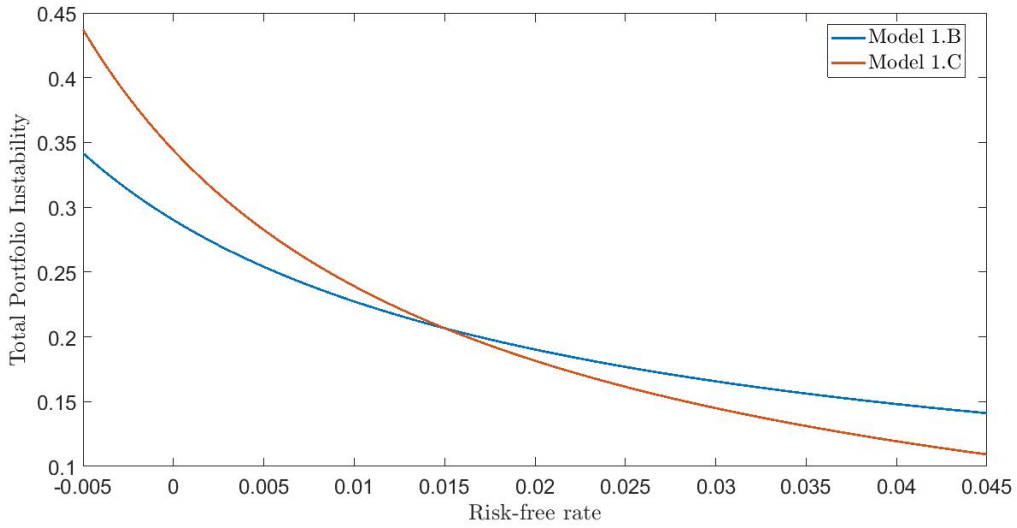

Figure 5.4: Total Portfolio Instability sensitivity evaluated from model 1.B (5.13) and model 1.C (5.14) for different values of the risk-free rate. For the calculations, we use the unconditional levels of σ^2 and $r - r^f$ given by models (5.13) and (5.14) for different levels of r^f.

Covid-19 crisis and the Markowitz investor. A counterfactual analysis

In this section, we deploy the estimated parameters from Section 5.3 to characterise the role of the low rates in the recent Covid-19 crisis. Our findings suggest that the pre-crisis low interest rate level have likely amplified the market turmoil in March 2020. We quantify by how much portfolio weights would have moved in models in the presence of higher interest rates in the framework presented by model 1.B (5.13) and model 1.C (5.14). We assume an investor solving the Markowitz problem in (5.6) as news about the risk-adjusted return of risky assets arrive $(\epsilon_t/\sqrt{\sigma_t^2})$. We do not consider the risk-free as a source of the shock but rather as a model parameter affecting the portfolio choices in real-time. As a first step, we use the observed EONIA rate to model a baseline scenario in which investors optimise their

portfolios in response to changes in the expected return and variance profile of the various asset classes induced by the Covid-19 pandemic.

Using model (5.13) and (5.14) we forecast the expected excess return and variance and, as a residual we get the out of sample residuals ϵ_t^{out}. In a second step, we create a counterfactual, augmented EONIA path. We denote with $r_t^{f,\text{CF}}$ as the counterfactual risk-free rate and calculate the corresponding expected excess return and variance as

$$\begin{aligned} E_{t-1}\left[(r_t - r_t^f)^{\text{CF}}\right] &= \mu_0 + \mu_M \sigma_t^{2,\text{CF}} + \mu_f r_t^{f,\text{CF}} \\ \sigma_t^{2,\text{CF}} &= \omega + \alpha\left(\epsilon_{t-1}^{\text{OOS}}\right)^2 + \beta\sigma_{t-1}^2 + \tilde{\lambda} r_t^{f,\text{CF}}, \end{aligned} \tag{5.16}$$

which brings to the portfolio weights ω^{CF} using the explicit portfolio solution (5.6). Note that $\mu_f = 0$ for model 1.B. As counterfactual we use a risk-free rate that is 1 percentage point higher than the actual EONIA rate. In our comparative analysis, we scale the portfolio weights to 0 at the beginning of January 2020 so that we can compare the relative portfolio shifts in the course of 2020.

Figure 5.5 shows the evolution of weights under model 1.B, while figure 5.6 shows the evolution under model 1.C. Their dynamics are very similar. The reduction in the equity share during March was in the order of 15% of the portfolio size prevailing in January 2020, with a higher risk-free rate, the reduction of the equity could have been reduced from -15% to -12.5% if the risk-free rate had been 50 basis points higher risk-free rate and to -10% in the presence of 100 basis point higher risk-free rate. As expected, these counterfactual scenarios show that monetary policy could have played a role to alleviate, but not deleting, the rebalancing pressure induced by the Covid-19 shock.

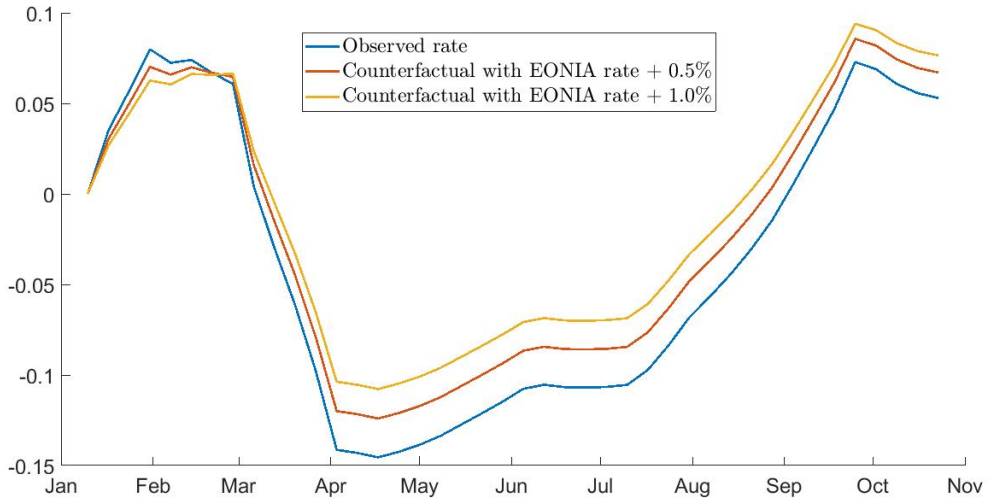

Figure 5.5: Change in the portfolio share since January 2020 given actual risk-free rates (in blue) and counterfactual risk-free rates (orange: EONIA + 0.5%, yellow: EONIA + 1%) in model 1.B (5.13). The measures are filtered using a 1-month moving window.

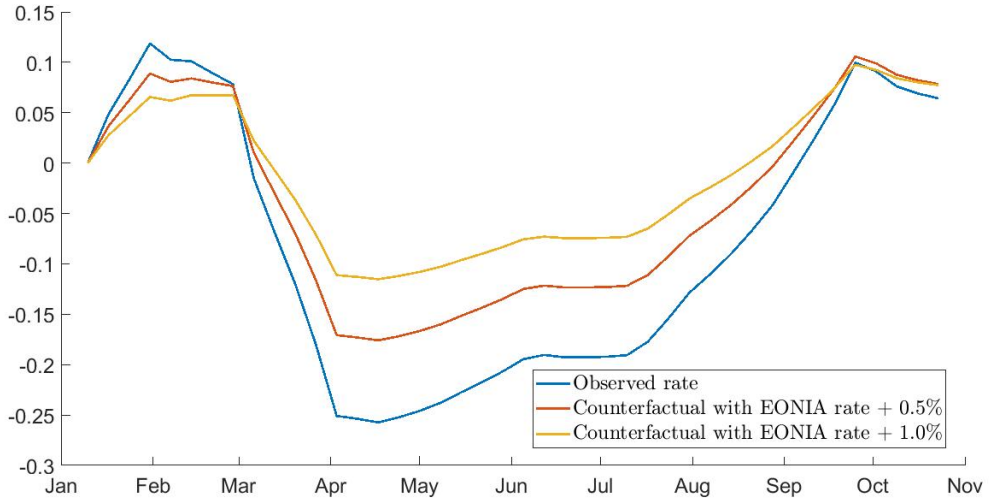

Figure 5.6: Observed sell-off (in blue) and counterfactual sell-off (in orange with 0.5% and yellow with 1%) evaluated from model 1.C (5.14). The measures are filtered using a 1-month moving window.

5.4 Risk Parity Investor

In the previous section, we analyze the behaviour of a Markowitz investor during the Covid-19 crisis and the increment of the Total Portfolio Instability in the last years of low rates. However, the Markowitz portfolio has two drawbacks. First, optimal Markowitz multi-asset portfolios typically allocate most of the value to only a few assets. Second, Markowitz weights are very sensitive to estimation errors, where small changes in the expected returns and variance estimation lead to very different optimal portfolios as documented in Best and Grauer (1991) and Chopra and Ziemba (1993). To avoid these drawbacks, a new strand of models which rely only on the risk measure is considered. Cesarone et al. (2020) shows that portfolios built using risk diversification strategies are the most robust to noise. In particular, they find that *Risk Parity Portfolios*, introduced by Maillard et al. (2010), are more stable with respect to estimation errors. Prior to the Covid-19 crisis, an estimated USD 300 billion invested was invested in funds following risk-parity strategies. The mere size of this market segment may render the portfolio flows induced by external shocks macro-critical and may thus be able to amplify the initial shock, e.g. by putting further downward pressure on the prices of riskier assets and/or raising market volatility. In this section, we replicate the Risk Parity portfolio to robustify the results presented in Section 5.3.

The Risk Parity Portfolio (RP) aims to spread the risk among all assets equally and each asset contributes to the total portfolio volatility, denoted as $\sigma(\omega_t)$, equally. More precisely, the volatility contribution of asset i should be equal to the volatility contribution of asset j, $\omega_t^i \frac{\partial \sigma(\omega_t)}{\partial \omega_t^i} = \omega_t^j \frac{\partial \sigma(\omega_t)}{\partial \omega_t^j}$ for any

asset i and j and any time t. That is,

$$\omega_t^i \frac{\partial \sigma(\omega_t)}{\partial \omega_t^i} = \omega_t^i \frac{(\Sigma_t \omega_t)_i}{\sigma(\omega_t)}. \tag{5.17}$$

At time $t-1$, the RP investor with a covariance matrix expectation $\hat{\Sigma}_t$, re-balances his/her portfolio weights ω_t solving the optimization problem:

$$\omega_t = \arg\min_{\omega} \sum_{i=i}^{K} \left[\frac{\sigma(\omega)}{K} - \omega^i \frac{(\Sigma_t \omega)_i}{\sigma(\omega)} \right]^2 \tag{5.18}$$

The portfolio obtained minimizing (5.18) is not well identified, indeed it has infinite solutions. As per the Markowitz portfolio (5.1) we can identify the optimal portfolio imposing that $\mathbf{1}\omega = 1$. However, we impose the volatility targeting constraint $\sigma(\omega) = \omega^*$. A key advantage of the volatility targeting constraint is that we can explore another dimension of the investor behaviour, where he/she increases leverage during low volatility periods and decreases the leverage during high volatility periods. The final optimization problem then reads

$$\begin{aligned} \omega_t = \arg\min_{\omega} \quad & \sum_{i=i}^{K} \left[\frac{\sigma(\omega)}{K} - \omega_i \frac{(\Sigma_t \omega)_i}{\sigma(\omega)} \right]^2 \\ \text{subject to} \quad & \sigma(\omega) = \sigma^* \end{aligned} \tag{5.19}$$

There is no closed-form solution for the risk parity portfolio, for this reason, there is no a priori mathematical relationship between the behaviour of the portfolio weights and the risk-free interest rate. In the following empirical exercise, we estimate the portfolio weights and the Portfolio Instability by means of numerical optimization and derivation.

Econometric model and Portfolio Instability

We consider a risk-parity investor with a target of 8% annual volatility and a universe of 3 risky assets, namely the equity (EUROSTOXX 600),

the long-term government bond (Markit iBoxx Sovereigns Eurozone Index) and the Corporate bond (Bloomberg Barclays Euro Aggregate Corporate Total Return Index).

Portfolio weights in a risk parity portfolio are set as to equalise the contribution of each asset (class) to overall portfolio volatility. Hence, the optimal weights depend on the level of the target volatility. We assume that the risk-parity investor allocates 50% of his/her risk in the Equity asset class and the 50% in the fixed income. Of the latter, 25% is allocated in the Government bond and the 25% is allocated in the corporate bonds. The investor forecasts the future covariance matrix using a Dynamic Conditional Correlation model (Engle, 2002b)

$$\begin{aligned} r_{i,t} - r_t^f &= \mu_i + \epsilon_{i,t} \\ \epsilon_{i,t} &= \sigma_{i,t} z_t, \quad E[z_t] = 0,\ E[z_t^2] = 1 \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 \\ V_t &= D_t R_t D_t \\ R_t &= \Delta_t^{-1} Q_t \Delta_t^{-1} \\ Q_t &= \bar{Q}(1 - \alpha_{DCC} - \beta_{DCC}) + \alpha_{DCC}(\epsilon_t \epsilon_t') + \beta_{DCC} Q_{t-1}. \end{aligned} \tag{5.20}$$

where $\bar{Q}$ is the unconditional mean of the pseudo-correlation matrix Q, D_t is the diagnoal matrix of conditional volatilities, and $\Delta_t = \text{diag}(Q_t)$ is the matrix with only the diagonal elements of the pseudo-correlation matrix. From now on we refer to model (5.20) as *Model 2.A*.

In the spirit of Section 5.3, we allow the risk-free rate to affect the

volatility dynamics. For this reason, we propose an alternative model

$$\begin{aligned} r_{i,t} - r_t^f &= \mu_i + \epsilon_{i,t} \\ \epsilon_{i,t} &= \sigma_{i,t} z_t, \quad E[z_t] = 0,\ E[z_t^2] = 1 \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 + \lambda_i r_t^f \\ V_t &= D_t R_t D_t \\ R_t &= \Delta_t^{-1} Q_t \Delta_t^{-1} \\ Q_t &= \bar{Q}(1 - \alpha_{DCC} - \beta_{DCC}) + \alpha_{DCC}(\epsilon_t \epsilon_t') + \beta_{DCC} Q_{t-1}, \end{aligned} \tag{5.21}$$

where λ_i captures the dependency between the risk-free rate and the volatility level. We denote this model as *Model 2.B*.

Table 5.2 shows the estimated parameters from model (5.20) and (5.21). The AIC shows that model 2.B provides the best fit. In addition, the likelihood ratio test rejects the null hypothesis of model 2.A with a p-value of 10^{-9}. The sensitivity of the volatility with respect to the risk-free level is positive and marginally significant for the equity and the corporate bonds, while it is negative and non-significant for the Government bond. In economic terms, low interest rates tend to mute the volatility in so-called risky asset classes (including equity and corporate bonds) but increase the volatility levels of benchmark bonds consistent with the notion of convexity in bond pricing. As in Section 5.3, we evaluate the behaviour of the total portfolio instability with respect to the risk-free rate. We define a range of possible risk-free rate levels ranging from -0.5% to 4.5%. In the GARCH-DCC-type model (5.21) the unconditional expected values are defined as

$$\begin{aligned} \bar{\sigma}_i^2(r_f^*) &= \frac{\omega_i + \lambda_i r_f^*}{1 - \alpha_i - \beta_i} \\ \bar{V} &= \bar{D}\bar{R}\bar{D}, \end{aligned} \tag{5.22}$$

where $\bar{D}$ is the diagonal matrix with the $\bar{\sigma}_i$ element on the i-th diagonal.

Parameters	Model 2.A			Model 2.B		
	Equity	Gov. Bond	Corp. Bond	Equity	Gov. Bond	Corp. Bond
ω	0.066	0.005	0.001	0.043	0.005	0.000
	(2.547)	(2.819)	(1.755)	(1.868)	(2.493)	(1.357)
α	0.176	0.092	0.091	0.165	0.093	0.082
	(3.060)	(3.634)	(3.700)	(3.242)	(3.681)	(3.345)
β	0.797	0.844	0.897	0.791	0.843	0.895
	(18.245)	(22.715)	(41.963)	(17.074)	(22.854)	(37.262)
λ				1.765	-0.012	0.024
				(1.707)	(-0.421)	(1.980)
α_{DCC}		0.039			0.038	
		(10.049)			(10.003)	
β_{DCC}		0.959			0.961	
		(240.477)			(252.157)	
AIC	1668.9			**1630.1**		

Table 5.2: Estimation of models 2.A and 2.B in equations (5.20) and (5.21). Any column represents a different asset class. In parenthesis we report the t-statistics In the last row we show the AIC and the best performing model according to this measure is bolded.

Using equation (5.22), we can study (T)PI of the risk-parity portfolio for varying levels of the risk-free rate. We study the portfolio shifts induced by a 20 percentage point volatility increase in all the asset classes and a parallel increase in all correlation coefficients by 10 percentage points. Figure 5.7 reports the results. The behaviour of the Total Portfolio Instability is qualitatively identical to the Total Portfolio Instability in figure 5.3. Portfolio Instability increases dis-proportionately as interest rates decline. Concerning the Portfolio Instability, the most susceptible asset class are corporate bonds whereas the share in government bonds is almost indifferent with respect to the interest rate level (as suggested by the estimated parameter λ in table 5.2). In addition, we see that the investor is susceptible to higher flight-to-safety behaviour for the low level of the risk-free rate. The results confirm the behaviour of the Total Portfolio Instability of the Markowitz investor, suggesting that our findings do not depend on the investors' utility optimization functions.

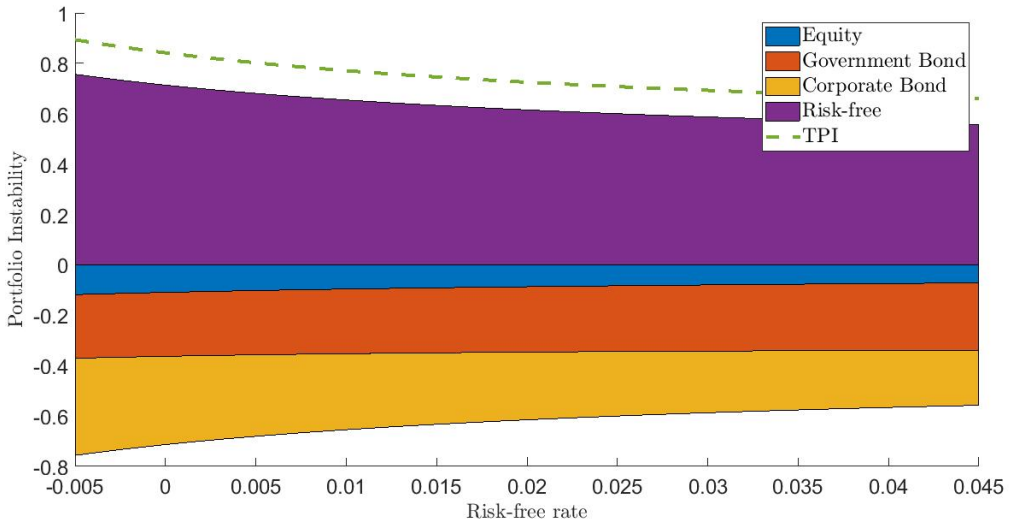

Figure 5.7: Total portfolio instability of the risk parity portfolio given different levels of the interest rate. In blue we report the TPI for the volatility shock scenario, in orange we report the TPI for the correlation shock scenario

Covid-19 crisis and the Risk parity investor. A counterfactual analysis

In this section, we perform the same analysis as in Section 5.3. The counterfactual analysis is based on the counterfactual rates $r_t^{f,\text{counter}}$ of 0.5 and 1% higher than the one observed in the EONIA rate. Figure 5.8 shows the results for a risk-parity investor during the Covid-19 crisis, with a volatility target of 8% annualized. Again, to make the result more realistic, we filter the weekly weights with a 4-period rolling window. The qualitative result confirms what we observe in figure 5.5 and 5.6, while the quantitative results differ a lot. The risk-parity investor with an 8% volatility target strongly uses the leverage in a period of low volatility, like the one before the Covid-19 pandemic. This high level of leverage translates into high portfolio rebalancing volumes during crises. In the baseline scenario, the

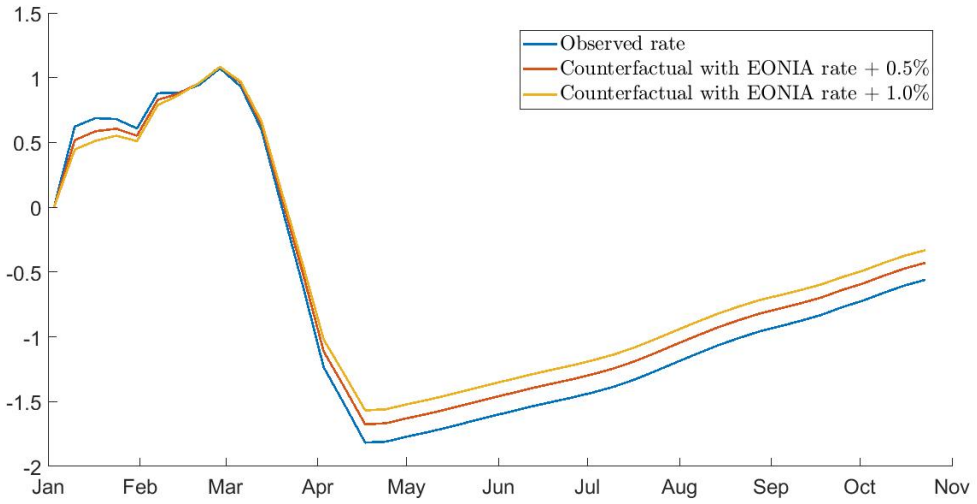

Figure 5.8: Observed sell-off (in blue) and counterfactual sell-off (in orange with 0.5% and yellow with 1%) evaluated from model 2.B (5.21). The measures are filtered using a 1-month moving window.

investor sells almost the 200% of the portfolio value of risky assets. This number would decrease to 170% and 150% in the Counterfactual scenario with a higher EONIA rate. For more information see Vassallo et al. (2020). Given that the risky portfolio is composed of 3 asset classes, we can assess the selling pressure on every component. Figure 5.9 shows the sell-off decomposed into the different components. In all three asset classes, a higher risk-free rate produces a smaller sell-off. The higher relative impact is on the Equity and the Corporate bond, which are the asset classes with a positive relation between variance and the risk-free rate. On the opposite, the government bond asset class shows to be more robust to differences in the risk-free level, as expected by the estimated parameter λ in table 5.2.

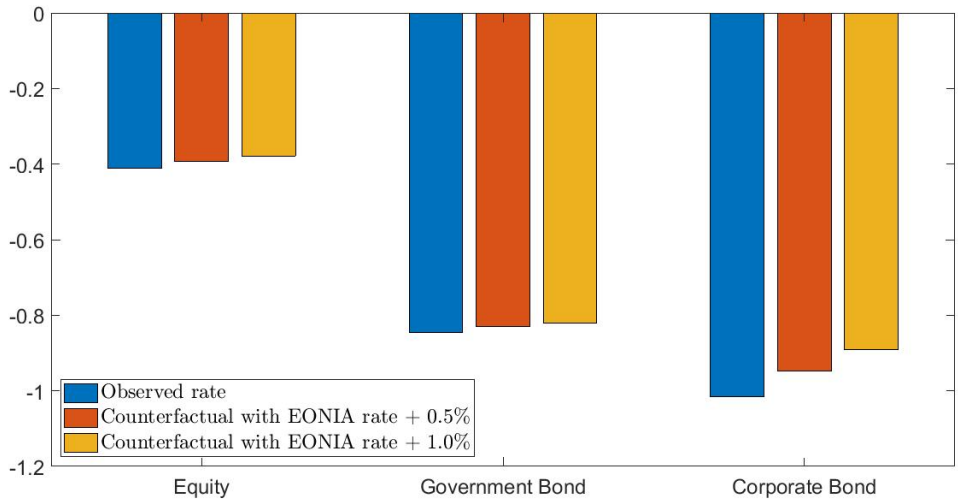

Figure 5.9: Observed sell-off (in blue) and counterfactual sell-off (in orange with 0.5% and yellow with 1%) evaluated from model 1.C (5.14). The measures are filtered using a 1-month moving window.

5.5 Conclusions

This chapter investigates the role of the low interest rate environment on portfolio reallocation following external shocks. A low level of interest rate increases exposures to risky assets. This effect is defined as *Risk-taking channel of monetary policy* and is crucial for the monetary policy transmission mechanism. The risk-taking behaviour of investors creates a potential financial stability vulnerability that we measure as *Portfolio Instability* which reflects the amount needed to rebalance the portfolio when a shock arrives. We show that in recent years the Total Portfolio Instability measure reached the maximum levels in recent history which can lead to severe financial stability issues following significant portfolio reallocations. Using both a mean-variance investor and a Risk Parity investor, we show that the portfolio rebalancing in the prevailing low rate environment during the Covid-19 crisis was significantly higher than would have occurred in a higher

rate environment.

Chapter 6

Conclusive remarks

The objective of this thesis is manifold.

In Chapter 3, we developed a filtering and modeling technique, named MLSS, to better extract information from multivariate sentiment time series. The filtered sentiment is decomposed into two components. The long-term sentiment captures the common trends and the long-term dynamics of the sentiment series, the short-term sentiment captures the fast trends of the investors' mood. Empirically, the extracted sentiment from the MLSS better explains contemporaneous and lagged returns with respect to other filtering techniques.

In Chapter 4, we proposed two score-driven realized covariance models. The first model is based on the Wishart distribution and the second one is based on the matrix-F distribution. The estimation procedure relies on two steps. In the first step, we estimate the volatility components, in the second step, we estimate the correlation dynamics. This two-step approach is more flexible to capture the different dynamics of the univariate and multivariate components. Using an out of sample exercise, we showed that the matrix-F distribution is more robust to outliers and brings to statistically significant

better forecasts. Finally, we showed that the 2-step methodology brings significant economic gains, where risk-averse investors are willing to switch from standard joint estimations to the proposed two-step approach in the portfolio construction.

In Chapter 5, we have studied the sensitivity of portfolio weights according to the risk-free rate level. We showed that low level of the risk-free rate brings to riskier portfolios. In addition, we introduced a measure of Portfolio Instability, where we showed that low level of the risk-free rate is associated with higher portfolio turnover following an exogenous shock. The recent Covid-19 crisis is used as a quasi-experiment to assess the economic impact of a low risk-free rate level on the market sell-off.

Bibliography

Abadir, K. M. and Magnus, J. R. (2005). *Matrix Algebra*. Econometric Exercises. Cambridge University Press.

Adrian, T. and Song Shin, H. (2010). Financial intermediaries and monetary economics. In Friedman, B. M. and Woodford, M., editors, *Handbook of MonetaryEconomics.*, volume 3 of *Handbook of Monetary Economics*, pages 601–650. Elsevier.

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics*, 31(3):282–299.

Algaba, A., Ardia, D., Bluteau, K., Borms, S., and Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.

Allen, D., McAleer, M., and Singh, A. (2015). Machine news and volatility: The Dow Jones Industrial Average and the TRNA real-time high-frequency sentiment series. *Handbook of High Frequency Trading*, pages 327–344.

Altavilla, C., Boucinha, M., and Peydró, J.-L. (2018). Monetary policy and bank profitability in a low interest rate environment. *Economic Policy*, 33(96):531–586.

Altunbas, Y., Gambacorta, L., and Marques-Ibanez, D. (2014). Does Monetary Policy Affect Bank Risk? *International Journal of Central Banking*, 10(1):95–136.

Andersen, T. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further u.s. evidence. *Journal of Financial Economics*, 91(1):1–23.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294.

Appel, G. (2003). Become your own technical analyst: How to identify significant market turning points using the moving average convergence-divergence indicator or macd. *The Journal of Wealth Management*, 6:27–36.

Audrino, F., Sigrist, F., and Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36(2):334 – 357.

Audrino, F. and Tetereva, A. (2019). Sentiment spillover effects for US and European companies. *Journal of Banking & Finance*, 106:542 – 567.

Avellaneda, M. and Lee, J.-H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7):761–782.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Bali, T. G. and Peng, L. (2006). Is there a risk-return trade-off? evidence from high-frequency data. *Journal of Applied Econometrics*, 21(8):1169–1198.

Banbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149 – 169.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93 – 101.

Bauwens, L., Braione, M., and Storti, G. (2016). Forecasting Comparison of Long Term Component Dynamic Models for Realized Covariance Matrices. *Annals of Economics and Statistics*, (123/124):103–134.

Bauwens, L., Braione, M., and Storti, G. (2017). A dynamic component model for forecasting high-dimensional realized covariance matrices. *Econometrics and Statistics*, 1(C):40–61.

Bauwens, L., Storti, G., Violante, F., et al. (2012). Dynamic conditional correlation models for realized covariance matrices. *CORE DP*, 60:104–108.

Beckers, S. (2018). Do social media trump news? The relative importance of social media and news based sentiment for market timing. *The Journal of Portfolio Management*, 45(2):58–67.

Bekierman, J. and Manner, H. (2018). Forecasting realized variance measures using time-varying coefficient models. *International Journal of Forecasting*, 34(2):276 – 287.

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The review of financial studies*, 4(2):315–342.

Blasques, F., Koopman, S. J., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1 – 18.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1):71 – 91.

Bonato, M., Caporin, M., and Ranaldo, A. (2012). A forecast-based comparison of restricted wishart autoregressive models for realized covariance matrices. *The European Journal of Finance*, 18(9):761–774.

Borio, C. and Zhu, H. (2012). Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism? *Journal of Financial stability*, 8(4):236–251.

Bork, L. (2009). Estimating US monetary policy shocks using a factor-augmented vector autoregression: an EM algorithm approach. *Available at SSRN 1358876*.

Borovkova, S. (2015). The role of news in commodity markets. *Available at SSRN 2587285*.

Borovkova, S., Garmaev, E., Lammers, P., and Rustige, J. (2017). SenSR: A sentiment-based systemic risk indicator. *Technical Report 553. De Nederlandsche Bank*.

Borovkova, S. and Lammers, P. (2017). Sector news sentiment indices. *Available at SSRN 3080318*.

Borovkova, S. and Mahakena, D. (2015). News, volatility and jumps: the case of natural gas futures. *Quantitative Finance*, 15(7):1217–1242.

Brandt, M. W. and Kang, Q. (2004). On the relationship between the conditional mean and volatility of stock returns: A latent var approach. *Journal of Financial Economics*, 72(2):217–257.

Brunnermeier, M. K. and Koby, Y. (2018). The reversal interest rate. *NBER Working Paper*.

Buccheri, G., Bormetti, G., Corsi, F., and Lillo, F. (2018). Filtering and smoothing with score-driven models. *arXiv preprint arXiv:1803.04874*.

Buccheri, G. and Corsi, F. (2019). HARK the SHARK: Realized Volatility Modeling with Measurement Errors and Nonlinear Dependencies*. *Journal of Financial Econometrics*.

Callot, L. A. F., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

Calomiris, C. W. and Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*, 133(2):299–336.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18(2):373 – 399.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281 – 318.

Campbell, J. Y. and Shiller, R. J. (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Shiller, R. J. (1988b). Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676.

Cavaliere, G. and Xu, F. (2014). Testing for unit roots in bounded time series. *Journal of Econometrics*, 178:259–272. Recent Advances in Time Series Econometrics.

Cesarone, F., Mango, F., Mottura, C. D., Ricci, J. M., and Tardella, F. (2020). On the stability of portfolio selection models. *Journal of Empirical Finance*, 59:210 – 234.

Chiang, T. C., Li, H., and Zheng, D. (2015). The intertemporal risk-return relationship: Evidence from international markets. *Journal of International Financial Markets, Institutions and Money*, 39(C):156–180.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223 – 236.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Peluso, S., and Audrino, F. (2015). Missing in asynchronicity: A Kalman-EM approach for multivariate realized covariance estimation. *Journal of Applied Econometrics*, 30(3):377–397.

Cox, D. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 8(2):93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *The Review of Economics and Statistics*, 96(5):898–915.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5):1461–1499.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven models for poisson counts. *Biometrika*, 90(4):777–790.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock Return Serial Dependence and Out-of-Sample Portfolio Performance. *The Review of Financial Studies*, 27(4):1031–1073.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.

Drechsler, I., Savov, A., and Schnabl, P. (2014). A model of monetary policy and risk premia. Working Paper 20141, National Bureau of Economic Research.

Drechsler, I., Savov, A., and Schnabl, P. (2018). A model of monetary policy and risk premia. *The Journal of Finance*, 73(1):317–373.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Engle, R. (2002a). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350.

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

Engle, R. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24:238–253.

Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30(2):212–228.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1):3 – 27.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.

Fama, E. F. (1970). Multiperiod consumption-investment decisions. *The American Economic Review*, 60(1):163–174.

Fischer, S., Hall, R. E., and Taylor, J. B. (1981). Relative shocks, relative price variability, and inflation. *Brookings Papers on Economic Activity*, 1981(2):381–441.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, 67(3):473 – 509.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

Gerber, A. S., Gimpel, J. G., Green, D. P., and Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *American Political Science Review*, 105(1):135–150.

Gilli, M. and Schumann, E. (2009). An empirical analysis of alternative portfolio selection criteria. *Swiss Finance Institute, Swiss Finance Institute Research Paper Series*.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1):211 – 223.

Gorgi, P., Hansen, P. R., Janus, P., and Koopman, S. J. (2018). Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model*. *Journal of Financial Econometrics*, 17(1):1–32.

Gourieroux, C., Jasiak, J., and Sufana, R. (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167 – 181.

Groß-Klußman, A. and Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, 18(2):321–340.

Gupta, A. and Nagar, D. (1999). *Matrix Variate Distributions*. PMS Series. Addison-Wesley Longman, Limited.

Hakansson, N. H. (1971). Multi-period mean-variance analysis: Toward a general theory of portfolio choice. *The Journal of Finance*, 26(4):857–884.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Harvey, A. C. (1990). *Estimation, prediction and smoothing for univariate structural time series models*, pages 168 – 233. Cambridge University Press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

Hill, S. J., Lo, J., Vavreck, L., and Zaller, J. (2013). How quickly we forget: The duration of persuasion effects from mass communication. *Political Communication*, 30(4):521–547.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Jungbacker, B. and Koopman, S. J. (2008). Likelihood-based analysis for dynamic factor models. *Tinbergen Institute, Tinbergen Institute Discussion Papers*.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448):1296–1310.

Laine, O.-M. (2020). Monetary policy and stock market valuation. Research Discussion Papers 16/2020, Bank of Finland.

Lemke, W. and Vladu, A. L. (2017). Below the zero lower bound: a shadow-rate term structure model for the euro area. Working Paper Series 1991, European Central Bank.

Lillo, F., Micciché, S., Tumminello, M., Piilo, J., and Mantegna, R. N. (2015). How news affect the trading behavior of different categories of investors in a financial market. *Quantitative Finance*, 15:213–229.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record*, 82:S2–S18.

Magnus, J. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. PROBABILISTICS AND STATISTICS. Wiley.

Maillard, S., Roncalli, T., and Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70.

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

McFadden, D. et al. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, pages 104–142. Academic Press.

Merton, R. (1972a). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4):1851–1872.

Merton, R. C. (1972b). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4):1851–1872.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42.

Neuenkirch, M. and Nöckel, M. (2018). The risk-taking channel of monetary policy transmission in the euro area. *Journal of Banking & Finance*, 93:71–91.

Oh, D. H. and Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics*, 36(2):181–195.

Opschoor, A., Janus, P., Lucas, A., and Dijk, D. V. (2017). New heavy models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 0(0):1–15.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246 – 256.

Patton, A. J. and Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In *Handbook of financial time series*, pages 801–838. Springer.

Peterson, R. (2016). *Trading on Sentiment: The Power of Minds Over Markets*. John Wiley & Sons, Ltd.

Picault, M. and Renault, T. (2017). Words are not all created equal: A new measure of ecb communication. *Journal of International Money and Finance*, 79:136 – 156.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Rajan, R. G. (2005). Has financial development made the world riskier? Working Paper 11728, National Bureau of Economic Research.

Ranco, G., Aleksovski, D., Caldarelli, G., Grcar, M., and Mozetic, I. (2015). The effects of Twitter sentiment on stock price returns. *PLOS ONE*, 10(9):e0138441.

Ranco, G., Bordino, I., Bormetti, G., Caldarelli, G., Lillo, F., and Treccani, M. (2016). Coupling news sentiment with web browsing data improves prediction of intra-day price dynamics. *PLOS ONE*, 11(1):e0146576.

Rebonato, R. and Jäckel, P. (2011). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Available at SSRN 1969689*.

Rivers, D. and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal*, 5(1):1–39.

Scruggs, J. T. (1998). Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach. *The Journal of Finance*, 53(2):575–603.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, 39(1):119–138.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264.

Smales, L. A. (2014). News sentiment in the gold futures market. *Journal of Banking & Finance*, 49:275 – 286.

Smales, L. A. (2015). Asymmetric volatility response to news sentiment in gold futures. *Journal of International Financial Markets, Institutions and Money*, 34:161–172.

Sun, L., Najand, M., and Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, 73:147 – 164.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, pages 1–17.

Vassallo, D., Bormetti, G., and Lillo, F. (2019). A tale of two sentiment scales: Disentangling short-run and long-run components in multivariate sentiment dynamics. *Available at SSRN 3463691*.

Vassallo, D., Buccheri, G., and Corsi, F. (2021). A dcc-type approach for realized covariance modeling with score-driven dynamics. *International Journal of Forecasting*, 37(2):569–586.

Vassallo, D., Hermans, L., and Kostka, T. (2020). Volatility-targeting strategies and the market sell-off. Financial stability review, box 2, European Central Bank.

Vohra, S. and Teraiya, J. (2013). A comparative study of sentiment analysis techniques. *Journal JIKRCE*, 2(2):313–317.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Wu, L. S.-Y., Pai, J. S., and Hosking, J. (1996). An algorithm for estimating parameters of state-space models. *Statistics & Probability Letters*, 28(2):99–106.

Appendix A

Appendix of Chapter 3

A.1 List of stocks

Table A.1 reports names and sectors of the 27 stocks considered in the empirical analysis.

Tickers	Name	Sector ticker	Sector name
VZ	Verizon	COM	Communication Services
CVX	Chevron	ENE	Energy
AXP	American Express Company	FIN	Financial
GS	Goldman Sachs	FIN	Financial
JPM	JPMorgan Chase	FIN	Financial
JNJ	Johnson & Johnson	HLC	Health Care
MRK	Merck	HLC	Health Care
PFE	Pfizer	HLC	Health Care
UNH	UnitedHealth	HLC	Health Care
BA	Boeing	IND	Industrials
CAT	Caterpillar	IND	Industrials
GE	General Electric	IND	Industrials
MMM	3M Co	IND	Industrials
UTX	United Technologies	IND	Industrials
XOM	XOMA Corp	MAT	Basic Materials
KO	Coca-Cola	NCY	Consumer Goods
PG	Procter & Gamble	NCY	Consumer Goods
AAPL	Apple	TEC	Technology
CSCO	Cisco	TEC	Technology
IBM	IBM	TEC	Technology
INTC	Intel	TEC	Technology
MSFT	Microsoft	TEC	Technology
DIS	Disney	YCY	Consumer Cyclical
HD	Home Depot	YCY	Consumer Cyclical
MCD	McDonalds	YCY	Consumer Cyclical
NKE	Nike	YCY	Consumer Cyclical
WMT	Wal-Mart	YCY	Consumer Cyclical

Table A.1: List of investigated stocks, their ticker, and the economic sector according to the classification of Yahoo Finance.

A.2 Signal-to-noise ratio and comparison with MLNSL

We compare how well the MLSS model fits the data with respect to the MLNSL model using the likelihood ratio test. Since the MLNSL model is nested into the MLSS model, we use the χ^2 distribution to test the null hypothesis (the MLSS model does not fit the data better than the MLNSL) against the alternative hypothesis (the MLSS model fits the data better than the MLNSL). The null hypothesis is rejected with a p-value smaller than 0.01 for both news and social sentiment.

In the last columns of Table 3.1 in the paper we report the signal-to-noise ratio for each asset obtained using the MLSS model and the signal-to-noise ratio obtained using the MLNSL model. The signal-to-noise ratio for the MLSS model, using the same notation of equation (3.6), is evaluated as

$$\mathsf{stn}(i)^{\text{MLSS}} = \frac{\mathsf{Var}\left(\Lambda(i,\cdot)v_t\right) + \mathsf{Var}\left(u_t^i\right)}{\mathsf{Var}\left(\epsilon_t^i\right)} = \frac{\sum_{j=1}^{q}(\Lambda(i,j))^2 + Q_{\text{short}}(i,i)}{R(i,i)} \tag{A.1}$$

while the signal to noise ratio for the MLNSL model, using the notation of equation (3.2), is evaluated as

$$\mathsf{stn}(i)^{\text{MLNSL}} = \frac{\mathsf{Var}\left(v_t^i\right)}{\mathsf{Var}\left(\epsilon_t^i\right)} = \frac{Q(i,i)}{R(i,i)} \tag{A.2}$$

When the MLSS model is estimated, the signal to noise ratio is on average around 0.8 for the news sentiment, while when the MLNSL model is estimated, the signal to noise ratio decreases to an average of 0.03.

Thus our proposed MLSS model has a signal-to-noise ratio approximately twenty times larger than the MLNSL. Our result also points out that the noise in social media is generally higher than the noise in newspapers.

A.3 Quantile regression: Contemporaneous effects

In this appendix we perform the same tests of Section 3.5 in the paper with contemporaneous return and sentiment. In particular, we compute the quantile regression (3.15) with $h = 0$, Table A.2 shows the values of

τ **quantiles**	$R^1(\tau)$ **measure**				
	MLSS	LSS	MLNSL	LNSL	Obs
0.01	16.2%***	6.1%**	1.4%	1.6%	0.6%
0.05	9.2%***	4.0%***	2.8%***	2.7%***	1.7%***
0.10	7.1%***	4.3%***	3.5%***	3.2%***	2.5%***
0.33	2.2%***	1.8%***	1.9%***	1.7%***	1.0%***
0.50	1.1%***	1.1%***	1.2%***	1.0%***	0.7%***
0.66	0.5%***	0.9%***	1.3%***	0.8%***	0.7%***
0.90	1.2%***	1.7%***	1.5%***	0.8%***	0.8%***
0.95	2.9%***	2.3%***	1.9%***	1.0%***	1.0%***
0.99	10.2%***	4.6%**	0.9%	0.6%	1.5%

Table A.2: The R^1 measure across the value τ. We denote with *** the significance at 1%, ** the significance at 5% and * the significance at 10%

the $R^1(\tau)$ measure for different values of τ. It is worth to notice that the quantile regressions are highly significant for every model, except for the 0.01 and 0.99 quantiles, where they are only significant for the MLSS and LSS models. There are three important findings. The first one is that, as in the lagged relation, for any model, the values of R^1 are higher in the tails and lower close to the median. The results are not symmetric around the median. The lower quantiles, which correspond to negative returns, have higher R^1 than the corresponding R^1 in the higher quantiles. This suggests that the sentiment series are powerful explanatory variables in bad times. This conclusion is in accordance with the results in (Garcia, 2013), which shows that investors' sensitivity to news is most pronounced going through hard times. The second result is that the models which exploit the

τ **quantiles**	p**-values**	
	L_t^{ST}	L_t^{LT}
0.01	0.005%	76.313%
0.05	0.000%	2.052%
0.10	0.000%	3.381%
0.33	0.000%	3.257%
0.50	0.000%	7.668%
0.66	1.487%	20.078%
0.90	0.189%	0.309%
0.95	0.007%	0.922%
0.99	0.006%	22.903%

Table A.3: p-values for the Wald test.

multivariate structure (MLSS and MLNSL) produce higher R^1 measures than the corresponding models which apply the cross-sectional averaging procedure on the sentiment series (LSS and LNSL models, respectively). This result confirms that the cross-sectional dependence structure is helpful in extracting a sensible signal. The last result is that the MLSS and LSS models, excluding few values around the median, have higher $R^1(\tau)$ values than other models. This suggests that disentangling the long-term and short-term sentiment components is the most important step to capture the contemporaneous relation with market returns. In particular, the MLSS model, which exploits both the separation in two components and the multivariate structure, strongly outperforms the benchmark model, which solely uses the observed noisy sentiment.

If we look at the contribution of the short and long-term sentiment separately, we again observe similar results with the one observed in Section 3.5 of the paper. Table A.3 reports the p-values of the Wald statistics extracted using a block bootstrap procedure and shows that the short-term sentiment is highly significant at any level of τ, while the long-term sentiment has lower p-values. In particular, the short-term sentiment, which

captures rapidly changing trends, is significant for extreme returns ($\tau = 0.01$ or $\tau = 0.99$) while the long-term sentiment is not. This result suggests that extreme market swings can be explained by unexpected and short-lasting news. Moreover, it further supports the importance of disentangling sentiment components which are sensitive to different time scales.

These findings show very strong contemporaneous relation between sentiment and market returns. We look at these results as a sanity check of our approach. Indeed, since we are not claiming that sentiment causes returns or viceversa, then it is reasonable to expect a significant contemporaneous relation at daily time scale. The sentiment explains returns and this could be due to the fact that the news, from which sentiment is computed, report and comment about the market performance. What is more promising is that the R^1 measure increases with the complexity of the model, and this is especially true for extreme market events – where the observed sentiment is not significant. Then, we conclude that an essential ingredient of the analysis is the combination of a multivariate model with the separation of sentiment in two components, the stochastic long-run trend (long-term sentiment) common to all assets and a fast changing and asset-specific trend (short-term sentiment).

A.4 Market absortion of news

The 1-lag and the h-lag sentiment series contain useful information to explain future returns. The market does not immediately digest all the news in the newspapers and social media but it takes few days to do it. It is worth to investigate whether, at the same time, the opposite relation may hold. In particular, do media and social networks immediately digest re-

turns that occurred in the previous days? Is this reaction dependent on the sign of returns? In order to answer these questions, we define the non-causal quantile regression

$$\hat{V}^{h,\mathrm{MLSS}}(\tau) = \min_{(\alpha^0,\alpha^1\beta^1\in\mathbb{R}^6)} \sum_{t=1}^{T} \rho_\tau \left(r_t^m - \alpha^0 - \beta^1 S_{t-h}^{\mathrm{MLSS}} \right), \tag{A.3}$$

where h can assume positive and negative values. From the $R^1(\tau)$ of this model, we can unravel the interplay among professional and social media news, and returns in the financial market. For $h > 0$, we measure the impact of the news on future returns, while for $h < 0$, we assess how fast past returns are digested in newspapers and social media. When $h = 0$, we evaluate to the immediate impact of the news on daily returns. The two cases $h = 0$ and $h = 1$ have been studied in the previous sections. In figure A.1, we show the evolution of $R^1(\tau)$ compared to the evolution of $R^1(1-\tau)$. In this manner, we observe how the market predicts or digests returns of comparable absolute value but opposite sign. We shrink to zero all values of $R^1(\tau)$ which are not significant with a p-value smaller than 0.05. For $h \geqslant 0$, the R^1 measures are statistically significantly and slightly increase as they approach $h = 1$. At this stage, the behavior of $R^1(\tau)$ is not different for values of $\tau < 1/2$ and $\tau > 1/2$. However, for $h \leq 0$, when a return appears, the response to a positive or a negative value is very different. For $h = 0$, $R^1(\tau)$ increases when the return is negative, confirming that, as reported in Table A.2, the current news have an higher explanatory power on the negative rather then positive returns. However, as mentioned previously, it is not possible to assess the causal relation among news and returns. Then, the $R^1(\tau)$ measure reaches its maximum when $h = -1$, suggesting that newspapers and social media keep talking about the previous day market performance. This is especially true when

the previous day return is negative. The impact of negative returns on sentiment slowly decreases but remains significant. On the contrary, when the return is positive, the effect is milder. This general picture is stronger for extreme quantile levels, e.g. $\tau = 0.01$ and 0.05, and becomes negligible when we move to the median region of returns.

Summarizing, positive returns are rapidly digested from newspaper and social media, while the echo of negative market performances persist for few days. As a final test of reliability of our approach, we perform the

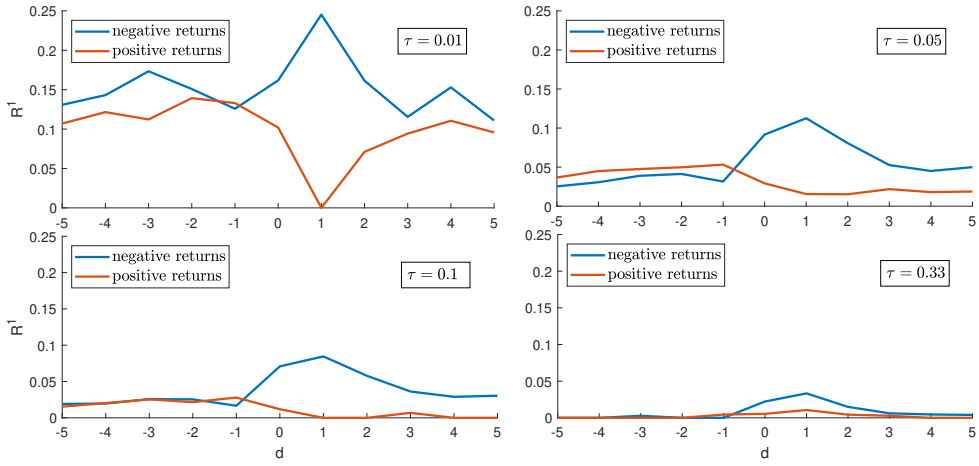

Figure A.1: R^1 measure of model (A.3) for different h and τ. Top left panel: $R^1(0.99)$ in orange and $R^1(0.01)$ in blue. Top right panel: $R^1(0.95)$ in orange and $R^1(0.05)$ in blue. Bottom left panel: $R^1(0.9)$ in orange and $R^1(0.1)$ in blue. Bottom right panel: $R^1(0.66)$ in orange and $R^1(0.33)$ in blue. The x-axis labeling in terms of $d = -h$ allows to have past sentiment on the left and future sentiment on the right.

same analysis using the noisy sentiment S^{Obs} instead of the filtered signal S^{MLSS}. Figure A.2 reports the results. No clear patterns arise and the h-lag observed sentiment has no statistical significance in explaining contemporaneous and future returns. Again, the proper filtering of sentiment time series appears essential to extract some meaningful information from the

mood of the market.

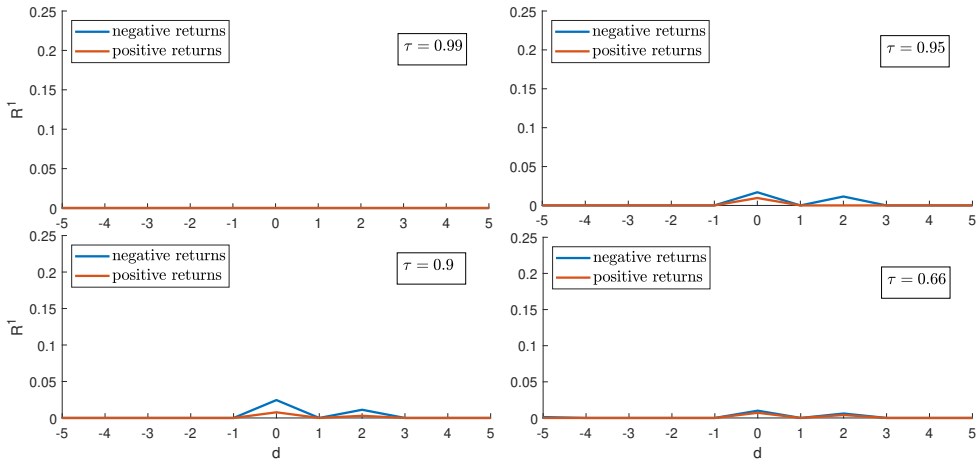

Figure A.2: Same analyses as in Figure A.1 but the filtered sentiment signal S^{MLSS} is replaced by the noisy sentiment S^{Obs}.

A.5 Portfolio allocation on February 2007-June 2017 without trading costs

Table A.4 reports the performances of the five sentiment strategies together with the buy-and-hold portfolio without trading costs. We notice that the qualitative results do not change. Given that the MLSS portfolio produces the higher number of trades, the performance gap with respect to the other strategies increase in size in terms of returns, Sharpe and Sortino ratios. Figure A.3 shows the evolution of the sentiment-based portfolios without trading costs. The MLSS portfolio, contrary to all the other portfolios, performs very well during the financial crisis and strongly outperforms the other sentiment-based portfolios and the $1/n$ portfolio. However, the gain reduces during the $2009 - 2017$ bull market period. Nonetheless, even if the absolute return reduces, the volatility is consistently lower during the whole period and the Sharpe and Sortino ratios are respectively 40% and 49% higher in the MLSS portfolio rather than the buy-and-hold portfolio.

Measures	BH	MLSS	LSS	MLNSL	LNSL	Obs
A. return (%)	8.972	**9.393**	7.650	9.091	8.308	7.33
A. volatility (%)	19.122	**14.080**	17.996	17.797	19.113	19.765
A. neg. volatility (%)	15.514	**10.932**	14.294	14.368	15.125	15.928
A. Sharpe ratio	0.469	**0.667**	0.425	0.511	0.435	0.371
A. Sortino ratio	0.578	**0.859**	0.535	0.633	0.549	0.46
MDD ($)	59377	54938	50182	**49397**	61921	61982

Table A.4: Performances of the six strategies without trading cost for the period February 2007 - June 2017. In bold, the best performance per row. BH is the buy-and-hold portfolio, while MLSS, LSS, MLNSL, LSNSL, and Obs correspond to portfolios built from the corresponding model for the sentiment time series.

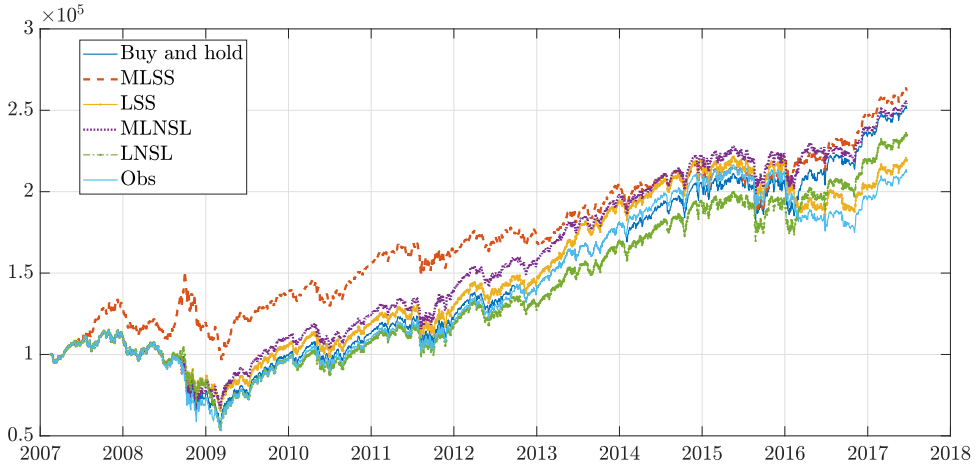

Figure A.3: Portfolio evolution of the sentiment based strategies built using equation (3.19) together with the buy-and-hold equally weighted portfolio in blue.

A.6 Robustness check: Portfolio allocation on July 2017 - December 2019

In this appendix, we use the parameter values estimated from the TRMI sentiment time series over the February 2007 - June 2017 to filter the sentiment signal in the July 2017 - December 2019 period. This procedure ensures that the filtered signals do not suffer from any forward-looking bias.

Table A.5 shows that the qualitative results do not change from the Section 3.6. The MLSS model outperforms the buy-and-hold portfolio with a relative gain of 14% in both Sharpe and Sortino ratio. Two main differences are visible from the February 2007 - June 2017 period. The LSS model slightly outperforms the buy-and-hold portfolio and it is the second best performing model, while in the previous case the second best performing model was the MLNSL. The Obs portfolio produces the same performance

Measures	BH	MLSS	LSS	MLNSL	LNSL	Obs
A. return (%)	13.56	**15.28**	14.316	12.584	9.72	13.56
A. volatility (%)	14.477	**14.311**	14.257	14.672	15.053	14.477
A. neg. volatility (%)	10.673	10.552	**10.489**	10.841	11.521	10.673
A. Sharpe ratio	0.937	**1.068**	1.004	0.858	0.646	0.937
A. Sortino ratio	1.27	**1.448**	1.365	1.161	0.844	1.27
MDD ($)	23489	**18871**	19631	26639	31207	23489

Table A.5: Performances of the six strategies without trading cost for the period July 2017 - December 2019. In bold, the best performance per row. BH is the buy-and-hold portfolio, while MLSS, LSS, MLNSL, LSNSL, and Obs correspond to portfolios built from the corresponding model for the sentiment time series.

Measures	BH	MLSS	LSS	MLNSL	LNSL	Obs
A. return (%)	13.566	**14.571**	13.907	12.518	9.492	13.566
A. volatility (%)	14.483	14.394	**14.328**	14.7	15.122	14.483
A. neg. volatility (%)	10.678	10.641	**10.564**	10.865	11.589	10.678
A. Sharpe ratio	0.937	**1.012**	0.971	0.852	0.628	0.937
A. Sortino ratio	1.27	**1.369**	1.317	1.152	0.819	1.27
MDD ($)	23489	**19319**	20330	26868	31898	23489
Number of trades	1	41	25	5	13	1
Transaction costs ($)	50	2460	1421	279	741	50

Table A.6: Performances of the six strategies with trading cost for the period July 2017 - December 2019. In bold, the best performance per row. BH is the buy-and-hold portfolio, while MLSS, LSS, MLNSL, LSNSL, and Obs correspond to portfolios built from the corresponding model for the sentiment time series.

of the buy-and-hold portfolio and the reason is that the selling signal from s^{Obs} is always negative. Then, the number of transaction is equal to 1. In table A.6, we see that the transaction costs do not change the qualitative results and again, the MLSS strategy is the one which produces the higher number of trades and, as a consequence, the higher transaction costs. As done in the main text, the performance of the MLSS strategy from table A.6 can be further improved by applying the penalization of the selling signal based on the Mc Fadden's R^2.

Appendix B

Appendix of Chapter 4

B.1 Statistical significance of the sentiment portfolios

Here, we assess the significance of the trades produced by the strategy (3.18) for the different sentiment filters. We design a Monte Carlo experiment where a trader follow a random selling signal. The selling signal is given by $s_t^{\text{shuffled, mod}}$, which is nothing more than a shuffled realization of s_t^{mod}. The number of random selling signals corresponds by construction to the number of selling signals produced by s^{mod}, which is reported in table 3.6 of the paper. We repeat the experiment $10,000$ times. The corresponding portfolios are then sorted according to their Sharpe and Sortino ratios and the p-value of each strategy is computed by comparison with the quantiles from the Monte Carlo experiment. Table B.1 shows the results over the period February 2007 - June 2017. The MLSS strategy significantly outperforms the random strategy with a p-value smaller than 5%. All the other strategies are not statistically different from a random strategy.

Strategies	Annual Sharpe ratio	Annual Sortino ratio
MLSS	0.519(96.9%)	0.725(96.4%)
best 5%	0.485	0.69
best 10%	0.431	0.611
best 25%	0.349	0.491
median	0.26	0.364
LSS	0.385(36.7%)	0.542(37.4%)
best 5%	0.529	0.749
best 10%	0.502	0.709
best 25%	0.457	0.644
median	0.409	0.574
MLNSL	0.495(86.2%)	0.7(86.9%)
best 5%	0.523	0.739
best 10%	0.504	0.711
best 25%	0.474	0.667
median	0.443	0.622
LNSL	0.419(27.0%)	0.59(27.6%)
best 5%	0.524	0.741
best 10%	0.505	0.713
best 25%	0.475	0.669
median	0.446	0.627

Table B.1: Performances of the sentiment strategies compared with the 95%, 90%, 75% and 50% quantiles from the shuffled strategy. Values in brackets are the percentages of randomly generated portfolios which perform worse than the sentiment-based strategy for the period February 2007 - June 2017.

The p-values of the MLSS trading strategy are even lower when the R^2-penalized trading strategy (3.20) is implemented. Table B.2 shows the p-values. When the number (100%) is reported, all the $10,000$ random strategies perform worse than the MLSS α strategy. The number of selling signals for the MLSS with $\alpha = 0.80$ is too small and the result may be not reliable.

Strategies	Annual Sharpe ratio	Annual Sortino ratio
MLSS(00)	0.519(96.9%)	0.725(96.4%)
best 5%	0.485	0.69
best 10%	0.431	0.611
best 25%	0.349	0.491
median	0.26	0.364
MLSS(20)	0.525(96.5%)	0.735(96.1%)
best 5%	0.502	0.712
best 10%	0.457	0.648
best 25%	0.383	0.539
median	0.306	0.428
MLSS(35)	0.68(99.9%)	0.967(99.8%)
best 5%	0.512	0.727
best 10%	0.473	0.67
best 25%	0.409	0.576
median	0.339	0.475
MLSS(50)	0.723(100.0%)	1.03(100.0%)
best 5%	0.531	0.752
best 10%	0.495	0.7
best 25%	0.436	0.615
median	0.373	0.523
MLSS(65)	0.757(100.0%)	1.09(100.0%)
best 5%	0.534	0.757
best 10%	0.504	0.711
best 25%	0.459	0.647
median	0.415	0.583
MLSS(80)	0.534(97.3%)	0.752(97.2%)
best 5%	0.519	0.733
best 10%	0.501	0.707
best 25%	0.477	0.671
median	0.45	0.634

Table B.2: Performances of the R^2 adjusted MLSS strategies for different values of α compared with the 95%, 90%, 75% and 50% quantiles from the random strategy for the period February 2007 - June 2017. The sentiment strategies are referred to as MLSS($\alpha \times 100$). Values in brackets are the percentages of randomly generated portfolios which perform worse than the sentiment-based strategy.

B.2 Computation of the scaled score in the univariate models

χ^2 density

We compute the scaled score appearing in the dynamic of the log-variance $\lambda_t^{(i)}$ in Eq. (4.13) in the case of the χ^2 density. To simplify the notation, we suppress the subscript i. The conditional log-likelihood is:

$$\log p_{W_1}(x_t; v_t, \nu) = \frac{1}{2}c(\nu) + \left(\frac{\nu}{2} - 1\right)\log(x_t) - \frac{\nu}{2}\log(v_t) - \frac{\nu}{2}\left(\frac{x_t}{v_t}\right) \quad \text{(B.1)}$$

where $c(\nu) = \nu\log(\nu/2) - 2\log\Gamma(\nu/2)$. We now prove the following result (recall that $v_t = e^{\lambda_t}$):

Proposition 7. *For the density in Eq. (B.1), the score* $\nabla_t^{W_1} = \frac{\partial \log p_{W_1}(x_t;\lambda_t,\nu)}{\partial \lambda_t}$ *is given by:*

$$\nabla_t^{W_1} = \frac{\nu}{2e^{\lambda_t}}\left[x_t - e^{\lambda_t}\right] \quad \text{(B.2)}$$

Proof.

$$\begin{aligned}\nabla_t^{W_1} &= \frac{\partial \log p_{W_1}(x_t;\lambda_t,\nu)}{\partial \lambda_t} = \frac{\partial \log p_{W_1}(x_t;\lambda_t,\nu)}{\partial v_t} \times \frac{\partial v_t}{\partial \lambda_t} \\ &= \frac{\nu}{2v_t^2}\left[x_t - v_t\right] \times v_t \\ &= \frac{\nu}{2e^{\lambda_t}}\left[x_t - e^{\lambda_t}\right]\end{aligned}$$

□

Then, we compute the information quantity:

Proposition 8. *For the density in Eq. (B.1), the Fisher information* $\mathcal{I}_{t|t-1}^{W_1} = \mathrm{E}_{t|t-1}[\nabla_t^2]$ *is given by:*

$$\mathcal{I}_{t|t-1}^{W_1} = \frac{\nu}{2} \quad \text{(B.3)}$$

Proof.

$$\mathcal{I}_{t|t-1}^{W_1} = E_{t-1}\left[(\nabla_t^{W_1})^2\right] = E_{t-1}\left[\frac{\nu^2}{4e^{2\lambda_t}}(x_t - e^{\lambda_t})^2\right]$$
$$= \frac{\nu^2}{4e^{2\lambda_t}} Var\left[x_t\right] =^* \frac{\nu^2}{4e^{2\lambda_t}} \frac{2e^{2\lambda_t}}{\nu} = \frac{\nu}{2}$$

where (*) follows from $Var\left[x\right] = Var\left[(v_t/\nu)k_\nu\right] = 2\frac{(v_t)^2}{\nu}$ if k_ν is distributed as a χ^2_ν. □

Finally, it is immediate to compute the scaled score:

Proposition 9. *For the density in Eq. (B.1), the scaled score* $s_t^{W_1} = (\mathcal{I}_{t|t-1}^{W_1})^{-1}\nabla_t^{W_1}$ *is given by:*

$$s_t^{W_1} = \frac{1}{e^{\lambda_t}}\left[x_t - e^{\lambda_t}\right] \tag{B.4}$$

F density

In the case of the univariate F density, the conditional log-likelihood is:

$$\log p_F\left(x_t; v_t, \nu_1, \nu_2\right) = d(\nu_1, \nu_2) - \frac{\nu_1}{2}\log(v_t) + \left(\frac{\nu_1}{2} - 1\right)\log(x_t) - \frac{\nu_1 + \nu_2}{2}\log\left(\tilde{w}_t\right) \tag{B.5}$$

where:

$$\tilde{w}_t = 1 + \frac{\nu_1 x_t}{(\nu_2 - 2)v_t} \tag{B.6}$$

$$d(\nu_1, \nu_2) = \frac{\nu_1}{2}\log\left(\frac{\nu_1}{\nu_2 - 2}\right) + \log\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right) - \log\Gamma\left(\frac{\nu_1}{2}\right) - \log\Gamma\left(\frac{\nu_2}{2}\right) \tag{B.7}$$

We compute now the score of the conditional log-likelihood.

Proposition 10. *For the density in Eq. (B.5), the score* $\nabla_t^F = \frac{\partial \log p_F(x_t; v_t, \nu_1, \nu_2)}{\partial \lambda_t}$ *is given by:*

$$\nabla_t^F = \frac{\nu_1}{2e^{\lambda_t}}\left[\frac{\nu_1 + \nu_2}{\nu_2 - 2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right] \tag{B.8}$$

Proof.

$$\begin{aligned}\nabla_t^F &= \frac{\partial \log p_F\left(x_t; v_t, \nu_1, \nu_2\right)}{\partial \lambda_t} = \frac{\partial \log p_F\left(x_t; v_t, \nu_1, \nu_2\right)}{\partial v_t} \times \frac{\partial v_t}{\partial \lambda_t} \\ &= -\frac{\nu_1}{2v_t} + \frac{\nu_1+\nu_2}{2}\left[\tilde{w}_t \frac{\nu_1 x_t}{(\nu_2-2)(v_t)^2}\right] \times v_t \\ &= \frac{\nu_1}{2e^{2\lambda_t}}\left[\frac{\nu_1+\nu_2}{\nu_2-2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right] \times e^{\lambda_t} \\ &= \frac{\nu_1}{2e^{\lambda_t}}\left[\frac{\nu_1+\nu_2}{\nu_2-2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right]\end{aligned}$$

□

As in the correlation model (cf. Section 4.2), we scale the score by the inverse of the Fisher information of the χ^2 density. Thus, we get:

Proposition 11. *For the density in Eq. (B.5), the scaled score* $s_t^F = (\mathcal{I}_{t|t-1}^{W_1})^{-1}\nabla_t^F$ *is given by:*

$$s_t^F = \frac{1}{e^{\lambda_t}}\left[\frac{\nu_1+\nu_2}{\nu_2-2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right] \tag{B.9}$$

B.3 Proposition 1

Proof. We need to compute $\frac{\partial l(X_t)}{\partial f_t'} = \frac{\partial l(X_t)}{\partial \mathsf{vech}(Q_t)'}$, where l is given by Eq. (4.17)

$$l\left(X_t\right) = \frac{1}{2}d_X(k,\nu) + \frac{\nu-k-1}{2}\log|X_t| - \frac{\nu}{2}\log|V_t| - \frac{\nu}{2}tr\left(V_t^{-1}X_t\right).$$

Thanks to the chain rule, we can split our equation as:

$$\frac{\partial l\left(X_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} = \frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'}\frac{\partial \mathsf{vec}\left(V_t\right)}{\partial \mathsf{vec}\left(R_t\right)'}\frac{\partial \mathsf{vec}\left(R_t\right)}{\partial \mathsf{vec}\left(Q_t\right)'}\frac{\partial \mathsf{vec}\left(Q_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'}.$$

Then, starting with the first term and considering $d\log|X| = tr\left(X^{-1}\right)dX$ and $d\left(X^{-1}\right) = -X^{-1}\left(dX\right)X^{-1}$, see Magnus and Neudecker (1999),

$$\begin{aligned}\frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'} &= -\frac{\nu}{2}\left[\mathsf{vec}\left(V_t^{-1}\right)' - \mathsf{vec}\left(X_t\right)'\left(V_t^{-1}\otimes V_t^{-1}\right)\right] \\ &= \frac{\nu}{2}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]'\left(V_t^{-1}\otimes V_t^{-1}\right).\end{aligned}$$

The second term is, thanks to the fact that $\mathsf{vec}\left(AXB\right) = \left(B'\otimes A\right)\mathsf{vec}\left(X\right)$

$$\frac{\partial \mathsf{vec}\left(V_t\right)}{\partial \mathsf{vec}\left(R_t\right)'} = \frac{\partial \mathsf{vec}\left(D_tR_tD_t\right)}{\partial \mathsf{vec}\left(R_t\right)'} = \left(D_t\otimes D_t\right).$$

By definition of duplication matrix, see Abadir and Magnus (2005), we have that

$$\frac{\partial \mathsf{vec}\left(Q_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} = \mathcal{D}_k.$$

The third term is a little bit more complicated, indeed defining $\Delta_t = \left(diag(Q_t)^{1/2}\right)$,

$$\begin{aligned}d\mathsf{vec}\left(R_t\right) =& d\mathsf{vec}\left(\Delta_t^{-1}Q_t\Delta_t^{-1}\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) + \mathsf{vec}\left(d\left(\Delta_t^{-1}\right)Q_t\Delta_t^{-1}\right) + \mathsf{vec}\left(\Delta_t^{-1}Q_t d\left(\Delta_t^{-1}\right)\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) + \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]d\mathsf{vec}\left(\Delta_t^{-1}\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) - \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]\Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(\Delta_t\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) - \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]\Delta_t^{-1}\otimes\Delta_t^{-1}W_Q d\mathsf{vec}\left(Q_t\right),\end{aligned}$$

where $q_t = \mathsf{vec}\left(\Delta_t\right)$ and W_Q is a diagonal matrix with its ith diagonal elements equal to $1/2\sqrt{q_t^{(i)}}$ if $q_t^{(i)} \neq 0$ and zero otherwise.

Given these four results combined with the fact that $\left(D_t\otimes D_t\right)$ and $\left(V_t^{-1}\otimes V_t^{-1}\right)$ are symmetric , we get

$$\begin{aligned}\frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'} =& \frac{\nu}{2}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]'\left(V_t^{-1}\otimes V_t^{-1}\right)\left(D_t\otimes D_t\right)\left(\Delta_t^{-1}\otimes\Delta_t^{-1}\right) \\ & \times\left[I - \left(Q_t\Delta_t^{-1}\otimes I\right) + \left(I\otimes Q_t\Delta_t^{-1}\right)\right]W_Q\mathcal{D}_k \\ =& \frac{\nu}{2}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]'\left(D_t^{-1}\Delta_tQ_t^{-1}\otimes D_t^{-1}\Delta_tQ_t^{-1}\right)\Psi_t\mathcal{D}_k,\end{aligned}$$

where $\Psi_t = \left[I - \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]W_Q\right]$. □

B.4 Proposition 2

Proof. Starting with the definition of Fisher information matrix:

$$
\begin{aligned}
E\left[\nabla_t^W \nabla_t^{W'} | \mathcal{F}_{t-1}\right] =& E\big[\frac{\nu^2}{4}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right)-\mathsf{vec}\left(V_t\right)\right] \\
&\times\left[\mathsf{vec}(X_t)-\mathsf{vec}(V_t)\right]'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\Psi_t\mathcal{D}_k|\mathcal{F}_{t-1}\big] \\
=& \frac{\nu^2}{4}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\times \\
&\times Var\left(\mathsf{vec}(X_t)-\mathsf{vec}(V_t)|\mathcal{F}_{t-1}\right)\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\Psi_t\mathcal{D}_k \\
=^{*}& \frac{\nu^2}{4}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\frac{1}{\nu}\left(I_{k^2}+K_k\right)\times \\
&\times\left(V_t\otimes V_t\right)\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\Psi_t\mathcal{D}_k \\
=^{**}& \frac{\nu}{2}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\right)\mathcal{D}_k\mathcal{D}_k^{+}\Psi_t\mathcal{D}_k
\end{aligned}
$$

where * is thank to the vech formulation of the Wishart variance in Abadir and Magnus (2005) and ** is given by $2\mathcal{D}_k\mathcal{D}_k^{+}=\left(I_k+K_k\right)$ (recall that $V_t=\left(D_t\Delta_t^{-1}Q_t\Delta_t^{-1}D_t\right)$)

$$
\begin{aligned}
A &= \left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left(V_t\otimes V_t\right)\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right) \\
&= \left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left(D_t\Delta_t^{-1}Q_t\Delta_t^{-1}D_tD_t^{-1}\Delta_t Q_t^{-1}\otimes D_t\Delta_t^{-1}Q_t\Delta_t^{-1}D_tD_t^{-1}\Delta_t Q_t^{-1}\right) \\
&= \left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left(D_t\Delta_t^{-1}\otimes D_t\Delta_t^{-1}\right) \\
&= \left(D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\right) \\
&= \left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)
\end{aligned}
$$

□

B.5 Proposition 3

Proof. Multiplying the inverse of 4.19 with the 4.18 we get

$$\begin{aligned}
s_t^W =& \left(\mathcal{D}_k{}'\Psi_t' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\times \left(H_t^{-1}Q_t^{-1} \otimes H_t^{-1}Q_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \left(\mathcal{D}_k{}'\Psi_t' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\times \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\left(H_t^{-1} \otimes H_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \left(\mathcal{D}_k{}'\Psi_t' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\times \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\left(H_t^{-1} \otimes H_t^{-1}\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \left(\mathcal{D}_k{}'\Psi_t' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\times \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\left(H_t^{-1} \otimes H_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]
\end{aligned}$$

Now let us simplify $\mathcal{I}_t^W = \frac{\nu}{2}\mathcal{D}_k'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k$. This simplification is not restrictive because the matrix Ψ_t is very sparse, moreover using the approximation $\Psi_t = I$ allows us to define the inverse of $\mathcal{I}_t$ which is not full-rank with the original representation.
We get

$$\begin{aligned}
s_t =& \mathcal{D}_k{}^{+}\left(H_t^{-1} \otimes H_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \mathcal{D}_k{}^{+}\left(H_t^{-1} \otimes H_t^{-1}\right)\mathsf{vec}(X_t) - \mathcal{D}_k{}^{+}\left(H_t^{-1} \otimes H_t^{-1}\right)\left(H_t \otimes H_t\right)\mathsf{vec}(Q_t) \\
=& \mathcal{D}_k{}^{+}\left(H_t^{-1} \otimes H_t^{-1}\right)\mathsf{vec}(X_t) - \mathsf{vech}\left(Q_t\right)
\end{aligned}$$

□

B.6 Proposition 4

Proof. As in the Wishart case, we can split

$$\frac{\partial l\left(X_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} = \frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'}\frac{\partial \mathsf{vec}\left(V_t\right)}{\partial \mathsf{vec}\left(R_t\right)'}\frac{\partial \mathsf{vec}\left(R_t\right)}{\partial \mathsf{vec}\left(Q_t\right)'}\frac{\partial \mathsf{vec}\left(Q_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'}.$$

The only terms which is different from the previous model is the first one. The log-likelihood function is and get, using again $d\log|X| = tr\left(X^{-1}\right)dX$ and $d\left(X^{-1}\right) = -X^{-1}\left(dX\right)X^{-1}$

$$\begin{aligned}
dl_{X_t} =& -\frac{\nu_1}{2}tr\left(V_t^{-1}dV_t\right) - \frac{\nu_1+\nu_2}{2}tr\left(\tilde{W}_t^{-1}d\tilde{W}_t\right) \\
=^* & -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)'d\mathsf{vec}V_t + \frac{\nu_1+\nu_2}{2}tr\left(\tilde{W}_t^{-1}\frac{\nu_1}{\nu_2-k-1}V_t^{-1}dV_tV_t^{-1}X_t\right) \\
=& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)'d\mathsf{vec}V_t + \frac{\nu_1+\nu_2}{2}tr\left(\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\tilde{W}_t^{-1}V_t^{-1}dV_t\right) \\
=& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)'d\mathsf{vec}V_t + \frac{\nu_1+\nu_2}{2}\mathsf{vec}\left(\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\tilde{W}_t^{-1}V_t^{-1}\right)'d\mathsf{vec}V_t
\end{aligned}$$

hence we obtain

$$\begin{aligned}
\frac{\partial l\left(X_t\right)}{\partial\mathsf{vec}\left(V_t\right)'} =& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)' + \frac{\nu_1+\nu_2}{2}\mathsf{vec}\left(\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\tilde{W}_t^{-1}V_t^{-1}\right)' \\
=& \frac{\nu_1}{2}\left[\left(V_t^{-1}\otimes V_t^{-1}\right)\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right)\right]'.
\end{aligned}$$

Combining all the formulas together we get

$$\begin{aligned}
\frac{\partial l\left(X_t\right)}{\partial\mathsf{vech}\left(Q_t\right)'} =& \frac{\nu_1}{2}\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right)'\left(V_t^{-1}\otimes V_t^{-1}\right)\left(D_t\otimes D_t\right) \\
& \times\left(\Delta_t^{-1}\otimes\Delta_t^{-1}\right)\Psi_t\mathcal{D}_k \\
=& \frac{\nu_1}{2}\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right)' \\
& \times\left(D_t^{-1}\Delta_tQ_t^{-1}\Delta_tD_t^{-1}\otimes D_t^{-1}\Delta_tQ_t^{-1}\Delta_tD_t^{-1}\right)\left(D_t\Delta_t^{-1}\otimes D_t\Delta_t^{-1}\right)\Psi_t'\mathcal{D}_k' \\
=& \frac{\nu_1}{2}\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right)'\left(H_t^{-1}Q_t^{-1}\otimes H_t^{-1}Q_t^{-1}\right)
\end{aligned}$$

□

B.7 Proposition 5

Proof. Consider the multiplication $\mathcal{I}_t^W \nabla_t^F$, considering the same approximation of $\mathcal{I}_t^W$ we used in B.5,

$$\begin{aligned}
s_t^F =& \left(\mathcal{D}_k{}' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\right)^{-1}\mathcal{D}_k' \\
&\times \left(H_t^{-1}Q_t^{-1} \otimes H_t^{-1}Q_t^{-1}\right)\left[\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \left(\mathcal{D}_k{}' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\right)^{-1}\mathcal{D}_k' \left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right) \\
&\times \mathcal{D}_k\mathcal{D}_k{}^+ \left(H_t^{-1} \otimes H_t^{-1}\right)\left[\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vech}\left(H_t^{-1}X_t\tilde{W}_t^{-1}H_t^{-1}\right) - \mathsf{vech}\left(Q_t\right)
\end{aligned}$$

□

B.8 Proposition 6

Proof. The proof is straightforward:

$$\begin{aligned}
\lim_{\nu_2\to\infty} s_t^F =& \lim_{\nu_2\to\infty} \frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vech}\left(H_t^{-1}X_t\tilde{W}_t^{-1}H_t^{-1}\right) - \mathsf{vech}\left(Q_t\right) \\
=& \mathsf{vech}\left(H_t^{-1}X_t\tilde{W}_t^{-1}H_t^{-1}\right) - \mathsf{vech}\left(V_t\right) = s_t^W
\end{aligned}$$

since $\tilde{W}_t = I_k + \frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t$ □

B.9 Figures of Section (4.3)

In this section, we report the figures of the experiment based on the misspecified DGP's presented in Section (4.3)

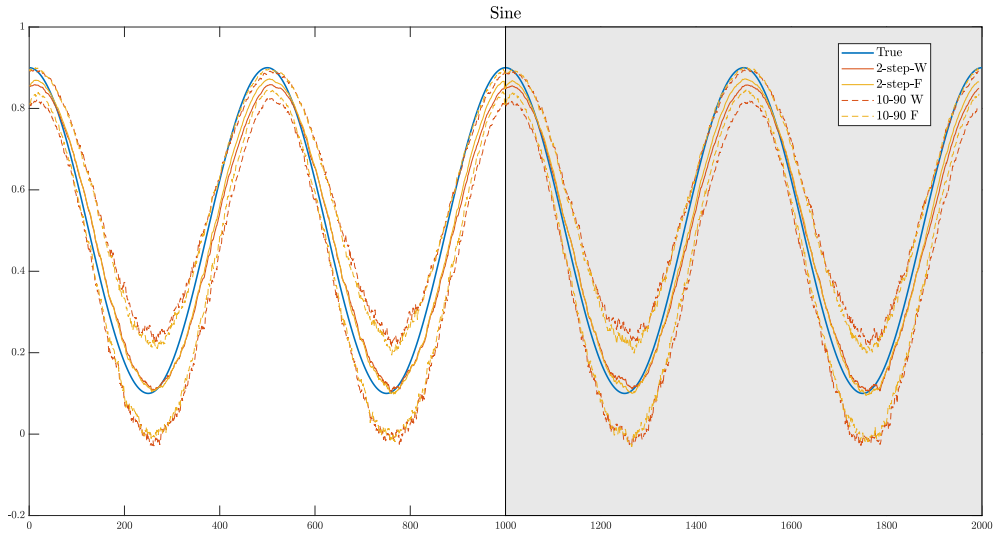

Figure B.1: In-sample and out-of-sample filtered estimates of $\rho_t^{(1)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

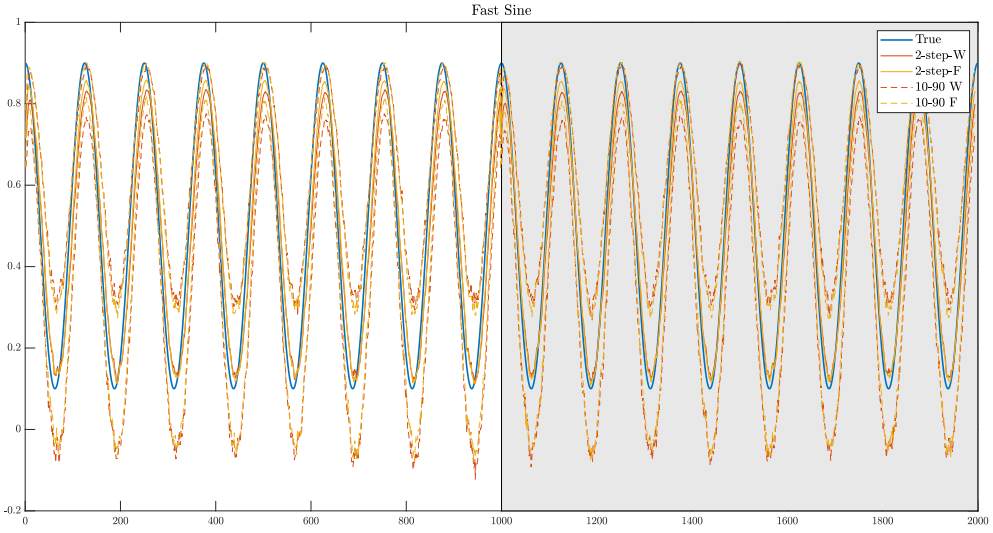

Figure B.2: In-sample and out-of-sample filtered estimates of $\rho_t^{(2)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

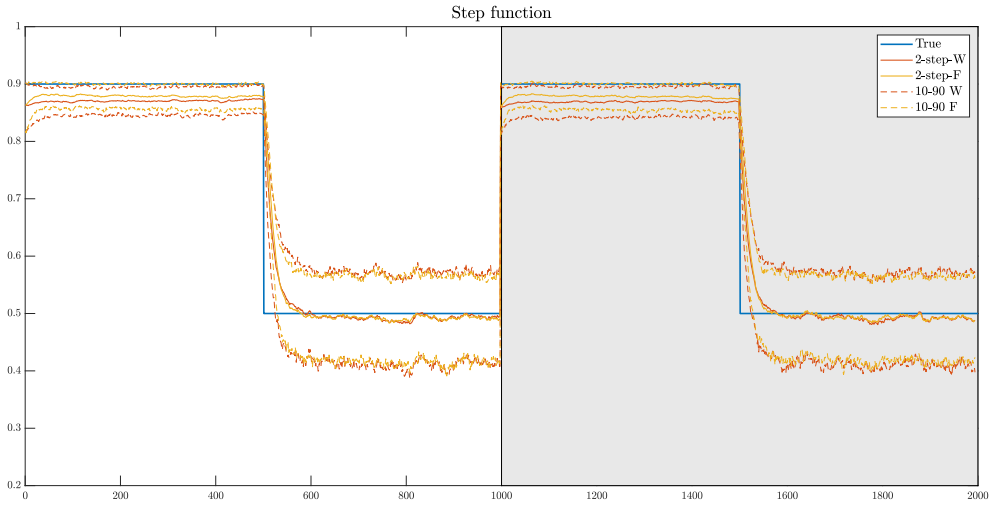

Figure B.3: In-sample and out-of-sample filtered estimates of $\rho_t^{(3)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

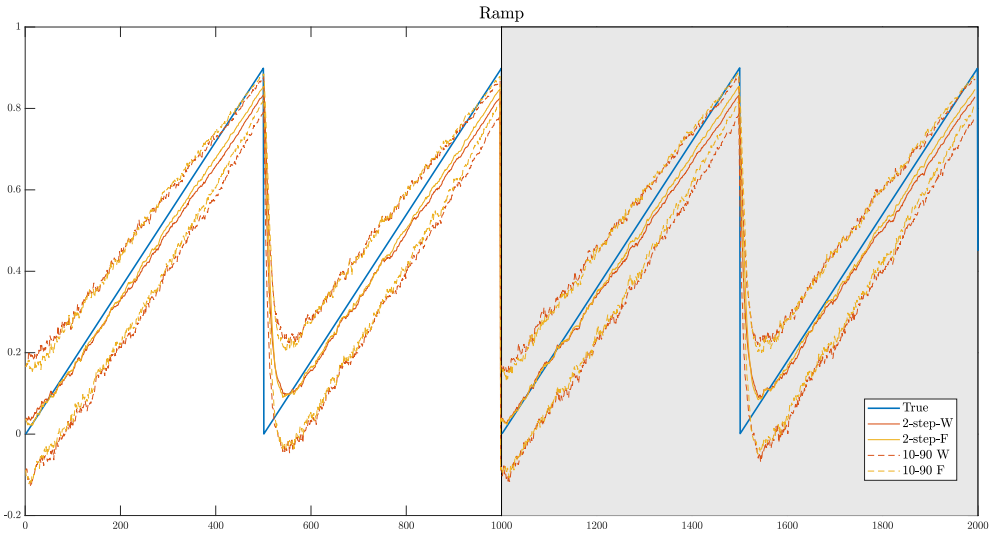

Figure B.4: In-sample and out-of-sample filtered estimates of $\rho_t^{(4)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

Appendix C

Appendix of Chapter 5

C.1 Proof Propositions 1 and 2

The proof of proposition 1 is trivial. Indeed the value of the total derivative is given by

$$\frac{\mathrm{d}\omega^1}{\mathrm{d}r^f} = \frac{\gamma\psi\sigma^2 - \lambda\gamma(r - r^f)}{(\sigma^2)^2} \tag{C.1}$$

Using equation (5.6) we get that

$$\frac{\mathrm{d}\omega^1}{\mathrm{d}r^f} = \frac{\gamma\psi - \omega^1}{(\sigma^2)} < 0 \iff \omega^1 > \frac{\gamma\psi}{\lambda} \tag{C.2}$$

which concludes the proof.

For Proposition 2, using the information set $\theta = \left[r - r^f, \sigma^2, \gamma\right]$ for the investor problem (5.6), we analytically evaluate the Portfolio Instability measure defined in (5.3). The Jacobian matrix reduces to

$$\begin{aligned} D\omega(\theta) &= \begin{bmatrix} \frac{\partial\omega^1}{\partial(r-r^f)} & \frac{\partial\omega^1}{\partial\sigma^2} & \frac{\partial\omega^1}{\partial\gamma} \\ \frac{\partial\omega^2}{\partial(r-r^f)} & \frac{\partial\omega^2}{\partial\sigma^2} & \frac{\partial\omega^2}{\partial\gamma} \end{bmatrix} = \begin{bmatrix} \frac{\partial\omega^1}{\partial(r-r^f)} & \frac{\partial\omega^1}{\partial\sigma^2} & \frac{\partial\omega^1}{\partial\gamma} \\ -\frac{\partial\omega^1}{\partial(r-r^f)} & -\frac{\partial\omega^1}{\partial\sigma^2} & -\frac{\partial\omega^1}{\partial\gamma} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\gamma}{\sigma^2}, & \frac{\gamma\mu_M\sigma^2 - \gamma(r-r^f)}{(\sigma^2)^2}, & \frac{(r-r^f)}{\sigma^2} \\ -\frac{\gamma}{\sigma^2}, & -\frac{\gamma\mu_M\sigma^2 - \gamma(r-r^f)}{(\sigma^2)^2}, & -\frac{(r-r^f)}{\sigma^2} \end{bmatrix} \end{aligned} \tag{C.3}$$

Firstly, we consider an exogenous change in the risky asset's expected excess return. We denote with v^{ret} the size of the exogenous shock, then

$$PI_{[v^{\text{ret}},0,0]} = D\omega(\theta) \begin{bmatrix} v^{\text{ret}} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\gamma v^{\text{ret}}}{\sigma^2} \\ -\frac{\gamma v^{\text{ret}}}{\sigma^2} \end{bmatrix} \tag{C.4}$$

and the TPI can be evaluated as

$$TPI_{[v^{\text{ret}},0,0]} = \sqrt{2}\frac{\gamma|v^{\text{ret}}|}{\sigma^2}. \tag{C.5}$$

The sensitivity of the TPI to the risk-free rate is given by

$$\frac{\mathrm{d}TPI_{[v^{\text{ret}},0,0]}}{\mathrm{d}r^f} = -\sqrt{2}\frac{\gamma\lambda|v^{\text{ret}}|}{(\sigma^2)^2}. \tag{C.6}$$

The second possible shock is a shock to the variance of the risky asset (v^{var}):

$$PI_{[0,v^{\text{var}},0]} = D\omega(\theta) \begin{bmatrix} 0 \\ v^{\text{var}} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\gamma\mu_M\sigma^2-\gamma(r-r^f)}{(\sigma^2)^2}v^{\text{var}} \\ -\frac{\gamma\mu_M\sigma^2-\gamma(r-r^f)}{(\sigma^2)^2}v^{\text{var}} \end{bmatrix} = \begin{bmatrix} \frac{\gamma\mu_M-\omega^1}{(\sigma^2)}v^{\text{var}} \\ -\frac{\gamma\mu_M-\omega^1}{(\sigma^2)}v^{\text{var}} \end{bmatrix} \tag{C.7}$$

The total portfolio instability is given by $TPI_{[0,v^{\text{var}},0]} = \sqrt{2}\frac{|\omega^1-\gamma\mu_M|}{(\sigma^2)}|v^{\text{var}}|$ and the sensitivity with respect to the risk-free rate

$$\frac{\mathrm{d}TPI_{[0,v^{\text{var}},0]}}{\mathrm{d}r^f} \begin{cases} \sqrt{2}\frac{\frac{\partial\omega^1}{\partial r^f}\sigma^2-\lambda(\omega^1-\gamma\mu_M)}{(\sigma^2)^2}|v^{\text{var}}|, & \text{if } \omega^1-\gamma\mu_M > 0, \\ \sqrt{2}\frac{-\frac{\partial\omega^1}{\partial r^f}\sigma^2-\lambda(\gamma\mu_M-\omega^1)}{(\sigma^2)^2}|v^{\text{var}}|, & \text{if } \omega^1-\gamma\mu_M < 0, \end{cases} \tag{C.8}$$

For the third possible shock in the investor preference, i.e. in the parameter γ, we find similar results as in the previous scenario

$$PI_{[0,0,v^{\gamma}]} = D\omega(\theta) \begin{bmatrix} 0 \\ 0 \\ v^{\gamma} \end{bmatrix} = \begin{bmatrix} \frac{(r-r^f)}{\sigma^2}v^{\gamma} \\ -\frac{(r-r^f)}{\sigma^2}v^{\gamma} \end{bmatrix} \tag{C.9}$$

$$TPI_{[0,0,v^\gamma]} = \sqrt{2}\frac{\omega^1}{\gamma}|v^\gamma|. \tag{C.10}$$

$$\frac{\partial TPI_{[0,0,v^\gamma]}}{\partial r^f} = \sqrt{2}\frac{\frac{\partial \omega^1}{\partial r^f}}{\gamma}|v^\gamma| < 0 \iff \frac{\partial \omega^1}{\partial r^f} < 0. \tag{C.11}$$

This concludes the proof.

C.2 Econometric model with US Data

In this appendix we test the theoretical findings in Section 5.3 using US monthly data from April 1951 to December 2019. The market proxy is the value-weighted portfolio provided in the French's website[1] and risk-free rate proxy is the one-month T-bill rate. Table C.1 shows the estimates for the models 1.A (5.12), 1.B (5.13) and 1.C (5.14). The results are consistent with the findings in table 5.1 with an higher significance, partially due to the longer time series and partially due to the lower frequency of the data. The only difference between the US estimates and the EU estimates is the sign of the parameter ψ for the model 1.B but in both cases the results are not significant.

Figure C.1 shows the same non-linear behaviour observed in figure 5.4. The results show that the portfolio instability behaviour is robust across countries.

[1]Kenneth R. French website.

Parameters	model 1.A	model 1.B	model 1.C
μ_0	0.038	0.084	0.109
	(0.979)	(4.732)	(2.836)
μ_M	0.168	-0.002	0.301
	(1.192)	(-0.636)	(1.768)
μ_f			-2.718
			(-3.325)
ω	0.014	0.011	0.010
	(2.274)	(1.569)	(1.667)
α	0.115	0.114	0.113
	(3.717)	(3.001)	(2.979)
β	0.837	0.798	0.800
	(20.213)	(11.834)	(13.233)
$\tilde{\lambda}$		0.345	0.341
		(2.410)	(2.557)
ψ		-0.001	-2.616
		(-0.557)	(-3.358)
$\frac{\partial \omega^1}{\partial r^f}$		< 0	< 0
$\frac{\partial TPI_{[v^{\text{ret}},0,0]}}{\partial r^f}$		< 0	< 0
$\frac{\partial TPI_{[0,v^{\text{var}},0]}}{\partial r^f}$		< 0	< 0
$\frac{\partial TPI_{[0,0,v^{\gamma}]}}{\partial r^f}$		< 0	< 0
AIC	1176.6	1164.0	**1150.6**

Table C.1: Estimation of models 1.A, 1.B and 1.C in equations (5.13) and (5.14) and model implied sensitivities to the interest rate. In the first section we report the estimates and the Robust t ratios in parenthesis. The estimates ψ is evaluated according to (5.5) and the Robust t-statistic is evaluated with the delta method. In the second section we report the different quantities in equations (5.7), (5.8), (5.9) and (5.10). When the sign is < 0 the partial derivative is negative in the whole sample, when the sign is > 0 the partial derivative is positive and when the sign is $\pm$ the partial derivative can be both positive and negative. In the last row we show the AIC and the best performing model according to this measure is bolded.

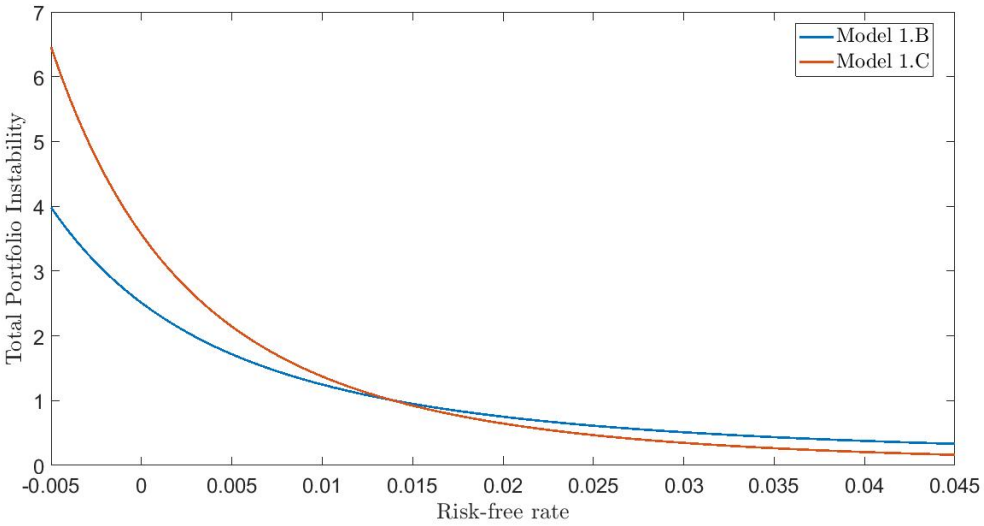

Figure C.1: Total Portfolio Instability sensitivity evaluated from model 1.B (5.13) and model 1.C (5.14) for different values of the risk-free rate. For the calculations, we use the unconditional levels of σ^2 and $r - r^f$ given by models (5.13) and (5.14) for different levels of r^f.