

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

LOUVAIN FINANCE

LOUVAIN INSTITUTE OF DATA ANALYSIS AND MODELING IN ECONOMICS AND STATISTICS



Essays on Financial Econometrics and Quantitative Finance

Linqi Wang

Doctoral Thesis

Supervisors:

Christian Hafner
(UCLouvain, Belgium)
Frédéric Vrans
(UCLouvain, Belgium)

Jury:

Damiano Brigo (Imperial College London, UK)
Pierre Devolder (UCLouvain, Belgium)
Donatien Hainaut (UCLouvain, Belgium)
Oliver Linton (University of Cambridge, UK)

President:

Pierre Devolder (UCLouvain, Belgium)

Abstract

Negative interest rates have significantly impacted multiple segments of financial markets and market participants' investment behavior. With nominal interest rates entering negative territory for the first time in history, a reassessment of the existing framework for interest rate modelling and derivative pricing is warranted. Moreover, lower bond yields induce investors to seek higher returns in other asset classes which are potentially associated with higher risks. Therefore, priority should be given to the development of new modelling solutions for the real-time monitoring of financial market volatility, correlation, and liquidity conditions given their central role in portfolio allocation and risk management.

This dissertation revolves around the development of modelling approaches with relevant applications to financial markets. In particular, Chapter 1 studies the implications of the negative interest rate environment for the modelling of short rates and the pricing of interest rate derivatives. In addition, we propose a model calibration approach that takes into account the whole implied forward rate distribution rather than sparse data points for derivative prices or implied volatilities. Chapter 2 focuses on the development of a Dynamic Conditional Score (DCS) model for the log correlation matrix which can accommodate fat tails in the conditional distributions and generalizes the Beta-t-EGARCH model of [Harvey \(2013\)](#), which uses the student-t distribution, to the multivariate case. In Chapter 3, we develop a new algorithm for dynamic portfolio selection which combines a Dynamic Conditional Correlation (DCC) model with non-linear shrinkage for the dynamics of asset returns with LASSO-type penalization schemes to control for within- and between-group variations in portfolio weights. In the application, we consider US stocks whose sectoral classification is used for group assignment. Other applications could also be considered. For instance, this framework could be used to efficiently exploit the diversification opportunities across regions and asset classes arising in a negative rate environment. Finally, Chapter 4 proposes a new class of dynamic autoregressive liquidity models for the Amihud illiquidity measure to capture both the long-run trend in illiquidity series with a nonparametric component and short-run illiquidity dynamics with an autoregressive component. This allows us to provide real-time monitoring of liquidity conditions for asset classes which might constitute attractive diversification opportunities in a negative interest rate environment.

Acknowledgments

This thesis would not have been possible without the support of many people. I thus would like to take this opportunity to express my gratitude to some of them.

First and foremost, I would like to thank my supervisors, Prof. Christian Hafner and Prof. Frédéric Vrms. Prof. Hafner, thank you for introducing me to the world of financial econometrics and for your continuous support from the very beginning. You were always available for in-depth discussions on research projects and to help me navigating the obstacles of the publication process for our papers. I benefited immensely from your career advice and support and this led me to pursue an academic career in financial econometrics. Prof. Vrms, thank you for your continuous support and encouragement. You have always been willing to provide me with your heartfelt guidance on our research projects and to support me throughout my thesis journey with your invaluable assistance and suggestions. You also gave me the opportunity to develop my teaching skills by sharing your own experience, which helped me better prepare for my academic career. This thesis would not have been possible without both of you and I could not have wished for better advisors. I am very grateful to have had the opportunity to collaborate with and learn from both of you.

Second, I would like to thank Prof. Oliver Linton for agreeing to be my host and advising me during my research visiting at the University of Cambridge. It was a wonderful experience to have the opportunity to collaborate with you on our joint project and to learn from your extensive knowledge and expertise in econometrics during our regular discussions. I would like to take this opportunity to also thank Prof. Andrew Harvey for his invaluable feedback and advice on DCS modelling. I benefited immensely from our discussions on potential extensions and applications of this framework.

I am also indebted to my committee members – Prof. Damiano Brigo, Prof. Pierre Devolder, Prof. Donatien Hainaut and Prof. Oliver Linton – for accepting to evaluate my thesis. I am thankful for your time, availability, and for your constructive comments and feedback which led to a significant improvement in the quality of the final version.

Over time, I had and continue to have great colleagues at LFIN and LIDAM, as well as those I met during my research visit at the University of Cambridge. You were always supportive and your valuable feedback during seminars and workshops helped me improve my work. A special thanks goes to the administrative team and IT support at Université catholique de Louvain

and University of Cambridge who were always available to help me whenever I encountered administrative difficulties or technology related issues.

I am grateful to the *Actions de Recherche Concertées* (ARC) for providing the financial support for my thesis.¹ I am also grateful to the *Fonds de la Recherche Scientifique* (FNRS) and the Erasmus Traineeship Program for having provided funding to support my research visit at the University of Cambridge.

Finally, I would like to thank my family and friends. It has been a challenging period due to COVID, but your continuous support, understanding and love made me strong and determined. Last but definitely not the least, I would like to thank my partner for his unwavering support, encouragement and love at every step of this journey. I am fortunate to have you in my life and I look forward to continuing our journey together.

¹The PhD thesis is funded by an ARC grant 18-23/089: “Negative and ultra-low interest rates: behavioral and quantitative modelling”.

Contents

Introduction	1
1 Asymmetric short-rate model without lower bound	13
1.1 Introduction	13
1.2 Model description	16
1.2.1 Baseline framework	17
1.2.2 Proposed extension	19
1.2.3 Forward rate dynamics and densities under the forward measure	19
1.3 Model calibration	22
1.3.1 Calibration of the implied volatility curve	23
1.3.2 Calibration of the forward rate density	23
1.3.3 Summary of Calibration Algorithm	25
1.4 Empirical application	27
1.4.1 Data description and volatility curve smoothing	27
1.4.2 Calibration results for the implied volatility curve	29
1.4.3 Calibration results for the forward rate density	31
1.5 Conclusions	38
Appendix 1.A Pricing formula for existing framework	40
Appendix 1.B Arbitrage-free smoothing of the volatility curve	41
Appendix 1.C Correlation between two factors	43
1.C.1 Introduce correlation within the affine framework	43
1.C.2 Introduce correlation by correlated Brownian Motions	44
Appendix 1.D Proof of Proposition 1	45

Appendix 1.E	Calibration results of implied volatility curve: 29 January 2021	47
Appendix 1.F	The G2++ model	48
2	A dynamic conditional score model for the log correlation matrix	51
2.1	Introduction	51
2.2	Matrix logarithm of a correlation matrix	54
2.3	The model and its estimation	55
2.3.1	One-step estimation	58
2.3.2	Two-step estimation	62
2.4	Empirical application	66
2.4.1	Data description	66
2.4.2	Volatility estimation results	67
2.4.3	Results for the conditional correlations	68
2.5	Conclusions	73
Appendix 2.A	Theoretical material	74
2.A.1	Lyapounov exponent of the SRE in the bivariate case	74
2.A.2	Proof of Theorem 1	75
2.A.3	Proof of Theorem 2	76
2.A.4	Proof of Theorem 3	78
Appendix 2.B	Tables and figures	78
Appendix 2.C	Example: the bivariate Gaussian case	82
Appendix 2.D	Lyapounov exponent of the SRE	85
Appendix 2.E	Expressions for the information matrix	91
Appendix 2.F	Monte Carlo simulation	93
Appendix 2.G	Additional tables and figures	95
3	Dynamic portfolio selection with sector-specific regularization	105
3.1	Introduction	105
3.2	The portfolio selection methodology	110
3.2.1	Notation	110
3.2.2	Dynamic conditional covariance matrix estimation	110
3.2.3	Global minimum-variance portfolio	113

3.2.4	The dynamic portfolio selection algorithm	116
3.3	Empirical application	117
3.3.1	Data description	118
3.3.2	Dynamic covariance matrix estimation	119
3.3.3	GMV portfolio performance without short-sale constraints	121
3.3.4	GMV portfolio performance with short-sale constraints	126
3.4	Conclusions and outlook	131
Appendix 3.A	Additional tables	132
4	Dynamic Autoregressive Liquidity (DARLiQ)	137
4.1	Introduction	137
4.2	Amihud illiquidity	140
4.3	The model	144
4.4	Estimation	145
4.4.1	Estimation based on conditional moment restriction	146
4.4.2	Estimation based on i.i.d. assumption	147
4.4.3	Sieve semiparametric estimation procedure	150
4.5	Large sample properties	152
4.5.1	Conditional moment restrictions	153
4.5.2	I.i.d. shocks	155
4.6	Testing for temporary and permanent shifts	157
4.7	Risk premium	160
4.8	Empirical application	160
4.8.1	Data description	161
4.8.2	Estimation results	162
4.8.3	Testing for permanent shifts: discontinuity in g function	170
4.8.4	Risk premium	176
4.9	Conclusions	177
Appendix 4.A	Lemmas	179
Appendix 4.B	Proof of main results	179
Appendix 4.C	Semiparametric efficiency	182
4.C.1	Known f	182

4.C.2	Parametric f	185
4.C.3	Unknown f	187
Appendix 4.D	Risk premium regressions	191
Appendix 4.E	Other tables and figures	191
4.E.1	Estimation based on conditional moment restrictions	191
4.E.2	Estimation: i.i.d. error term with parametric density	193
4.E.3	Estimation: i.i.d. error term with nonparametric density	200
4.E.4	Risk premium	203

List of Figures

1	S&P 500 index daily log illiquidity.	10
1.1	Interpolated caplet prices and market-implied probability density.	28
1.2	Volatility curve analysis of calibration approach minimizing implied volatility error.	29
1.3	Pricing accuracy analysis of calibration approach minimizing implied volatility error.	30
1.4	Forward density analysis of calibration approach minimizing implied volatility error.	31
1.5	Forward density analysis of calibration approach minimizing JS divergence.	32
1.6	Pricing accuracy analysis of calibration approach minimizing JS divergence.	33
1.7	Volatility curve analysis of calibration approach minimizing JS divergence.	34
1.8	Evolution over time of JS divergence resulting from density matching calibration.	35
1.9	Time series of the market-implied and modelled conditional moments of L(6M,1Y).	36
1.10	Volatility curve analysis based on implied volatility error minimization on 29 January 2021.	47
1.11	Forward density analysis of calibration approach minimizing JS divergence under the G2++ and the VaCIR++ models.	49
1.12	Evolution over time of JS divergence resulting from density matching calibration under the Hull-White, G2++ and VaCIR++ models.	50
2.1	The correlation ρ of the transformed r.v. as a function of the original correlation $\tilde{\rho}$	64
2.2	Time series of daily asset prices.	80
2.3	Estimated conditional correlation series.	81
2.4	The combinations of (ϕ, κ) for which $\mathbb{E}[\log \Lambda(\psi)] = 0$	86
2.5	Simulated $\lambda_{min}(\Theta_t)$ for $N = 6$ and sample sizes up to 1,000,000 observations.	90

2.6	Time series of daily log-returns.	101
2.7	Estimated volatility series using the Beta-t-EGARCH model.	102
2.8	Histograms of PITs of standardized residuals of the Beta-t-EGARCH model. . .	103
3.1	Boxplots of GARCH(1,1) parameter estimates.	120
3.2	Out-of-sample period weight evolution.	128
4.1	Fab 5 daily log illiquidity – $\log \ell_t$	141
4.2	S&P 500 index and Bitcoin asset daily log illiquidity – $\log \ell_t$	143
4.3	Fab 5 and Bitcoin illiquidity series and trend functions ($\times 10^{10}$).	163
4.4	Fab 5 and Bitcoin log illiquidity series and trend functions.	164
4.5	Test statistics for detecting permanent breaks in the illiquidity series – Facebook, Amazon, Apple.	173
4.6	Test statistics for detecting permanent breaks in the illiquidity series – Google, Microsoft, Bitcoin.	174
4.7	Fab 5 and Bitcoin log illiquidity and updated trend function based on the GMM estimator of λ_t parameters.	192
4.8	ACF of ζ_t under the symmetric specification for λ_t	193
4.9	ACF of ζ_t under the asymmetric specification for λ_t	194
4.10	ACF of ζ_t^2 under the symmetric specification for λ_t	195
4.11	ACF of ζ_t^2 under the asymmetric specification for λ_t	196
4.12	PIT of ζ_t under the symmetric specification for λ_t	197
4.13	PIT of ζ_t under the asymmetric specification for λ_t	198
4.14	Fab 5 and Bitcoin log illiquidity and updated trend function based on the semi- parametric ML estimator of λ_t parameters where the error term ζ_t follows a Weibull distribution.	199
4.15	Comparison between the kernel density estimate of ζ_t and the Weibull density under the symmetric specification for λ_t	200
4.16	Comparison between the kernel density estimate of ζ_t and the Weibull density under the asymmetric specification for λ_t	201

4.17	Fab 5 and Bitcoin log illiquidity and updated trend function based on the semi-parametric ML estimator of λ_t parameters where the density of the error term ζ_t is estimated nonparametrically.	202
4.18	S&P 500 index daily (log) illiquidity series and return data.	203
4.19	S&P 500 index weekly (log) illiquidity series and return data.	204
4.20	S&P 500 index monthly (log) illiquidity series and return data.	205

List of Tables

2.1	Diagnostics and model selection criteria for multivariate models.	78
2.2	Modelling comparison for dynamic correlation matrix.	79
2.3	Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio.	79
2.4	Parameters used in the data generating process.	93
2.5	Bias and RMSE of parameter estimates.	94
2.6	Description of the assets considered in the empirical application.	95
2.7	Descriptive statistics for historical log-returns.	95
2.8	The sample correlation matrix of returns.	96
2.9	Parameter estimates of the volatility using Beta-t-EGARCH(1,1).	96
2.10	Volatility model diagnostics.	97
2.11	Descriptive statistics for standardized residuals.	97
2.12	Two-step parameter estimates for the correlation part of the model.	98
2.13	Model descriptions.	99
2.14	Parameter estimates for the dynamic correlation models.	99
2.15	Parameter estimates for the dynamic volatility and correlation models.	100
3.1	DCC parameter estimates for five sample sizes and for four investment universes.	121
3.2	Out-of-sample annualized portfolio standard deviation and Sharpe ratio.	123
3.3	Statistics of portfolio weights.	127
3.4	Out-of-sample annualized portfolio standard deviation and Sharpe ratio (with short-sale constraint).	129
3.5	Statistics of portfolio weights (with short-sale constraint).	130
3.6	Logarithm of penalty tuning parameters $\log(\lambda)$ (without short-sale constraint).	132

3.7	Logarithm of penalty tuning parameters $\log(\lambda)$ (with short-sale constraint).	132
3.8	P-values for statistical test of difference in volatility (without short-sale constraint).	133
3.9	P-values for statistical test of difference in volatility (without short-sale constraint).	134
3.10	Running time for the out-of-sample period (in minutes).	135
4.1	Estimated parameters of an AR(5) with trend.	142
4.2	Summary statistics for daily illiquidity $-\ell_t \times 10^{10}$.	161
4.3	Estimated parameters of the λ_t process based on first moment restriction.	165
4.4	Fully efficient estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Weibull distribution.	167
4.5	Fully efficient estimates of the parameters for λ_t process when using the nonparametric estimates of the density of the error term ζ_t .	169
4.6	Log-likelihood comparison between models using the parametric (Weibull) and nonparametric estimates of the ζ_t density.	169
4.7	Average statistics for testing permanent breaks in the liquidity series.	170
4.8	Test statistics for detecting permanent and temporary breaks in the liquidity series of Amazon and Apple stocks.	172
4.9	Test statistics for detecting permanent and temporary breaks in the liquidity series of Google and Microsoft stocks.	175
4.10	Coefficient estimates for regressions using daily, weekly and monthly observations.	177

Introduction

Low interest rate environment and its implications

The aftermath of the 2007-2009 Global Financial Crisis (GFC) left most advanced economies in an environment with weak growth prospects and low inflation. Central banks across the world had cut their policy rates during the crisis to levels approaching zero – which was previously considered as a lower bound on nominal interest rates. Therefore, they needed to resort to unconventional monetary policy tools to provide further stimulus in this low interest rate environment. One such tool was negative policy rates which were adopted by the European Central Bank and other central banks in Denmark, Switzerland, Sweden, and Japan. This new policy tool has had a significant impact on financial markets. Firstly, negative rates directly impact bond markets by reducing yields on short term bonds and giving investors incentives, together with access to cheaper funding, to seek higher yields in riskier segments, such as government or corporate bonds. Moreover, there are also implications for other asset classes as low bond yields drive investors into riskier asset classes. In equity markets for instance, both value and growth stocks become more attractive – the former because they usually pay the highest dividends and the latter due to the boost in valuation resulting from a lower discount factor on future cash flows. Secondly, negative rates impact more complex financial products whose pricing requires interest rates as an input. In particular, negative rates have direct implications for the pricing of interest rate derivatives whose underlyings are rates, and indirect implications for equity derivatives where interest rates affect pricing through the discount factor.

The multiple financial market implications of the low interest rate environment outlined above open interesting avenues of research. First, negative interest rates call for a reassessment of the existing toolkit for interest rate modelling as several widely used approaches, with the Cox-Ingersoll-Ross (CIR) short-rate model being the most prominent example, explicitly restrict

interest rates to be non-negative. The implications of this unsuitable restriction for derivative pricing also warrant further investigation. Chapter 1, which is based on [Vrins and Wang \(2021\)](#), study the implications of negative interest rates on short-rate modelling and develop a new calibration method for the pricing of interest rate derivatives.

Second, low interest rates induce a search-for-yield environment where investors consider riskier assets or investment strategies to reap higher returns. For instance, they may consider not only classical financial assets such as stocks and bonds, but also alternative investments such as commodities and cryptocurrencies. Empirical evidence suggests that alternative investment assets exhibit higher volatility than classic asset classes. Moreover, these assets provide interesting diversification opportunities due to their low correlation with stock markets but the diversification benefits tend to vary over the business and financial cycles. Therefore, an accurate modelling of conditional volatilities and correlations is crucial to fully exploit the diversification opportunities arising in a negative interest rate environment. In Chapter 2, [Hafner and Wang \(2021\)](#) focus on the development of a new dynamic conditional score model for the log correlation matrix which is then applied to financial and alternative assets data.

Third, the low interest rate environment – which mainly affects advanced economies – also creates diversification opportunities to other geographical areas where rates are higher such as emerging markets. Against this backdrop of search-for-yield across both asset classes and geographical regions, taking the perspective of a portfolio manager facing an environment with negative interest rates is an interesting jumping-off point to develop a flexible cross-asset investment framework. A key element to capture the constraints inherent in such an environment is to augment the standard mean-variance framework for asset allocation with a set of criteria motivated by investors' preferences or constraints. For example, investors might face scenarios where they would like to cap their overall exposure to a given asset class or geographical area. They might also want to control the amount of turnover in their portfolio on rebalancing dates to limit transaction costs. In Chapter 3, [Hafner and Wang \(2022\)](#) propose a new algorithm for dynamic portfolio selection which can accommodate such preferences or constraints and apply it to US stocks by exploiting the sector structure in the data.

Finally, liquidity is a key factor to consider for an investor facing a negative interest rate environment for two reasons. Firstly, the search-for-yield behavior induced by low interest rates can lead investors towards riskier market segments, e.g. high-yield corporate bonds, which

exhibit both higher credit and liquidity risks than higher-rated bonds. Secondly, a potential rapid lift-off in central banks' policy rates could trigger redemption waves in bond funds as bond prices drop, and the resulting selling pressures might in turn create a more widespread liquidity spiral in financial markets. It is therefore important to have a modelling framework which allows to accurately measure and forecast liquidity conditions in real time for different asset classes. In Chapter 4, [Hafner et al. \(2022\)](#) contribute to this line of research by developing a new class of Dynamic Autoregressive Liquidity (DARLiQ) models for the Amihud illiquidity measure.

Outline of the thesis

The four main chapters of this dissertation are self-contained research articles that can be read independently from each other. They all focus on the development of new modelling approaches to better capture the key features of financial market data.² The outline for the dissertation is as follows.

Chapter 1: Asymmetric short-rate model without lower bound

The Vasicek model ([Vasicek \(1977\)](#)) and the Cox-Ingersoll-Ross model ([Cox et al. \(1985\)](#); CIR hereafter) are considered as standard benchmarks for the modelling of short-rate dynamics due to their ability to capture mean-reverting dynamics in the data and their analytical tractability – as explicit solutions can be obtained for bond prices and bond options. The short-rate dynamics under the Vasicek model is specified as

$$dr_t = \kappa (\theta - r_t) dt + \sigma dW_t, \quad r_0 \in \mathbb{R}, \quad (1)$$

where κ , θ , and σ are positive constants; and W_t is a Brownian motion under the risk-neutral measure $\mathbb{Q}$. However, this process implies a Gaussian distribution for the short rate which contradicts the empirical evidence that interest rates display conditional asymmetry in their dynamics. Therefore, the CIR model is often favored as it can provide the desired feature of

²The main computations for the thesis were carried out using the R programming language on a 2020 iMac with the following specifications: Intel Core i9 10-Core processor (3.6GHz) and 72 GB RAM (2133 MHz DDR4). We also implemented the non-sequential calculations with parallel programming techniques to shorten the running time.

positive skewness in the short-rate distribution. The specification of the CIR model is given by

$$dr_t = \kappa (\theta - r_t) dt + \sigma \sqrt{r_t} dW_t, \quad r_0 \in \mathbb{R}^+, \quad (2)$$

where, as above, κ , θ , and σ are positive constants; and W_t is a Brownian motion under $\mathbb{Q}$. Moreover, the parameters have to satisfy the Feller condition $2\kappa\theta > \sigma^2$ for the r_t process to be strictly positive. This positivity constraint of the square-root process used to be an advantage of the model prior to the negative interest rate period, but this is no longer the case in the current environment.

Negative interest rates call for a reassessment of the existing toolkit for interest rate modelling as both the Vasicek and CIR models exhibit undesirable features for the current environment. Indeed, the most direct solution for the negative rate environment would be to return to the Vasicek model. However, the process dynamics would fail to capture the desired non-Gaussian features in the data. Another straightforward solution is to introduce a constant negative shift to the CIR process. The resulting model would nevertheless still exhibit a strict lower bound (the shift itself). In addition, although one could treat the shift as an additional parameter to be estimated from the data, this approach does not specify how to reliably estimate this lower bound parameter from a theoretical or empirical perspective.

This motivates us to propose a new short-rate process in Chapter 1 to appropriately capture the salient features of the negative interest rate environment. Our proposed model combines the advantages of the Vasicek and Cox-Ingersoll-Ross (CIR) dynamics – it is flexible, tractable and displays positive skewness without imposing a strict lower bound. Furthermore, the tractability of our framework allows us to obtain semi-analytical expressions for the forward rate density and caplet pricing. We also propose a novel approach to calibrate short-rate models by forward rate density matching, which minimizes the distance between the market- and model-implied densities. To this end, we select as our calibration criteria a symmetrized and smoothed version of the Kullback-Leibler (KL) divergence, the Jensen–Shannon (JS) divergence.

We use our framework to analyze the implications of the negative interest rate environment for the market-implied distribution of EURIBOR forward rates and its conditional moments. Moreover, we provide a thorough empirical analysis demonstrating the superior performance of our model compared to the Vasicek and CIR models. Additionally, we document that our proposed calibration procedure based the JS divergence minimization better captures the entire forward rate distribution compared to competing approaches – which minimize the price or

volatility discrepancies – while maintaining a good fit in terms of pricing and implied volatility errors. Therefore, the calibrated parameters obtained based on this approach can then be used to price new, complex, and illiquid derivatives with the same underlying forward rate.

Chapter 2: A dynamic conditional score model for the log correlation matrix

The measurement and estimation of volatilities and correlations are essential tools in the functioning of financial markets as many decisions faced by investors require a quantification of expected risks and expected payoffs or returns. Salient examples in which volatilities and correlations play a central role are portfolio optimization, risk management and option pricing. To be relevant for these empirical applications, a model should be able to capture the main stylized facts from financial market data: volatility clustering, mean-reverting dynamics of volatility, the leverage effect, non-Gaussian distribution of the standardized residuals, and time-varying correlations.

A popular choice in the literature for the modelling of conditional volatilities and correlations is the Dynamic Conditional Correlation (DCC) model of [Engle \(2002\)](#). One reason for the popularity of this approach is that it allows to decouple the specification of the volatility dynamics from the correlation dynamics thanks to the following representation of the conditional covariance matrix H_t

$$H_t = D_t R_t D_t , \quad (3)$$

where D_t is a diagonal matrix containing the conditional volatility components and R_t is the conditional correlation matrix. For the volatility components, the class of (Generalized) AutoRegressive Conditional Heteroscedasticity, or (G)ARCH models, introduced in the seminal work by [Engle \(1982\)](#) and [Bollerslev \(1986\)](#), emerged as the standard approach for dynamic volatility modelling. Suppose $y_t = \sigma_t \varepsilon_t$ is the asset return at time t where ε_t is a random variable with mean zero and variance one. We can specify the dynamics of the univariate asset volatility using a GARCH(1,1) process as

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 . \quad (4)$$

For the dynamics of the conditional correlation matrix, a DCC(1,1) model can be specified for the conditional pseudo-correlation matrix Q_t as

$$Q_t = (1 - \alpha - \beta) C + \alpha \varepsilon_{t-1} \varepsilon_{t-1}' + \beta Q_{t-1} , \quad (5)$$

where ε_t is the vector of devolatilized residuals, with elements y_{it}/σ_{it} , and C is the unconditional correlation matrix of ε_t . Note that the matrix Q_t can also be interpreted as the conditional covariance matrix of the devolatilized residuals. Since the diagonal elements of Q_t are close to one but not exactly equal to one, the conditional correlation matrix can be obtained by rescaling the Q_t matrix, $R_t := \text{Diag}(Q_t)^{-1/2} Q_t \text{Diag}(Q_t)^{-1/2}$.

Despite the flexibility of the modelling framework outlined above and its satisfactory empirical performances, it suffers from some limitations. The GARCH model specified in Equation (4) features lagged squared observations which make the updating of the volatility process overly responsive to extreme realizations. These extreme realizations are common in financial data whose error distribution displays fat-tails. To address this issue, Dynamic Conditional Score (DCS) models, also called Generalized Autoregressive Score (GAS) models, are introduced in the seminal works of [Harvey \(2013\)](#) and [Creal et al. \(2011\)](#). They propose to replace the squared past observations in the updating equation for the volatility dynamics by the score of the conditional density of the standardized errors. This leads to an intuitive updating mechanism for the time-varying parameters as the conditional score indicates the direction of steepest ascent to update the model's local fit in terms of likelihood, given the current position of the parameters. In addition, the score depends on the complete density, and not only on the first- or second-order moments of the observations. Furthermore, [Harvey \(2013\)](#) assumes a student-t distribution in his Beta-t-EGARCH model to better match the fat-tail feature in financial asset return data and he uses an exponential link function to ensure positivity of the volatility process. It is interesting to note that, by imposing the use of the conditional score in the updating equation of the time-varying parameters, the DCS approach introduces a strong link between the conditional distribution of the error term and the updating mechanism for the model's parameters. While the updating based on the student-t distribution appears appropriate for the modelling of asset returns, testing for the empirical relevance of this modelling assumption might be warranted in other applications. We do not investigate this avenue in the current thesis but the interested reader is referred to [Blasques et al. \(2022\)](#) for a first attempt at generalizing score-driven methods.

In Chapter 2, we aim to extend the Beta-t-EGARCH model of [Harvey \(2013\)](#) to a multivariate setting. One key element to achieve this goal is to ensure the positive definiteness of the conditional correlation matrix in arbitrary dimensions. We achieve this through exponenti-

ation of the conditional log correlation matrix thanks to the new parametrization proposed by [Archakov and Hansen \(2021\)](#) which generalizes the Fisher z-transformation from the bivariate case to higher dimensions. Our proposed dynamic model for the log correlation matrix, whose dynamics are driven by the likelihood score with respect to the log parameters defined from the matrix logarithm of the correlation matrix, accommodates fat tails in the conditional distribution of returns thanks to a student-t copula capturing the dependence structure and univariate student-t for the marginals – with potentially different degrees of freedom.

We derive estimation theory for both a one-step and a two-step estimator which allows for separation into volatility and correlation parts and remains computationally feasible in settings with higher dimensions. An empirical application using financial and alternative assets data demonstrates the model performance in terms of in-sample estimation and out-of-sample forecasting when compared to alternative modelling approaches in the literature. We demonstrate that our model provides the best fit to the data in terms of likelihood, AIC and BIC criteria. We use the minimum variance portfolio selection application to demonstrate the out-of-sample forecasting performance where we show that a portfolio constructed based on the log correlation model has comparable volatility and Sharpe ratio to alternative DCS parameterizations and superior performance when compared with the corrected DCC (cDCC) model of [Aielli \(2013\)](#).

Chapter 3: Dynamic portfolio selection with sector-specific regularization

Portfolio selection has been a central topic in empirical finance since it was first proposed by [Markowitz \(1952\)](#). Typically, portfolio performance is evaluated based on the maximization of a measure of risk-adjusted performance, for instance the Sharpe ratio criteria. However, it has been shown in the literature that the estimation and prediction of asset mean returns can be challenging, see e.g. [Best and Grauer \(1991\)](#). Furthermore, studies such as [Jagannathan and Ma \(2003\)](#) show that prioritizing the minimization of portfolio variance can lead to portfolio allocations outperforming, in terms of out-of-sample Sharpe ratios, other competing approaches which involve the estimation of mean returns. Therefore, we refrain from modelling asset mean returns and focus on the global minimum variance (GMV) portfolio.

The GMV portfolio problem for a given covariance matrix H_t is formulated as a constrained

minimization problem, i.e.

$$\begin{aligned} \min_w w' H_t w \\ \text{s.t. } w' \mathbf{1} = 1 . \end{aligned} \tag{6}$$

This optimization problem has an analytical solution for the portfolio weights, $w_t = \frac{H_t^{-1} \mathbf{1}}{\mathbf{1}' H_t^{-1} \mathbf{1}}$. Optimal portfolio weights can thus be obtained for a given estimate of the covariance matrix, e.g. using the sample estimate. However, this naive approach leads to poor out-of-sample performance, especially in large dimensional cases, due to unstable and ill-conditioned sample estimates of the covariance matrix.

Solutions have been proposed to address this issue, such as the shrinkage approach (see e.g. [Ledoit and Wolf \(2004\)](#) and [Ledoit and Wolf \(2017\)](#)) and weight regularization ([Jagannathan and Ma \(2003\)](#); [DeMiguel et al. \(2009\)](#); and [Fan et al. \(2012\)](#)). In addition, [Engle et al. \(2019\)](#) show that, although dynamic frameworks also face potential challenges in large dimensional settings, the GARCH-DCC model (see Equations (4)-(5) above in the overview section on Chapter 2) implemented with composite likelihood estimation and nonlinear shrinkage technique provides superior out-of-sample portfolio performance.

In Chapter 3, we use the dynamic GARCH-DCC framework as a starting point to investigate how a portfolio manager could efficiently exploit relevant information about the structure of the investment universe – e.g. based on assets belonging to different asset classes, geographical regions, and industry sectors – to further improve the resulting portfolio allocation. We propose a new algorithm for dynamic portfolio selection that construct asset allocation strategies subject to a set of criteria motivated by investment preferences or constraints. Our proposed framework is very general but we focus our empirical application on a dataset of 500 US stocks, equally split across 10 sectors based on the Global Industry Classification Standard (GICS), to highlight that our approach is particularly well-suited to large-dimensional applications. We incorporate directly a sector-specific structure into our portfolio optimization procedure, which is defined through regularizations with respect to within- and between-sector variations of portfolio weights. The framework includes two special cases as benchmarks: the GARCH-DCC approach and the equally-weighted portfolio. An algorithm is proposed for the estimation of the model parameters and the calibration of the penalty terms based on cross-validation approach in a dynamic setting.

In practice, to avoid unreasonable amount of transaction costs and maintain a lower turnover, it is more common to re-balance the portfolios at a monthly frequency. However, we rely on daily data for the estimation of the conditional covariance matrix as it leads to more accurate estimates and forecasts. This creates a mismatch between the frequencies for modelling and portfolio construction as the dynamic model provides a forecast of the covariance matrix for the next day, while we want to construct a portfolio based on its predicted performance over the next month. To address this mismatch, we use an “averaged-forecasting” approach to obtain the covariance matrix under the GARCH-DCC model. For each portfolio construction date, we forecast the covariance matrix for all days in the upcoming month and then average those forecasts to construct the portfolio at this date.³

We provide a comprehensive performance analysis of our framework considering four different investment universes with portfolio size $N \in \{50, 100, 250, 500\}$. We document that among the three sector-wise regularizations considered, controlling for within-sector weight variation has the highest contribution to reducing the out-of-sample portfolio volatility. This observation indicates that within-sector penalization is more important than between-sector diversification or concentrating the allocation on a few selected sectors. Additionally, we show that introducing sparsity in asset allocation weights and promoting stable weight evolution over time also significantly improve the out-of-sample portfolio volatility.

Chapter 4: Dynamic Autoregressive Liquidity (DARLiQ)

Liquidity is a key property of financial assets and plays a central role in the functioning of financial markets. [Amihud \(2002\)](#) defines the daily illiquidity ratio as $\ell_t = \frac{|R_t|}{V_t}$, where R_t is the stock return and V_t is the (dollar) trading volume at time t . The Amihud illiquidity measure reflects the fact that a stock is less liquid if a given trading volume generates a larger move in its price. Empirical evidence points to the existence of factors driving low-frequency variations in illiquidity dynamics in addition to higher-frequency variations. To illustrate this point, [Figure 1](#) plots the evolution of the daily log Amihud illiquidity measure for the S&P 500 stock market index over the period of 1950–2021. We observe that the illiquidity series exhibit a strong downward trend over time. It is worth noting that there is a temporary uptick in the

³The reader is referred to Section 3.2 of Chapter 3 for details on how to obtain the conditional covariance matrix forecast to construct a monthly re-balanced portfolio.

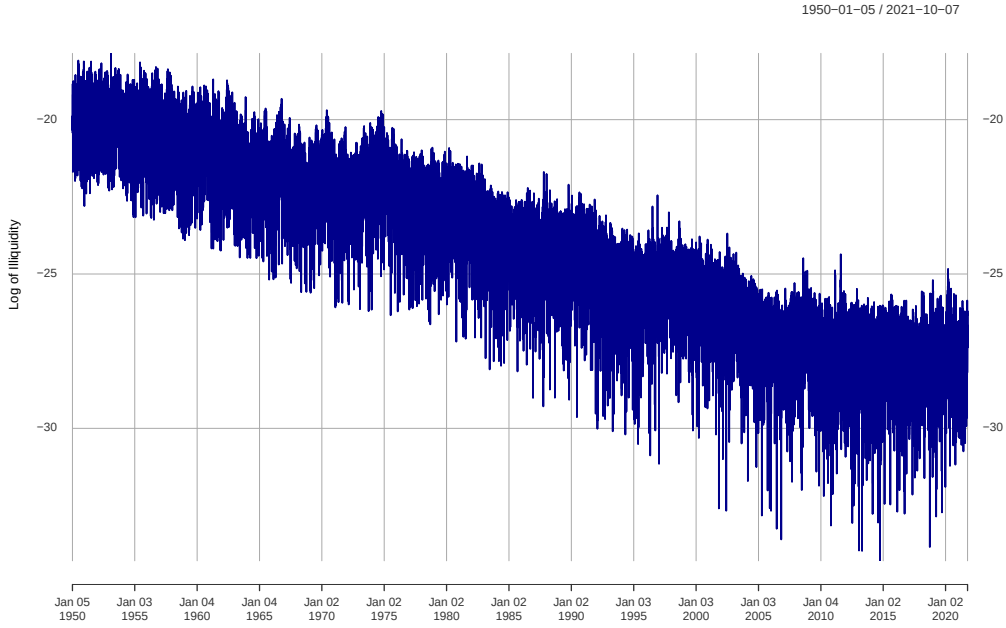


Figure 1: S&P 500 index daily log illiquidity.

illiquidity level during the 2007-2009 GFC, indicating a worsening in market liquidity conditions. Furthermore, market liquidity fluctuated at a more stable level in the post-GFC period.⁴ This motivates our modelling approach which weakens the requirement on stationarity.

Our approach builds on Multiplicative Error Models (MEMs), which have been developed to model the evolution of a non-negative process $x_t = \mu_t \varepsilon_t$ exhibiting some degree of persistence in its dynamics. It is specified as the product of a time-varying autoregressive scale factor μ_t – which depends upon the recent past of the x_t series – and a standard positive valued random variable ε_t with unit mean. The scale factor μ_t captures the conditional mean of x_t and its dynamics can be specified as

$$\mu_t = \omega + \beta \mu_{t-1} + \alpha x_{t-1} , \quad (7)$$

where the $\beta \mu_{t-1}$ term captures the persistence in the dynamics and the αx_{t-1} term captures the contribution of the more recent observation. The MEM approach has been successfully applied to the modelling of duration between irregularly spaced transaction data ([Engle and Russell](#)

⁴One potential reason for this stabilization in market liquidity after the GFC is the regulatory tightening on security brokers and dealers which might have led to a higher cost of market making – through increased cost of capital – and a decrease in risk taking due to leverage constraints. See [Adrian et al. \(2017\)](#) for a detailed account of the potential impact of regulation on the post-GFC trading environment for US Treasuries and corporate bonds.

(1998)); the joint modelling of duration, transaction volume, and volatility (Manganelli (2005)); and volatility modelling using multiple volatility proxies (Engle and Gallo (2006)).

As mentioned above, nonstationarity or trend in the illiquidity dynamics might be an important feature of financial market data. We account for this element by augmenting the standard MEM specification with a nonparametric trend component as follows

$$x_t = g(t/T)\mu_t\varepsilon_t, \quad (8)$$

where $g(\cdot)$ is a smooth but unknown function of rescaled time.⁵ Our new class of semiparametric Dynamic Autoregressive Liquidity (DARLiQ) models therefore incorporates both a nonparametric component capturing the long-run trend of the illiquidity series and an autoregressive component capturing the short-run dynamics of illiquidity. We develop a GMM estimator based on conditional moment restrictions and an efficient semiparametric maximum likelihood estimator based on an i.i.d. assumption. We derive large sample properties for both estimators. We further develop a methodology to detect the occurrence of permanent breaks captured by potential discontinuities in the nonparametric trend function $g(t/T)$ and transitory breaks in the illiquidity process which are captured by dummy variables in the short-run dynamics of μ_t .

Finally, we demonstrate the model performance and its empirical relevance on two applications. First, we study the impact of stock splits on the illiquidity dynamics of the five largest US technology company stocks. We document that stock splits significantly decrease the long-run trend level of liquidity but not the short-run illiquidity dynamics for our panel of US technology stocks. Second, we investigate how the different components of the illiquidity process obtained from our model relate to the stock market risk premium using data on the S&P 500 stock market index. Our regression results suggest that excess returns increase with respect to the expected illiquidity component while the unexpected illiquidity shocks have a negative impact on stock prices and returns.

⁵The reader is referred to Section 4.3 of Chapter 4 for the complete specification of the model and identification restrictions.

Chapter 1

Asymmetric short-rate model without lower bound¹

1.1 Introduction

The Global Financial Crisis of 2007-2008 prompted major central banks to cut their policy rates in order to mitigate the contractionary effects of the crisis on the real economy. Faced with the inability to further decrease their policy rates to support the economic recovery due to the (Zero) Lower Bound constraint, central banks resorted to unconventional monetary policies (UMPs), such as quantitative easing and negative interest rate policies, in an attempt to further stimulate the economic recovery in a low interest rate environment. Research has mainly focused on assessing the effectiveness of UMPs in stimulating the real economy; see e.g. [Acharya et al. \(2019\)](#) for an analysis of the real effects of the ECB's Outright Monetary Transactions (OMT) programme and [Heider et al. \(2019\)](#) for a study of the transmission of negative interest rates to the real economy through the bank lending channel.

Fewer papers focus on the implications of the low interest rate environment, which results partly from UMPs implemented by central banks, for the modelling and pricing of interest rate derivatives. [Filipović et al. \(2017\)](#) introduce the linear-rational term structure framework for the pricing of swaps and European swaptions and [Filipović and Kitapbayev \(2018\)](#) extend it to American swaption pricing. The authors' contributions are however limited to an environment

¹This chapter corresponds to Vrins, F. and L. Wang (2021): "Asymmetric short-rate model without lower bound", submitted to *Quantitative Finance*.

with a strict Zero Lower Bound and thus provide limited insights on challenges associated with derivative pricing in a negative interest rate environment. [Recchioni et al. \(2017\)](#) show empirically that allowing for negative interest rates within the multi-factor Heston stochastic volatility model of [Recchioni and Sun \(2016\)](#) leads to more accurate pricing both in-sample and out-of-sample for European call and put equity options.

We make a contribution to this research agenda by developing a short-rate model which appropriately captures the salient features of the negative interest rate environment while delivering tractable semi-analytical expressions for the pricing of caplets. Mean-reverting processes such as the Vasicek model ([Vasicek \(1977\)](#)) and the Cox-Ingersoll-Ross model ([Cox et al. \(1985\)](#); CIR hereafter) are considered as standard benchmarks for the modelling of short rate dynamics. The latter was often favored before the negative interest rate period as it precludes negative values, in contrast to the Vasicek model which postulates a Gaussian distribution. Nevertheless, short-rate modelling frameworks need to be reviewed in order to better reflect the recent interest rate environment. The most direct solution would be returning to models exhibiting Gaussian dynamics, such as the Vasicek model, which do not preclude negative values. However, this approach might be at odds with empirical evidence as interest rates display conditional asymmetry in their dynamics; see e.g. [Bauer and Chernov \(2021\)](#) who document pronounced variations in US yield conditional skewness over the business cycle using option-implied skewness for future Treasury yields. Square-root dynamics, as found in the CIR model, can provide the desired positive skewness but at the cost of non-negativity, which is not appropriate in a negative rate environment. A straightforward solution is to introduce a constant negative shift to the CIR process. Nevertheless, the resulting model would exhibit a strict lower bound (the shift itself). In addition, although one could treat the shift as an additional parameter to be estimated from the data, this approach does not prescribe how to reliably estimate this lower bound parameter and guidelines from the empirical or theoretical literature are currently scarce.

In light of these limitations, this paper makes four contributions: First, we propose a two-factor model (VaCIR hereafter) which combines the desirable features of the Vasicek and CIR models. We also consider the *deterministic shift* extension of these models proposed by [Brigo and Mercurio \(2001\)](#) which guarantees a perfect fit to the term structure of interest rates. The shifted versions of the Vasicek and CIR models are known respectively as the Hull-White model ([Hull and White \(1990\)](#)) and the CIR++ model ([Brigo and Mercurio \(2001\)](#)). These two models are

then combined to obtain the shifted version of our VaCIR model which we denote as VaCIR++.

The term structure literature has previously resorted to similar two-factor models to obtain decompositions of risky interest rates into their risk-free components, modelled by a Vasicek process, and compensations for their exposures to credit and/or liquidity risks, modelled by CIR processes; see for example Longstaff et al. (2005) and Filipović and Trolle (2013). Instead, our purpose is to propose a two-factor approach that would better characterize interest rate dynamics in a negative interest rate environment. The tractability of our framework allows us to obtain semi-analytical expressions for the forward rate density and caplet pricing.

Second, we propose a novel approach to calibrate short rate models by forward rate density matching. This is achieved by minimizing the distance between the market- and model-implied densities, measured by the Jensen–Shannon (JS) divergence.² More precisely, we choose the model parameters such that the associated forward rate distribution, computed under the corresponding forward measure, is the *closest* to the one implied by derivative market prices. The key advantage over the price or volatility mean absolute error (MAE hereafter) minimization is that our approach analyzes the fit based on the full distribution rather than focusing on a few data points. An accurate calibration of the density tends to ensure a good fit in terms of prices or volatilities. However, the converse is not necessarily true.

Third, we study the evolution of the market-implied EURIBOR forward rate densities during the recent period of negative interest rates. To this end, we collect a time series of normal volatility curves of caps and apply a stripping procedure to obtain caplet prices. The model-free approach of Breeden and Litzenberger (1978) is then used in combination with the arbitrage-free smoothing technique introduced in Fong (2009) to get smooth non-negative market implied forward rate densities. These densities are the key input to derivatives pricing and will serve as the calibration target in the density matching approach we developed. Two papers propose an analysis related to our empirical study of market-implied forward rate densities: Li and Zhao (2009) study the probability density functions of future US LIBOR rates implied from caps data. Similarly, Trolle and Schwartz (2014) provide stylized facts about higher-order swap rate moments based on the implied density obtained from the swaption cube data for Europe and

²By switching the focus of the calibration approach to the minimization of the relative entropy (JS divergence) between the market- and model-implied densities, our contribution builds directly on the literature applying information theory to quantitative finance, see e.g. Brody and Hughston (2002) for an application to term structure calibration.

the US. However, both papers focus on the period preceding the low interest rate regime and their analysis does not provide relevant insights about the negative interest rate environment.

Fourth, we provide a thorough empirical study investigating the performance of the Hull-White, the CIR++ and the VaCIR++ models during the recent period of negative interest rates in the euro area. We then evaluate the performance of these models when calibrated using both the volatility MAE and density matching criteria. We find that the VaCIR++ model, which combines the features of the Hull-White and CIR++ models, consistently exhibits the best performance regardless of the target criterion. Our results further demonstrate that the JS divergence calibration approach provides a more robust way of calibrating interest rate models to fixed-income derivatives data since it has superior performance in capturing the entire forward distribution, high pricing accuracy and comparable fitting quality in terms of matching implied volatility curves. Therefore, this approach can then be used to price new, complex, and illiquid derivatives with the same underlying forward rate.

The remainder of the paper is organized as follows: Section 1.2 introduces the modelling framework which combines the Hull-White and CIR++ models into our proposed two-factor VaCIR++ model. Section 1.3 details the different steps of our calibration algorithm. Section 1.4 presents the data for our empirical application and the results. Section 1.5 concludes.

1.2 Model description

In this section, we focus on analytically tractable homogeneous short rate models existing in the literature and explain how to extend them to obtain a more flexible and general framework that will better account for the negative interest rate environment. For all three considered cases, we use the *deterministic shift* extension illustrated by Brigo and Mercurio (2001). The idea is that by shifting a latent model with time-homogeneous dynamics with a deterministic function $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$, one obtains a short-rate model that is analytically tractable and able to perfectly fit to the term structure of interest rates. Moreover, the shift function ϕ is directly given by the difference between the instantaneous forward rate curve implied from the market and that generated by the latent process, possibly available in closed-form.

We start with two classic short rate models which will serve as benchmarks. The obvious choice is the Hull-White model. We consider an Ornstein-Uhlenbeck (Gaussian) latent process,

and shift it with the function ϕ . The other natural candidate is the CIR model. Applying a deterministic shift function to the latter yields the CIR++ model. This model features skewness and is compatible with the negative rates environment thanks to the shift. Note that this model exhibits a strict lower bound up to any time horizon T , given by the minimum of the function ϕ on the interval $[0, T]$. The alternative we propose is a combination of the aforementioned models, in an attempt to address their documented shortcomings while preserving the desired features.

1.2.1 Baseline framework

We postulate a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ where $\mathbb{Q}$ is a probability measure and $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$ the filtration, satisfying the usual conditions. All processes considered in this paper are assumed to be $\mathbb{F}$ -adapted.

We consider two standard interest-rate models that we take as benchmark, namely the Vasicek model (noted x hereafter) and the CIR model (noted y hereafter). The short-rate dynamics are respectively given by

$$dx_t = \kappa_1(\theta_1 - x_t)dt + \sigma_1 dW_t^x, \quad x_0 \in \mathbb{R}, \quad (1.1)$$

$$dy_t = \kappa_2(\theta_2 - y_t)dt + \sigma_2 \sqrt{y_t} dW_t^y, \quad y_0 \in \mathbb{R}_0^+. \quad (1.2)$$

where W^x and W^y are two independent Brownian motions under $\mathbb{Q}$.

These models belong to the class of homogeneous one-factor affine diffusions, with dynamics³

$$dz_t = (a + bz_t)dt + \sqrt{c + dz_t} dW_t, \quad z_0 > c/d, \quad \forall d > 0, \quad (1.3)$$

where W is a Brownian motion. We denote the parameters as $\Psi = (a, b, c, d, z_0)$. The stochastic differential equations (SDEs) associated with the Vasicek and CIR models are recovered from Equation (1.3) using respectively $\Psi^x := (\kappa_1\theta_1, -\kappa_1, \sigma_1^2, 0, x_0)$ and $\Psi^y := (\kappa_2\theta_2, -\kappa_2, 0, \sigma_2^2, y_0)$.⁴

Homogeneous affine diffusions, as in Equation (1.3), are popular due to their analytical tractability which simplifies model calibration. For instance, denoting by $\mathbb{E}$ the expectation

³While we focus the exposition on the univariate case, multi-dimensional jump-diffusion models can be considered.

⁴In these expressions, the constraints on the parameters are as follows: $\kappa_1, \sigma_1, y_0, \kappa_2, \theta_2, \sigma_2$ are positive constants and $2\kappa_2\theta_2 > \sigma_2^2$ (Feller condition).

under $\mathbb{Q}$, it is well-known that for such processes,

$$P_t^z(T) := \mathbb{E} \left[e^{-\int_t^T z_s ds} \middle| \mathcal{F}_t \right] = A^z(T-t) e^{-B^z(T-t)z_t}, \quad (1.4)$$

where A^z, B^z are deterministic functions satisfying Riccati equations, for which analytical solutions can be found.⁵

When $\mathbb{Q}$ stands for the risk-neutral measure, $\mathbb{F}$ is the information available to investors and z depicts the short-rate process, $P_t^z(T)$ agrees with $P_t(T)$, the no-arbitrage price at time t of a risk-free zero-coupon bond with maturity $T \geq t$:

$$P_t(T) := \mathbb{E} \left[\frac{M_t^z}{M_T^z} \middle| \mathcal{F}_t \right] = \mathbb{E} \left[e^{-\int_t^T z_s ds} \middle| \mathcal{F}_t \right] = P_t^z(T), \quad (1.5)$$

where $M^z := (M_t^z)_{t \geq 0}$ and $M_t^z := e^{\int_0^t z_s ds}$ is the money market account numéraire.

The existence of analytical expressions for zero-coupon bond prices explains why the Vasicek and CIR models are so popular for interest rates modelling. Furthermore, their deterministic shift extensions make it possible to replicate the observed term structure of discount factors at time t . More specifically, the short rate is expressed as $z_t^\phi = z_t + \phi^z(t)$, $z \in \{x, y\}$. The corresponding zero-coupon bond price takes the form

$$P_t^{z^\phi}(T) = \exp \left\{ - \int_t^T \phi^z(s) ds \right\} P_t^z(T). \quad (1.6)$$

One can then choose the function ϕ^z such that the discount curve implied by the z^ϕ model perfectly agrees with the market discount curve, i.e., $P_t^{z^\phi}(T) = P_t^M(T)$. It can be shown that this choice corresponds to

$$\phi^z(t) = f_0^M(t) - f_0^z(t), \quad (1.7)$$

$$f_0^x(t) = \left(\frac{\sigma_1^2}{2\kappa_1^2} e^{-\kappa_1 t} + \frac{\kappa_1^2 \theta_1 - \sigma_1^2/2}{\kappa_1^2} \right) (1 - e^{-\kappa_1 t}) + x_0 e^{-\kappa_1 t}, \quad (1.8)$$

$$f_0^y(t) = \frac{2\kappa_2 \theta_2 (e^{th} - 1)}{2h + (\kappa_2 + h)(e^{th} - 1)} + y_0 \frac{4h^2 e^{th}}{[2h + (\kappa_2 + h)(e^{th} - 1)]^2}, \quad (1.9)$$

where $f_t^M(\cdot)$ is the time- t instantaneous forward curve associated with the current market term structure, $f_t^z(\cdot)$ is the corresponding curve implied by the model $z \in \{x, y\}$ and $h := \sqrt{\kappa_2^2 + 2\sigma_2^2}$.

⁵They are recalled in [section 1.A](#) for $z \in \{x, y\}$.

1.2.2 Proposed extension

Recall that our goal is to design a model that would comply with negative interest rates. As discussed in the introduction, the Vasicek model (x) has the ability to deal with negative values, but at the expense of assuming a Normal distribution for the short rate. The shifted CIR model may comply with the negative interest rate environment whenever the shift is negative. However, as explained above, such a process would exhibit a time-dependent lower bound (the shift function itself) which would be associated with additional challenges on how to reliably estimate a meaningful lower bound on interest rates from available data.

The model we propose aims at capturing skewness without imposing a strict lower bound by combining two latent processes: a Vasicek process and a CIR process. We denote the resulting model VaCIR hereafter. Mathematically, we consider the following dynamics for the short rate in this model:

$$r_t := x_t + y_t, \quad (1.10)$$

where the Brownian motions W^x, W^y in Equation (1.1) and Equation (1.2) are assumed to be independent. This assumption preserves the analytical tractability of zero-coupon bond prices

$$P_t^r(T) = P_t^x(T)P_t^y(T) = A^x(T-t)A^y(T-t)e^{-[B^x(T-t)x_t + B^y(T-t)y_t]}. \quad (1.11)$$

Similarly, we extend the combined VaCIR model to perfectly replicate the current observed discount curve (denoted VaCIR++ hereafter), i.e., $r_t^\phi = r_t + \phi^r(t)$, where $\phi^r(t) = \phi^x(t) + \phi^y(t) - f_0^M(t)$. Additionally, the zero-coupon bond is defined in the same way as in Equation (1.6). Introducing correlation between the two latent processes would come at the cost of analytical tractability of A, B since the resulting model would not be part of the affine model class, see [section 1.C](#) for details. One consequence of assuming independence between x and y is that our proposed model might yield unsatisfactory calibrations results for the term structure of caplet volatilities exhibiting a large hump shape. As the main focus of our analysis is not the calibration of the market-implied volatility surface, we leave these considerations for future research and refer the interested reader to Chapter 4 of [Brigo and Mercurio \(2007\)](#) for more details.

1.2.3 Forward rate dynamics and densities under the forward measure

Due to the central role played by the forward rate in the price of derivative products (e.g., caps and floors), we derive the dynamics and densities related to the forward rate in the various

models⁶

$$F_t^z(T, S) = \frac{1}{\Delta} \left(\frac{P_t^z(T)}{P_t^z(S)} - 1 \right), \quad z \in \{x, y, r\}. \quad (1.12)$$

For convenience, we formulate the dynamics of $F(T, S) = (F_t(T, S))_{0 \leq t \leq T < S}$ under the S -forward measure, $\mathbb{Q}^S$.

Proposition 1. *Let $t \leq T < S$ and $\Delta := S - T$. Then,*

(a) *The forward rate $F^r(T, S)$ associated with the model specified in Equation (1.10) reads as*

$$F_t^r(T, S) = \frac{1}{\Delta} \left((1 + \Delta F_t^x(T, S))(1 + \Delta F_t^y(T, S)) - 1 \right). \quad (1.13)$$

Moreover, let $\sigma_t^x := \sigma_1$, $\sigma_t^y := \sigma_2 \sqrt{y_t}$ be the diffusion coefficients associated with models x and y , respectively. For $z \in \{x, y\}$, define $W_t^{z,S} := W_t^z + \int_0^t \zeta_s^{z,S} ds$ with $\zeta_t^{z,S} := \sigma_t^z B^z(S - t)$. Define $\mathbb{Q}^S$ using the random variable

$$\left. \frac{d\mathbb{Q}^S}{d\mathbb{Q}} \right|_{\mathcal{F}_S} = \frac{P_S^r(S) M_0^r}{P_0^r(S) M_S^r} = \frac{P_S^r(S)}{P_0^r(S) M_S^r}.$$

Then,

(b) $F^x(T, S)$, $F^y(T, S)$ and $F^r(T, S)$ are $\mathbb{Q}^S$ -martingales;

(c) *The dynamics of the forward rates associated with x and y are given by*

$$dF_t^x(T, S) = \sigma^x(t, F_t^x(T, S)) dW_t^{x,S}, \quad dF_t^y(T, S) = \sigma^y(t, F_t^y(T, S)) dW_t^{y,S}, \quad (1.14)$$

where, $W^{x,S}, W^{y,S}$ are independent $\mathbb{Q}^S$ -Brownian motions and

$$\begin{aligned} \sigma^x(t, z) &:= \sigma_1 \left[z + \frac{1}{\Delta} \right] (B^x(S - t) - B^x(T - t)), \\ \sigma^y(t, z) &:= \sigma_2 \left[z + \frac{1}{\Delta} \right] \sqrt{(B^y(S - t) - B^y(T - t)) (G_t(T, S, z))}, \\ G_t(T, S, z) &:= \ln \left([1 + \Delta z] \frac{A^y(S - t)}{A^y(T - t)} \right). \end{aligned}$$

Proof: see [section 1.D](#).

Because these processes are independent, the conditional density of r_t given (x_s, y_s) is obtained by convolution of the corresponding conditional densities for the Vasicek and the CIR processes sampled at t . The $\mathbb{Q}^S$ -conditional densities of x and y can be found in [Brigo and Mercurio \(2007\)](#).

⁶Of course, with the deterministic shift extension, the expression for forward rates should also be adjusted for the shift term, which can be re-written as $F_t^{z^\phi}(T, S) = \frac{1}{\Delta} \left(e^{-\int_T^S \varphi^z(s) ds} \frac{P_t^z(T)}{P_t^z(S)} - 1 \right)$, $z \in \{x, y, r\}$.

Proposition 2. *Let $t \leq T \leq S$. Under $\mathbb{Q}^S$, the density of the spot rate z_T^ϕ conditional on z_t^ϕ , for $z \in \{x, y, r\}$, is given by ,*

$$f_{z_T^\phi|z_t^\phi}^S(u) = f_{z_T|z_t}^S(u - \phi^z(t)) . \quad (1.15)$$

Note that when $z = r$, the condition $z_T|z_t$ becomes $r_T|(x_t, y_t)$.

(a) the density of x_T conditional on x_t is given by

$$f_{x_T|x_t}^S(u) = \frac{1}{\sigma^x(T-t)} \varphi\left(\frac{u - \mu^x(t, T)}{\sigma^x(T-t)}\right) , \quad (1.16)$$

$$\mu^x(t, T) = x_t e^{-\kappa_1(T-t)} + M^S(t, T) ,$$

$$\sigma^x(\tau) = \sqrt{\frac{\sigma_1^2}{2\kappa_1} [1 - e^{-2\kappa_1\tau}] } ,$$

$$M^S(t, T) = \left(\theta_1 - \frac{\sigma_1^2}{\kappa_1^2}\right) (1 - e^{-\kappa_1(T-t)}) + \frac{e^{-\kappa_1(S-T)}}{\kappa_1} (\sigma^x(T-t))^2 ,$$

where φ is the standard Normal density.

(b) the density of y_T conditional on y_t is given by

$$f_{y_T|y_t}^S(u) = f_{\chi^2(\nu, \delta(t, T, S))/q(t, T, S)}(u) = q(t, T, S) f_{\chi^2(\nu, \delta(t, T, S))}(q(t, T, S)u) , \quad u \geq 0 , \quad (1.17)$$

where $f_{\chi^2(\nu, \delta)}$ is the density of a non-central chi-squared random variable with ν degrees of freedom and non-centrality parameter δ . We introduced the following notation

$$q(t, T, S) := 2 \left(\rho(T-t) + \frac{\kappa_2 + h}{\sigma_2^2} + B^y(\Delta) \right) ,$$

$$\delta(t, T, S) := \frac{4\rho(T-t)^2 y_t e^{h(T-t)}}{q(t, T, S)} ,$$

$$\rho(\tau) := \frac{2h}{\sigma_2^2 (\exp[h\tau] - 1)} ,$$

with $\nu := 4\kappa_2\theta_2/\sigma_2^2$ and $h := \sqrt{\kappa_2^2 + 2\sigma_2^2}$.

(c) the density of r_T conditional on (x_t, y_t) is given by

$$f_{r_T|(x_t, y_t)}^S(u) = \int_0^\infty f_{x_T|x_t}^S(u-v) f_{y_T|y_t}^S(v) dv . \quad (1.18)$$

From the expressions in Proposition 2, we can now provide a characterization of the conditional density of the zero-coupon bond price under the S -forward measure.

Proposition 3. *Let $t \leq T \leq S$. Under $\mathbb{Q}^S$, the density of the zero-coupon bond price, $P_T^{z^\phi}(S)$, conditional upon z_t^ϕ , for models of $z \in \{x, y, r\}$ with shift extensions, is given by*

$$f_{P_T^{z^\phi}(S)|z_t^\phi}^S(u) = e^{\int_t^T \phi^z(s) ds} f_{P_T^z(S)|z_t}^S\left(u e^{\int_t^T \phi^z(s) ds}\right) , \quad u \geq 0 . \quad (1.19)$$

Note that when $z = r$, the condition $P_T^z(S)|z_t$ becomes $P_T^r(S)|(x_t, y_t)$.

(a) the density of the zero-coupon bond price $P_T^z(S)$ conditional upon z_t , for the Vasicek model ($z = x$) and the CIR model ($z = y$) are given by

$$f_{P_T^z(S)|z_t}^S(u) = \frac{1}{uB^z(\Delta)} f_{z_T|z_t}^S \left(\frac{1}{B^z(\Delta)} \ln \frac{A^z(\Delta)}{u} \right), \quad (1.20)$$

where $u \geq 0$ for the Vasicek model and $0 \leq u \leq 1$ for the CIR model.

As for the combined model, it reads as

$$f_{P_T^z(S)|\mathcal{F}_t}^S(u) = \int_0^1 \frac{1}{v} f_{P_T^x(S)|x_t}^S(u/v) f_{P_T^y(S)|y_t}^S(v) dv, \quad u \geq 0. \quad (1.21)$$

(b) the density of the spot rate $L^{z^\phi}(T, S) := F_T^{z^\phi}(T, S)$ conditional upon $\mathcal{F}_t = \mathcal{F}_t^x \vee \mathcal{F}_t^y$,⁷ $z \in \{x, y, r\}$, is given by

$$f_{L^{z^\phi}(T, S)|\mathcal{F}_t}^S(u) = \frac{\Delta}{(1 + \Delta u)^2} f_{P_T^{z^\phi}(S)|\mathcal{F}_t}^S \left(\frac{1}{1 + \Delta u} \right). \quad (1.22)$$

1.3 Model calibration

Calibrating a short rate model z to the market amounts to (i) select a set of n financial products, called calibration instruments hereafter, and (ii) look for the set of parameters Ψ^z minimizing the discrepancies between the market- and model-implied prices of the calibration instruments. The time- t no-arbitrage price of a caplet on the spot rate $L(T, S) = F_T(T, S)$, with strike K , fixing date T and payment date $S > T$ can be obtained from the $\mathbb{Q}^S$ density of $L(T, S)$:

$$\begin{aligned} \text{Cpl}_t(T, S, K) &= M_t \mathbb{E} \left[\frac{(L(T, S) - K)^+ \Delta}{M_S} \middle| \mathcal{F}_t \right] \\ &= P_t(S) \mathbb{E}^S \left[(L(T, S) - K)^+ \Delta \middle| \mathcal{F}_t \right] \\ &= \Delta P_t(S) \int_K^{+\infty} (v - K) f_{L(T, S)|\mathcal{F}_t}^S(v) dv, \end{aligned} \quad (1.23)$$

where $P_t(S)$ is the price of the zero-coupon bond with maturity S at time t and $f_{L(T, S)}^S$ is the density of the spot rate $L(T, S)$ provided in Equation (1.22). The model prices are obtained by evaluating the right-hand side of Equation (1.23), replacing $f_{L(T, S)}^S$ by $f_{L^z(T, S)}^S$ in Proposition (3) for each of the considered models. Calibration based on the minimization of the mean absolute error of implied volatilities is often favored in practice and we cover this approach in Section 1.3.1. In Section 1.3.2, we introduce a novel calibration technique whose aim is to minimize the

⁷Since x and y are independent, $L^x(T, S)|\mathcal{F}_t$ is equivalent to $L^x(T, S)|x_t$, $L^y(T, S)|\mathcal{F}_t$ is equivalent to $L^y(T, S)|y_t$ and $L^r(T, S)|\mathcal{F}_t$ is equivalent to $L^r(T, S)|(x_t, y_t)$.

distance between the model-implied density, $f_{L^z(T,S)}^S$, and the market-implied one, $f_{L(T,S)}^S$.

1.3.1 Calibration of the implied volatility curve

The caplet pricing equation provides us with the formula to compute the caplet prices under the three considered models by plugging in the density function specified in Equation (1.22). Under the assumption of normal volatility and the Bachelier formula, we can obtain the implied volatility associated with caplet products. Based on the implied volatility, the considered models are calibrated to the market caplet volatility surface, and the fitting performance is measured in terms of the mean absolute error of implied volatilities. We thus calibrate each of the considered models by looking for the parameters Ψ^z which minimize the MAE between the model-implied and market-implied volatilities:

$$\Psi^{z*} := \arg \min_{\Psi^z} \frac{1}{n} \sum_{i=1}^n \left| \sigma_i^M - \sigma_i^z \right| \quad (1.24)$$

where σ_i^M and σ_i^z are the market- and model-implied volatilities for caplet i , respectively.

1.3.2 Calibration of the forward rate density

The caplet pricing equation allows us to infer the implied forward rate density from market data in a model-free way.⁸ Indeed, differentiating twice the expression for the caplet price in Equation (1.23) with respect to the strike K , we obtain

$$f_{L(T,S)|\mathcal{F}_t}^S(K) = \frac{\frac{\partial^2}{\partial K^2} \text{Cpl}_t(T, S, K)}{\Delta P_t(S)}. \quad (1.25)$$

In order to get an accurate estimation of the density based on Equation (1.25), we need to obtain the market discount curve and the price quotations for caplets on a fine grid of strikes. Since only a limited number of products are actively traded and quoted on the fixed income market, we have to resort to interpolation methods. Special care needs to be taken in order to obtain a valid (i.e., non-negative) and smooth density.

To this end, we use the arbitrage-free smoothing technique of the implied volatility surface developed by [Fengler \(2009\)](#) which we apply to the caplet prices. The procedure takes the quoted caps data and obtain the caplet prices via caplet stripping. Then, a cubic spline smoothing technique is applied to caplet prices, augmented with appropriately chosen linear constraints

⁸See [Breedon and Litzenberger \(1978\)](#) for a seminal contribution and [Brigo and Mercurio \(2007\)](#) for a textbook treatment in the context of caplet pricing.

to enforce no-arbitrage restrictions. By eliminating arbitrage opportunities, this approach precludes the occurrence of negative local volatilities or negative transition probabilities and thus ensures that the forward rate density is well specified.⁹ The method proposed by [Fengler \(2009\)](#) is closely related to the literature on non-parametric estimation of risk-neutral transition densities under shape constraints ([Aït-Sahalia and Duarte \(2003\)](#); [Yatchew and Härdle \(2006\)](#)). We decided to adopt it instead of other competing approaches due to its ease of implementation; its good theoretical properties - inherited from the theory on natural smoothing splines under suitable shape constraints; and its applicability in contexts with scarce calibration data.

In order to calibrate the model-implied density to match the market-implied one, we need a criterion assessing how far the model-implied density is from the reference one. This problem has been extensively studied in the field of information theory which has already been applied to relevant research areas in quantitative finance, see e.g. [Brody and Hughston \(2002\)](#) for an application of the Shannon entropy to term structure modelling. We build on this existing literature and select the measure based on Kullback-Leibler (KL) divergence as our calibration criteria.^{10,11} Specifically, the KL divergence between two densities (p, q) is the relative entropy, and is based on (differential) Shannon's entropy:

$$\langle p|q \rangle := \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx, \quad (1.26)$$

where p is assumed to be absolutely continuous with respect to q .

Note that, given the asymmetry of the KL divergence, $\langle p|q \rangle$ will typically be different from $\langle q|p \rangle$. As a consequence, basing the optimization program on only one of them would result in different outcomes. Additionally, optimization based on the KL divergence might face some challenges in practice as the KL divergence goes to infinity when p is not absolutely continuous with respect to q . In our application, we circumvent these issues by considering the Jensen–Shannon (JS) divergence which is a symmetrized and smoothed version of the KL divergence. Specifically,

⁹A detailed outline of this procedure is provided in Appendix 1.B.

¹⁰The KL divergence is not a distance because the triangular inequality may fail to hold. See [Basseville \(2013\)](#) for a review of statistical applications using divergence measures.

¹¹We opt for the KL divergence compared to the L^2 metric for density calibration as this measure does not properly scale with the distribution – exacerbating (resp. minimising) discrepancies between densities in high (resp. low) density regions – which would impact calibration accuracy. Other choices, such as e.g. the Hellinger distance, could be entertained and are left for future research.

the JS divergence between two densities (p, q) is defined as

$$\text{JSD}(p|q) = \frac{1}{2} \langle p|m \rangle + \frac{1}{2} \langle q|m \rangle, \quad (1.27)$$

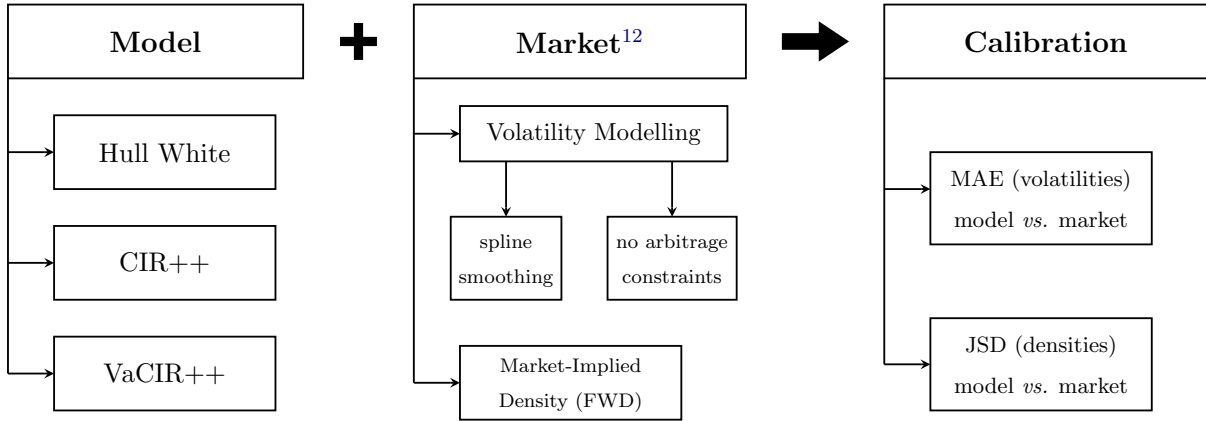
where $m = \frac{1}{2}(p + q)$. By comparing both densities to their average, the JS divergence is always ensured to be bounded as the denominator in the log is always nonzero whenever the numerator is nonzero. Therefore, the JS divergence can be used to compare densities with different supports. Additionally, the measure is bounded between zero and $\log 2$ and it reaches the lower bound only when $p = q$ almost everywhere.

Equipped with a measure accurately quantifying the goodness of fit between densities, we can calibrate each of the considered models by choosing as parameters the set Ψ^z which minimize the JS divergence between the model-implied and the market-implied densities

$$\Psi^{z*} := \arg \min_{\Psi^z} \text{JSD}(f_{L(T,S)}^S | f_{L^z(T,S)}^S) \quad (1.28)$$

where $f_{L(T,S)}^S$ and $f_{L^z(T,S)}^S$ are respectively the market- and model-implied forward rate densities.

1.3.3 Summary of Calibration Algorithm



As summarized in the diagram 1.3.3, we use Algorithm 1 and Algorithm 2 to calibrate the model over a discrete grid of strikes $K \in \{K_0, K_1, K_2 \cdots K_{N-1}, K_N\}$ for certain lower and upper bounds of K_0 and K_N .

¹²Volatility modelling and market-implied density are only needed under the calibration method using JS divergence. The calibration method minimizing the mean absolute error is based on the implied volatility data directly obtained from caplet stripping.

Algorithm 1 Calibration of the implied (normal) volatility curve

- 1: Calculate the caplet prices under the chosen model (Hull-White, CIR++, or VaCIR++) according to Equation (1.23).
 - 2: Obtain the market implied (normal) volatilities for the corresponding caplet products.
 - 3: Use mean absolute error as the target criterion to calibrate the three models by matching the market volatility curve according to Equation (1.24).
-

Algorithm 2 Calibration of the forward rate density

- 1: Calculate the model-implied density of the forward rate under the chosen model (Hull-White, CIR++, or VaCIR++) according to Proposition 2 and Proposition 3.
 - 2: Compute the market-implied density for the forward rate.
 - Retrieve caplet prices from market cap data via caplet stripping.
 - Implement arbitrage-free smoothing technique of volatility according to the algorithm presented in section 1.B.
 - Obtain the market-implied density over a fine grid according to Equation (1.25).
 - 3: Use JS divergence as the target criterion to calibrate the three models by matching the market-implied density according to Equation (1.28).
-

1.4 Empirical application

In this section, we present an empirical application where we conduct a thorough performance comparison of the three competing models (namely, Hull-White, CIR++, and VaCIR++) under the two calibration schemes introduced in Section 1.3. More specifically, we take the 6-month forward rate as an example to evaluate the pricing accuracy and the quality of fitting the market implied volatility curve. We further investigate the distribution of the forward rate in 6 months and illustrate the model performance in terms of density matching under the three considered frameworks in the negative interest rate environment. The rest of this section is organized as follows: Section 1.4.1 presents the data description and sample coverage used for our empirical application and illustrates the non-arbitrage smoothing technique implemented to construct the market-implied density curve; Section 1.4.2 covers the calibration results based on the minimization of implied volatility errors; Section 1.4.3 covers the calibration results based on the minimization of JS divergence where the modelled forward density is matched to the market-implied one.

1.4.1 Data description and volatility curve smoothing

We use historical data for the 6-month EURIBOR forward rate $F_0(6M, 1Y)$ and the interest rate derivatives whose underlying rate is the 6-month forward rate in 6 months, i.e. $L(6M, 1Y) = F_{6M}(6M, 1Y)$.¹⁴ Observe that the market provides quotations for the implied volatilities of caps whereas, in order to compute the market-implied density for the forward rate $L(6M, 1Y)$, we need caplet prices associated with the latter.

Therefore, we start by collecting monthly normal volatility quotes from Refinitiv (formerly known as Eikon, Thomson Reuters) for the 6-month EURIBOR caps quoted over a grid composed of 13 different strikes. The sample period spans from March 31, 2016 to January 29, 2021 during which the 6-month EURIBOR and 6-month forward rate are in the negative territory. Then we use Bachelier formula to get the cap prices, and apply the stripping approach proposed in Hagan and Konikov (2004) to extract caplet prices.

The market implied volatilities for caplets are obtained by inverting the Bachelier formula. The implied forward rate density can be computed from Equation (1.25), which involves the

¹⁴The discount curve and all the related interest derivatives are based on the 6-month EURIBOR rates for consistency.

numerical estimation of the second-order derivatives. Thus, caplet prices for a given fine grid of strikes are required to guarantee an accurate approximation. As explained in Section 1.3.2, we follow the procedure introduced in Fengler (2009) to ensure a well-specified forward rate density.

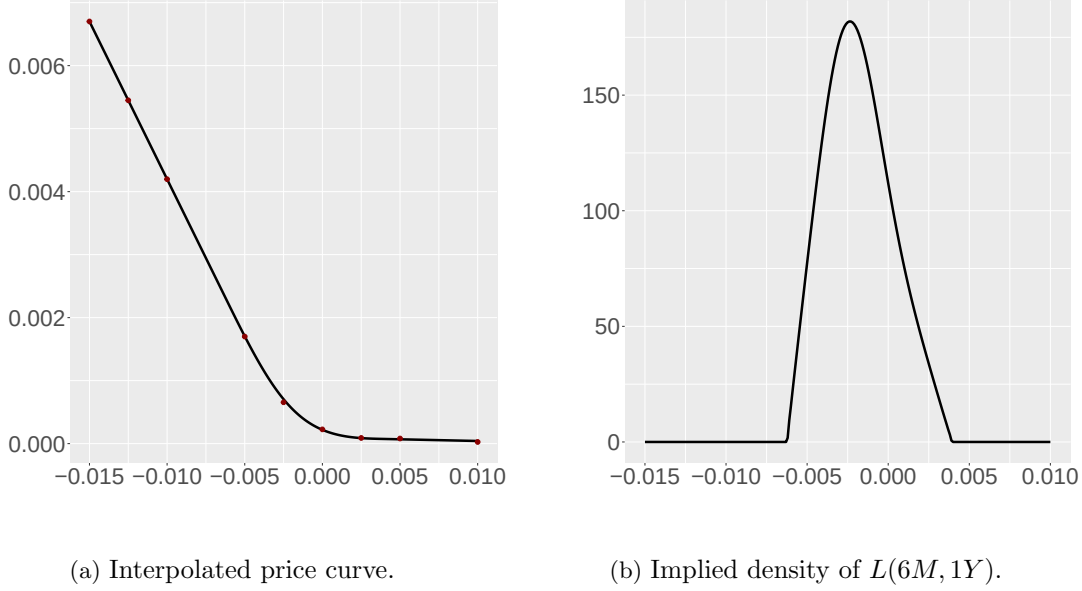


Figure 1.1: Interpolated caplet prices and market-implied probability density under $\mathbb{Q}^S$ constructed based on the cap market data (quoted in normal volatility) on March 31, 2016. Panel (a) gives the market caplet prices (dots) computed by Bachelier formula alongside with interpolated price curve obtained by no-arbitrage cubic spline smoothing technique. Panel (b) provides the $L(6M, 1Y)$ forward rate density implied from the interpolated caplet prices using Equation (1.25).

In Figure 1.1, we illustrate the performance of the smoothing approach from two perspectives, the fit of market prices and the behavior of the probability density. Figure 1.1a presents the interpolated prices and the market quoted prices for the caplet products of interest, $\text{Cpl}(0, 6M, 1Y)$. It can be observed that all market quotations are in line with the smoothed price curve, indicating that the interpolation technique is performing well as an accurate caplet valuation tool. Figure 1.1b shows that the market-implied density is non-negative, and is positive in the range from approximately -0.7% to 0.3% . This indicates that the 6-month forward rate in 6 months is distributed in a quite narrow range and the probability of the potential value is the highest around -0.22% which is consistent with the level of the current forward rate. This illustrates the property that the expectation under the corresponding forward measure of the future forward rate matches the current forward rate.

1.4.2 Calibration results for the implied volatility curve

Equation (1.23) provides the model-implied caplet prices. The corresponding caplet volatilities, obtained via Bachelier formula, can thus be compared with those extracted from market quotes as explained in Section 1.3.1. The model is then calibrated according to Equation (1.24). We provide the calibration performance on the first available date from our sample as an illustration.

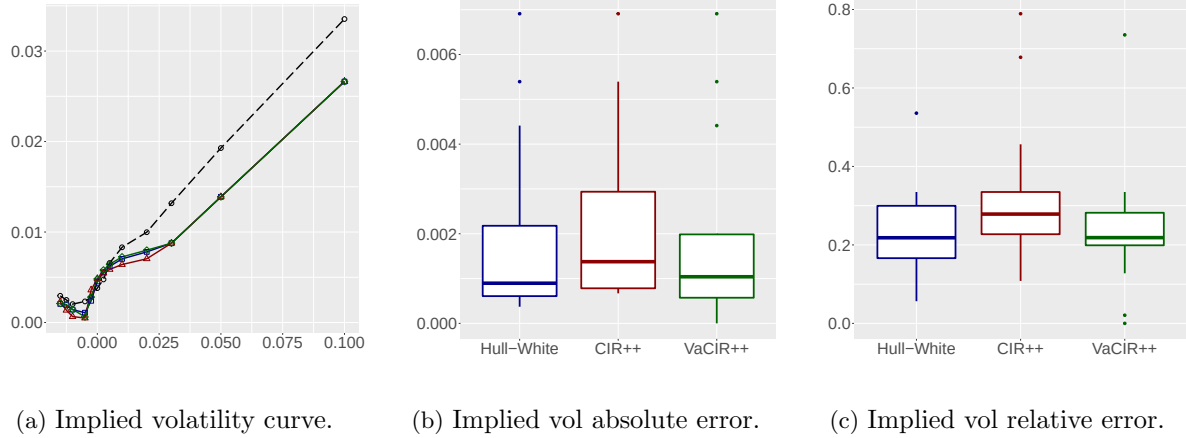


Figure 1.2: Implied volatility curves with associated absolute and relative volatility errors on 31 March 2016 where the models are calibrated on the criteria of implied volatility error minimization. Panel (a) shows the implied volatility curves with respect to the strike for the market (dashed), Hull-White ($\square$), CIR++ ($\triangle$) and VaCIR++ ($\diamond$) models. Panel (b) and Panel (c) provide respectively the box plots of the associated absolute and relative errors for the three candidate models.

We plot in Figure 1.2 the modelled volatilities with associated absolute and relative errors using the parameters calibrated based on fitting the market-implied volatilities. Figure 1.2a shows the implied volatility curves for each of the three calibrated models as well as the implied volatilities extracted from stripped market quotes, which serve here as calibration target. We observe that all three considered models capture quite well the in/at the money caplet volatilities but fail to capture accurately the data of the deep out of the money products. The box plots of the associated absolute and relative errors, displayed in Figure 1.2b and Figure 1.2c respectively, provide further details on the relative performance of the three considered candidates. The CIR++ model has the largest median in terms of the volatility absolute and relative errors while the error distribution of the VaCIR++ model is close to the Hull-White model but exhibits less dispersion. On this date, the Hull-White model outperforms the CIR++ model, but this is not always the case. For instance, the opposite situation happens on 29 January 2021 (see Figure 1.10 in Appendix 1.E). The persistent outperformance of the VaCIR++ model over both component

processes illustrates the ability of the framework to accommodate different circumstances and to deliver a better fit of the market data.

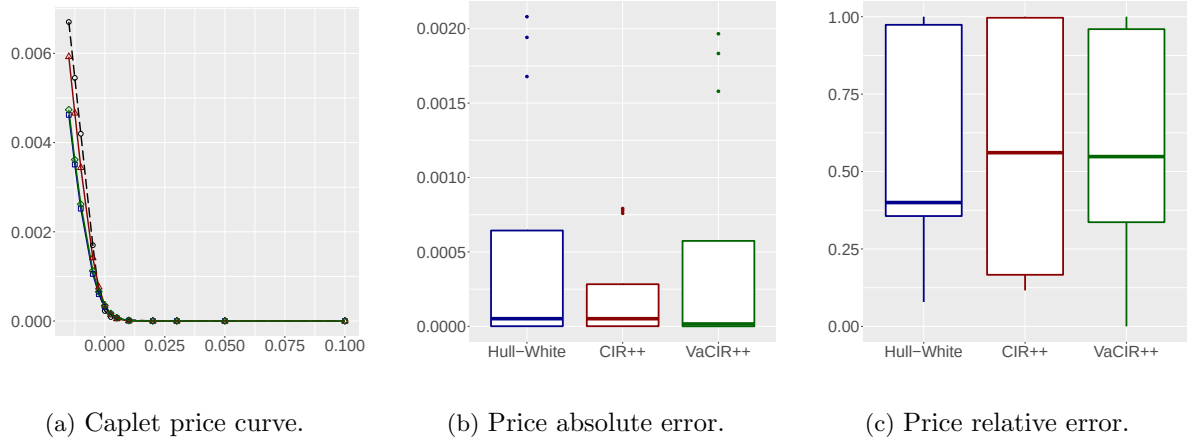


Figure 1.3: Caplet price curves with corresponding absolute and relative pricing errors on 31 March 2016 where models are calibrated on the criteria of implied volatility error minimization. Panel (a) displays the market and modelled caplet prices with respect to the strike for the market (dashed), Hull-White ($\square$), CIR++ ($\triangle$) and VaCIR++ ($\diamond$) models. Panel (b) and panel (c) provide respectively the box plots of the corresponding absolute errors and relative errors for the three candidate models.

Figure 1.3 shows the same results as Figure 1.2 but expressed in terms of prices rather than volatilities. It can be observed that although the Hull-White model provides a better fit for the implied volatility curve, this does not directly translate into higher pricing accuracy. In particular, the Hull-White model has the highest value in both absolute and relative errors while the CIR++ model and the VaCIR++ model provide, respectively, the best performance in terms of the absolute and the relative errors.

Not surprisingly, Figure 1.3a indicates that a small implied volatility error for the in/at the money caplets can lead to a sizable discrepancy in prices while the out of the money caplets are subject to large differences in volatility but have almost indistinguishable price curves. This observation highlights the drawbacks of the calibration approaches fitting the implied volatilities or caplet prices. More specifically, the far out of the money options tend to have extremely low prices but rather high implied volatilities. Therefore, putting equal weight on the error of those products in the calibration target could sacrifice the accuracy in modelling the exact prices and volatilities for the at/in the money products. Thus, a weighting scheme should be incorporated such that the information in such options would be substantially discounted. However, this approach would still be essentially focusing on a subset of products and would likely result in a

poor overall fit.

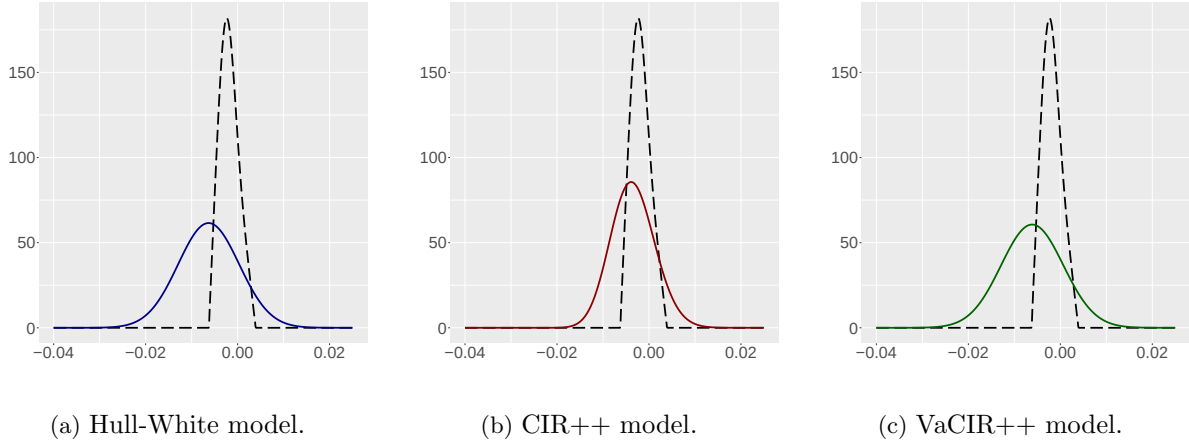


Figure 1.4: Density curves of the forward rate $L(6M, 1Y)$ on 31 March 2016 where models are calibrated on the criteria of implied volatility error minimization. Panel (a), panel (b) and panel (c) exhibit the market-implied forward density curve and the modelled forward density curves (dashed lines) and under the Hull-White, the CIR++ and VaCIR++ model, respectively (solid lines).

We further investigate the distribution of the forward rate under the three considered models compared to the market-implied one using the same parameters. [Figure 1.4](#) displays respectively the density curves modelled by the Hull-White, the CIR++ and the VaCIR++ models against the market-implied one. We observe that in general, the three models have a higher peak and their densities approach zero at a slower speed when the strikes decrease to deeply negative values or increase to highly positive rates. Moreover, it can be observed from [Figure 1.4a](#) to [Figure 1.4c](#) that the densities under all three models, compared to the market density, are more flat with a shift to the left side. This observation suggests that the calibration approach focusing on a few data points of the implied volatilities does not guarantee the match of the entire forward rate distribution.

1.4.3 Calibration results for the forward rate density

Equipped with the smoothed caplet price curve computed over a fine grid of strikes, we now turn to the estimation of the market-implied forward rate distribution according to Equation (1.25). We then calibrate the three considered models according to Algorithm 2 to closely match the market-implied density by minimizing the discrepancy (measured by JS divergence) between the model-implied and the market-implied curves. As in the previous section, we take the first available date from our sample and investigate the distribution of the forward rate under the

three candidates compared to the market-implied one.

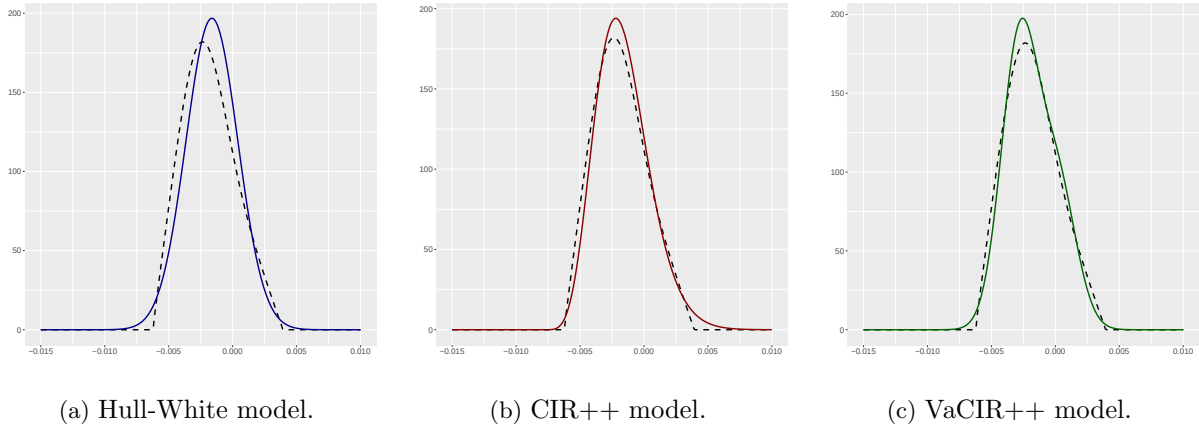


Figure 1.5: Density curves of the forward rate $L(6M, 1Y)$ on 31 March 2016 where models are calibrated on the criteria of JS divergence minimization. Panel (a), panel (b) and panel (c) exhibit the market-implied forward density curve (dashed lines) and the modelled forward density curves under the Hull-White, the CIR++ and the VaCIR++ model, respectively (solid lines).

We plot in Figure 1.5 the density curves obtained by calibrating each of the three models (solid lines). Figure 1.5a shows that the calibrated density under the Hull-White model has a right-shifted distribution and is more symmetric, compared to the market-implied one. Furthermore, it provides a closer fit to the market-implied density on the right-hand side but deviates significantly on the left-hand side. In addition, we plot in Figure 1.5b the density under the CIR++ model, we observe that it has a slightly higher peak than the market-implied density but it captures quite well the skewness exhibited by the market density. In particular, it provides an almost perfect fit on the left-hand side although it fails to match the right-hand tail in the forward rate distribution. Lastly, the calibrated density obtained from the VaCIR++ model is displayed in Figure 1.5c, and we conclude that the combined model matches well both tails and captures the overall market-implied density curve better than the Hull-White and CIR++ models. Furthermore, the superior performance of the VaCIR++ is confirmed by the significantly lower JS divergence value compared to the alternative models (see Figure 1.8). This observation is consistent with the argument that the VaCIR++ model will perform at least as well as its underlying component processes and it will be inclined to place a higher weight on the component process exhibiting better calibration performance, the CIR++ process in this case. It is worth noting that the model-implied densities obtained from the density matching calibration approach provide a tighter fit to the market-implied one compared to the results displayed

in Figure 1.4 which are based on the implied volatility MAE criterion. Thus far, our analysis focused on benchmarking our two-factor VaCIR++ model with the one-factor Hull-White and CIR++ models. Therefore, we also include another two-factor model in the comparison to capture the added flexibility associated with the use of a two-factor model. We opt for the G2++ model which is a two-factor Gaussian model introducing correlation between its two Brownian motions through an additional parameter ρ .¹⁵ The details of this model and additional results are presented in section 1.F. Our main conclusions regarding the superior performance of the VaCIR++ are confirmed and we also document the limited added value of the G2++ model over its one-factor counterpart (Hull-White) for our application.¹⁶

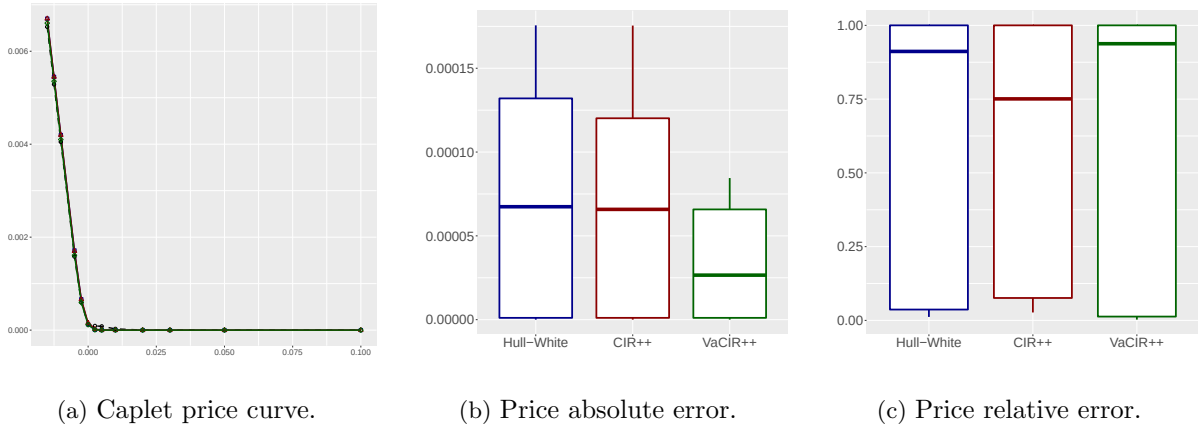


Figure 1.6: Caplet price curves with corresponding absolute and relative pricing errors on 31 March 2016 where models are calibrated on the criteria of JS divergence minimization. Panel (a) displays the market and modelled caplet prices with respect to the strike for the market (dashed), Hull-White ($\square$), CIR++ ($\triangle$) and VaCIR++ ($\diamond$) models. Panel (b) and panel (c) provide respectively the box plots of the corresponding absolute errors and relative errors for the three candidate models.

We further illustrate in Figure 1.6 the pricing accuracy of the three models using the calibrated parameters based on density matching. Figure 1.6a compares the model-implied caplet prices to the market data.¹⁷ It can be concluded that all three models are able to provide a

¹⁵The one-factor Hull-White and CIR++ models have each 4 parameters while the VaCIR++ model has 8 parameters to estimate. The G2++ model is parameterized with 6 parameters.

¹⁶Although the two-factor G2++ model has limited contribution in our application compared to the one-factor Hull-White model, it might lead to improved performance in empirical applications where the correlation between forward rates plays an essential role in the calibration; e.g. for swaption contracts.

¹⁷We note that the market prices are recovered by computing the risk-neutral expectation of the discounted payoffs using the market-implied density as the resulting prices are more comparable with the model-implied ones. This is due to the fact that the prices consistent with the market-implied density, obtained as a by-product of

quite accurate valuation for the caplet products. The absolute and relative pricing errors under the three models are plotted in Figure 1.6b and Figure 1.6c. The bar in the boxplots represents the median of the errors while the lower and upper bounds of each box represent the first and third quartiles respectively. We observe that the VaCIR++ model provides the lowest pricing errors while the Hull-White model and the CIR++ model have comparable pricing accuracy. Additionally, we note that high absolute pricing accuracy does not necessarily translate to better performance in terms of relative pricing errors as close to zero prices might lead to high relative pricing errors. Furthermore, high pricing accuracy indicates that the better density fit obtained when adopting the JS divergence calibration technique instead of the price or implied volatility MAE criterion does not come at the cost of larger pricing errors.

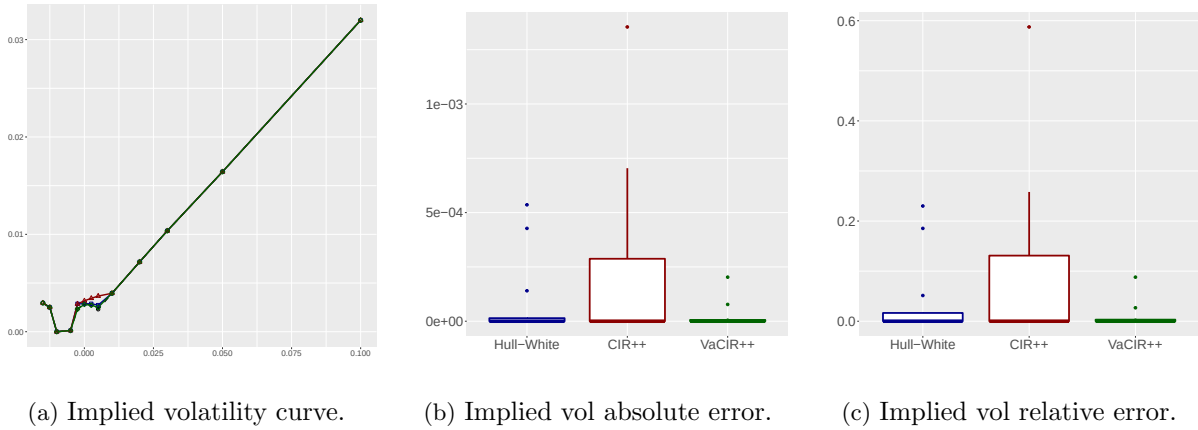


Figure 1.7: Implied volatility curves with associated absolute and relative volatility errors on 31 March 2016 where the models are calibrated on the criteria of JS divergence minimization. Panel (a) contains the implied volatility curves with respect to the strike for the market (dashed), Hull-White ($\square$), CIR++ ($\triangle$) and VaCIR++ ($\diamond$) models. Panel (b) and Panel (c) provide respectively the box plots of the associated absolute errors and relative errors for the three candidate models.

Additionally, we assess in Figure 1.7 the performance of the three models in terms of how close their implied volatility curves, arising from minimizing the JS divergence, are to the market-based implied volatility curve. Figure 1.7a shows that there are visible discrepancies between the implied volatilities under the CIR++ model and the market ones for caplet products with strikes between 0% and 1%. The matching performance of the implied volatility curve outside this range under the three frameworks are indistinguishable. Furthermore, we note that calibration focused

the no-arbitrage volatility curve smoothing, might slightly differ from the actual market prices, which are not necessarily free from arbitrage opportunities.

on density matching also provides convincing fitting performance in terms of implied volatility errors. Therefore, we conclude that the density matching approach provides a more robust way of calibrating interest rate models to fixed income derivative data since it has superior performance in capturing the entire forward distribution, high pricing accuracy and comparable fitting quality in terms of matching implied volatility curves.

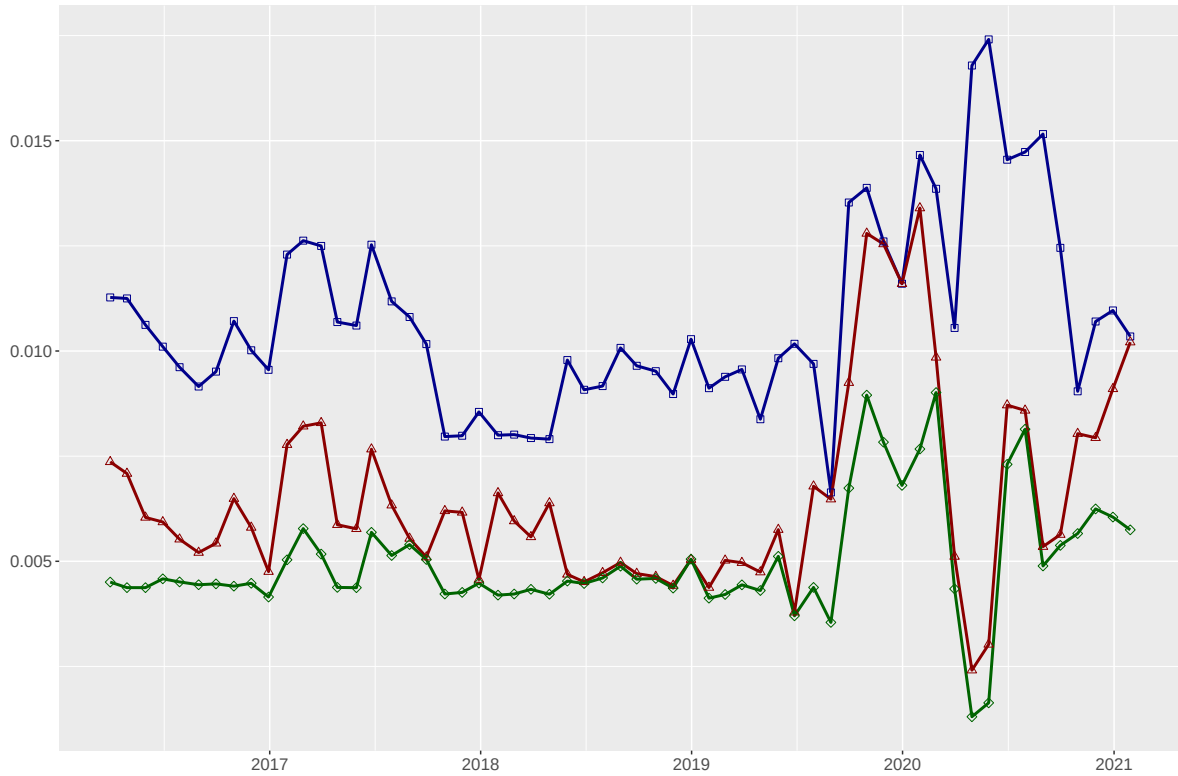


Figure 1.8: Evolution over time of the JS divergence resulting from the density matching calibration under the Hull-White ($\square$), CIR++ ($\triangle$) and VaCIR++ ($\diamond$) models.

We now move to the evaluation of model performance in terms of matching the market-implied density curve over time. To this end, the three models considered are calibrated on a monthly basis and we report in Figure 1.8 the JS divergence for each of them. The best-in-class property of VaCIR++ is preserved through time. In particular, it always has the lowest JS divergence, indicating closest matching to the market-implied density curve. Also, we can observe that the CIR++ model exhibits better performance than the Hull-White model which makes the CIR++ model an attractive candidate as a single-factor model – with the shift providing some flexibility to deal with negative rates. To summarize, the VaCIR++ model is more advantageous both theoretically – allowing for negative values, skewness, and no strict

lower bound – and empirically as it enhances the fitting performance and provides more stability in the calibration over time compared to each component.

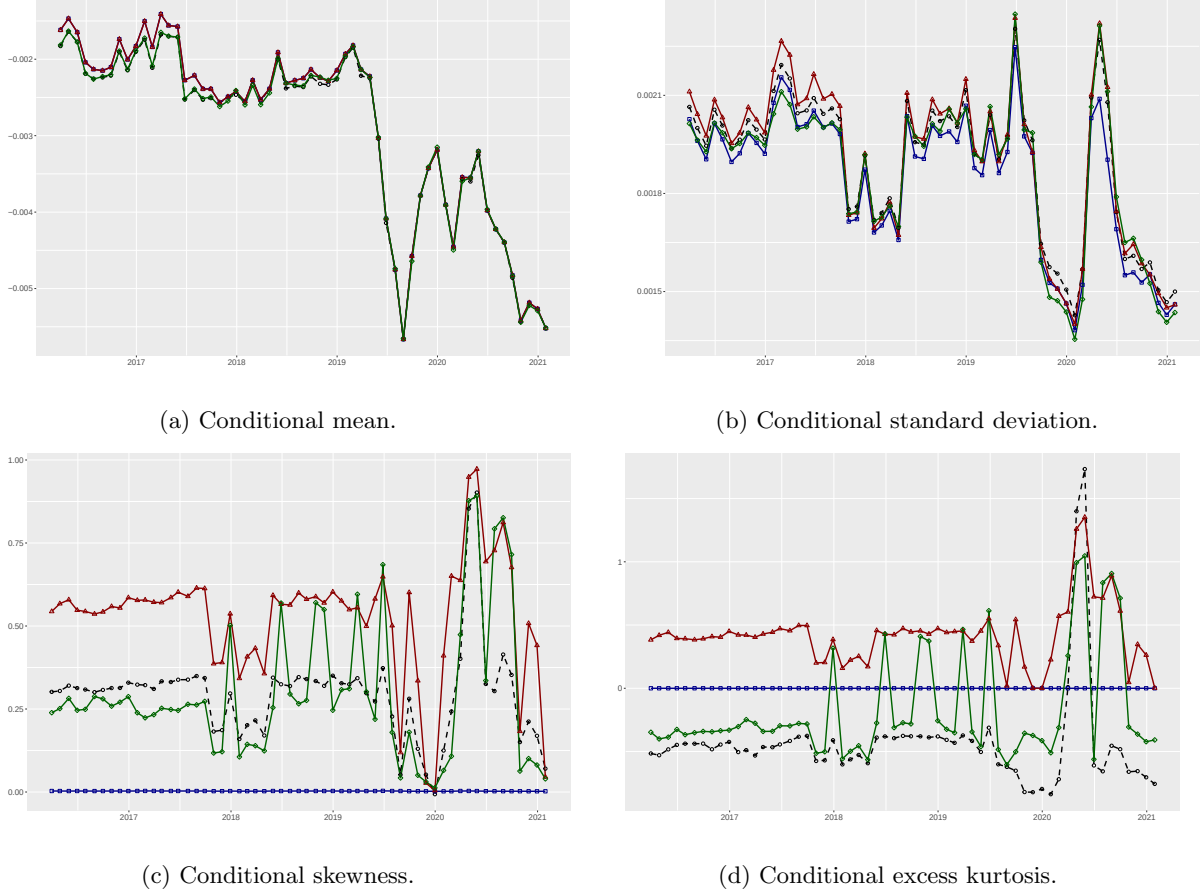


Figure 1.9: Time series of the $L(6M, 1Y)$ conditional moments for the market ($\circ$) and for the Hull-White ($\square$), CIR++ ($\triangle$) and VaCIR++ ($\diamond$) models.

Based on the density obtained, we characterize the variation of the implied conditional moments of the forward rate $L(T, S)$ (under the S -forward measure) over time.¹⁸ In this manner, we compute the first four conditional moments for the 6-month forward rate in 6 months over the whole sample period and investigate how well each of the three competing models can match the market-implied moments (Figure 1.9). It can be observed from Figure 1.9a that the three models accurately track the market-implied mean overall – with the Hull-White and CIR++ models exhibiting slight discrepancies in the first two years of the sample period. Figure 1.9b shows that the Hull-White has the worst performance in terms of tracking the market-implied

¹⁸The market- and model-implied conditional moments are computed numerically using the corresponding forward rate density.

standard deviation while the VaCIR++ model provides the best fitting in general. For the skewness, illustrated in [Figure 1.9c](#), we observe that the conditional skewness of the forward rate distribution under the Hull-White model is almost zero indicating that this model is unable to capture the skewness featured in the data. The conditional skewness under the CIR++ is positive, though higher than the market-implied one in general. The conditional skewness under the VaCIR++ model is the closest to the market-implied one, except for a few months in 2018 and 2019 where the model performance is comparable to the CIR++ model. Lastly, [Figure 1.9d](#) illustrates that the excess kurtosis under the Hull-White model is close to zero while the density implied by the CIR++ model always exhibits positive excess kurtosis. In both cases, the model-implied excess kurtosis is overall higher than the market implied one. The excess kurtosis implied by the VaCIR++ model provides however a better tracking of the market-implied one, especially in the first part of the sample. The jumps observed in the conditional skewness and excess kurtosis for the VaCIR++ model might be caused by local minima in the optimization and warrant further investigation. We also note that the CIR++ model and the VaCIR++ model have similar levels of JS divergence during those months but this does not necessarily imply that the two models have similar densities or conditional moments. Indeed, the two model-implied distributions could be equally similar to the market one while being noticeably different from each other.

Finally, focusing on the forward rate density and its conditional moments provides us with relevant market insights for risk management and hedging applications. As shown in [Figure 1.9a](#) and [Figure 1.9b](#), the mean and standard deviation of the future forward rate are stable at the beginning of the sample period. In particular, for the level of the mean, a significant downturn followed by an increasing trend is observed during 2019 while the period from 2020 to 2021 is dominated by a downward shift. The standard deviation, on the other hand, exhibits a significant upward movement followed by a short decline during 2019 and is then strongly trending upward during the period from 2020 to 2021. This more volatile evolution of interest rates during the COVID-19 period is likely due to the increased uncertainty in financial markets and concerns for funding conditions resulting from the crisis. Furthermore, [Figure 1.9c](#) shows that the conditional skewness is generally positive with a sharp increase at the onset of the COVID-19 crisis followed by a steady decrease over the rest of the sample period. We observe a similar behavior in [Figure 1.9d](#) for the excess kurtosis during COVID. Overall, excess kurtosis is negative over our

sample, indicating a lower tail-thickness compared to a Gaussian distribution.

1.5 Conclusions

We propose a simple asymmetric short-rate model that does not display a strict lower bound and is well-suited to capture the salient features of the negative interest rate environment. Our two-factor model, called the VaCIR++ model, combines the advantages of the Hull-White model (no strict lower bound) and the CIR++ model (positive skewness) while maintaining analytical tractability. Our framework delivers semi-analytical expressions for the forward rate density and caplet pricing which allows us to tackle the challenges encountered by standard benchmark models for the modelling and pricing of interest rate derivatives in a negative interest rate environment.

In addition, we introduce a new calibration procedure based on density matching which alleviates the drawbacks inherent to standard model calibration procedures based on prices or volatilities mean absolute error. Precisely, the model is calibrated such that the model-implied density is the closest from the market-implied one in terms of Kullback-Leibler divergence. While in this context we match the forward rate densities computed under the forward measure, our approach can be broadly applied to other settings where the interest lies in closely matching a chosen target density.

Finally, we illustrate the benefits of our framework in a financial application featuring a time series of caplets whose prices are retrieved by stripping cap implied volatilities. We provide a comparative study of calibration performance in the period of negative interest rates under two calibration criteria - the volatility mean absolute error minimization and our proposed density matching approach based on the minimization of the Kullback-Leibler divergence. We notice the outperformance of our model relative to the Hull-White and the CIR++ models. Moreover, the calibration procedure based on the Kullback-Leibler divergence significantly enhances the matching of the forward densities while preserving high pricing accuracy and comparable fitting quality for implied volatility curves.

Our work can be extended in several ways. First, our modelling framework can be extended to consider simultaneously multiple forward rate maturities and contract expiry lengths. This would allow us to explore the benefits of extracting information relevant for model calibration

using liquid products and apply this calibration to the pricing of less liquid products within a consistent framework. Second, the strength of the economic recovery following the COVID-19 crisis, combined with the escalation of the conflict between Russia and Ukraine, led to a strong revival of inflationary pressures. As major central banks reacted to these events by raising their policy rates, it would be interesting to evaluate how our proposed model and calibration approach perform in a period where interest rates revert back to higher levels.

1.A Pricing formula for existing framework

We provide below the A^z and B^z functions in Equation (1.3) for $z \in \{x, y, y^\phi\}$. For conciseness, we omit the subscript in the κ, θ, σ parameters.

For the Vasicek model specified in Equation (1.1), we have

$$A^x(\tau) = \exp \left\{ \left(\theta - \frac{\sigma^2}{2k^2} \right) [B(\tau) - \tau] - \frac{\sigma^2}{4k} B^x(\tau)^2 \right\} ,$$

$$B^x(\tau) = \frac{1}{k} \left[1 - e^{-k\tau} \right] .$$

For the CIR model specified in Equation (1.2),

$$A^y(\tau) = \left[\frac{2h \exp\{(k+h)\tau/2\}}{2h + (k+h)(\exp\{h\tau\} - 1)} \right]^{2k\theta/\sigma^2} ,$$

$$B^y(\tau) = \frac{2(\exp\{h\tau\} - 1)}{2h + (k+h)(\exp\{h\tau\} - 1)} .$$

where $h = \sqrt{k^2 + 2\sigma^2}$.

The shifted CIR model is a special case of a general model proposed in [Brigo and Mercurio \(2001\)](#). It can be shown that

$$A^{y^\phi}(\tau) = e^{-\phi\tau} A^y(\tau) , \quad B^{y^\phi}(\tau) = B^y(\tau) .$$

1.B Arbitrage-free smoothing of the volatility curve

Following the notations in [Fengler \(2009\)](#), assume that we observe the caplet prices $\mathbf{y} = (y_1, \dots, y_n)$ at the strikes $a = u_0, \dots, u_{n+1} = b$, and the function g represents the natural cubic spline function. For the value and second derivative representation, we set $g_i = g(u_i)$ and $\gamma_i = g''(u_i)$, for $i = 1, \dots, n$. Furthermore, we define $\mathbf{g} = (g_1, \dots, g_n)^\top$ and $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{n-1})^\top$. By definition, $\gamma_1 = \gamma_n = 0$.

We formulate the sufficient and necessary conditions to ensure a valid cubic spline using the two matrices $\mathbf{Q}$ and $\mathbf{R}$ defined below. Let $h_i = u_{i+1} - u_i$ for $i = 1, \dots, n-1$, and the elements $q_{i,j}$, of matrix $\mathbf{Q}$, for $i = 1, \dots, n$ and $j = 2, \dots, n-1$, are given by

$$q_{j-1,j} = h_{j-1}^{-1}, q_{j,j} = -h_{j-1}^{-1} - h_j^{-1} \text{ and } q_{j+1,j} = h_j^{-1},$$

for $j = 2, \dots, n-1$ and $q_{i,j} = 0$ for $|i - j| \geq 2$.

The matrix $\mathbf{R}$ is symmetric and its elements $r_{i,j}$, for $i, j = 2, \dots, n-1$, are defined by

$$\begin{aligned} r_{i,i} &= \frac{1}{3}(h_{i-1} + h_i), & \text{for } i = 2, \dots, n-1 \\ r_{i,i+1} = r_{i+1,i} &= \frac{1}{6}h_i, & \text{for } i = 2, \dots, n-2 \\ r_{i,j} &= 0, & \text{for } |i - j| \geq 2 \end{aligned}$$

Furthermore, we formulate the spline smoothing problem as a quadratic minimization program. Define vector $\mathbf{y} = (w_1 y_1, \dots, w_n y_n, 0, \dots, 0)^\top$, where the w_i are strictly positive weights and vector $\mathbf{x} = (\mathbf{g}^\top, \boldsymbol{\gamma}^\top)^\top$. Furthermore, define the matrices, $\mathbf{A} = (\mathbf{Q}, -\mathbf{R}^\top)$ and

$$\mathbf{B} = \begin{pmatrix} \mathbf{W}_n & 0 \\ 0 & \lambda \mathbf{R} \end{pmatrix}$$

where $\mathbf{W}_n = \text{diag}(w_1, \dots, w_n)$. The smoothing algorithm of the volatility curve is summarized below.

1. Estimate the volatility curve via an initial interpolation of caplet prices with respect to the moneyness, which is defined as the strikes in excess to the forward rate.
2. Obtain the implied volatility curve using the spline smoothing technique under no-arbitrage

constraints by solving the quadratic program formulated as follows

$$\begin{aligned}
& \min_{\mathbf{x}} \quad -\mathbf{y}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \mathbf{B} \mathbf{x} , \\
& \text{subject to} \quad \mathbf{A}^\top \mathbf{x} = 0 , \\
& \qquad \qquad \qquad \gamma_i \geq 0 , \\
& \qquad \qquad \qquad \frac{g_2 - g_1}{h_1} - \frac{h_1}{6} \gamma_2 \geq P_t(S)(S - T) , \\
& \qquad \qquad \qquad -\frac{g_n - g_{n-1}}{h_{n-1}} - \frac{h_{n-1}}{6} \gamma_{n-1} \geq 0 .
\end{aligned}$$

1.C Correlation between two factors

1.C.1 Introduce correlation within the affine framework

Assume short rate is defined as $r_t = X_t^1 + X_t^2$ and the two factors in the model $X_t = (X_t^1, X_t^2)$ are described as following system of SDEs under the risk-neutral Q -measure:

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \end{pmatrix} = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} \left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \sqrt{X_t^1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix},$$

which is equivalent to

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \end{pmatrix} = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} \left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11}\sqrt{X_t^1} & 0 \\ \sigma_{21}\sqrt{X_t^1} & \sigma_{22} \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix}.$$

We can re-write it as

$$\begin{aligned} dX_t^1 &= \kappa_1(\theta_1 - X_t^1)dt + \sigma_{11}\sqrt{X_t^1}dW_t^1, \\ dX_t^2 &= \kappa_2(\theta_2 - X_t^2)dt + \sigma_{21}\sqrt{X_t^1}dW_t^1 + \sigma_{22}dW_t^2. \end{aligned}$$

Here, correlation between the two factors X_t^1, X_t^2 is introduced by σ_{21} . If we set $\sigma_{21} = 0$, this framework reduces to the case where we have two independent processes with X_t^1 being a CIR process and X_t^2 being a Vasicek process. In addition, this belongs to the Affine framework given that

$$\Sigma\Sigma' = \begin{pmatrix} \sigma_{11}\sqrt{X_t^1} & 0 \\ \sigma_{21}\sqrt{X_t^1} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \sigma_{11}\sqrt{X_t^1} & \sigma_{21}\sqrt{X_t^1} \\ 0 & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_{11}^2 X_t^1 & \sigma_{11}\sigma_{21}X_t^1 \\ \sigma_{11}\sigma_{21}X_t^1 & \sigma_{21}^2 X_t^1 + \sigma_{22}^2 \end{pmatrix}.$$

The zero coupon bond price is defined as

$$\begin{aligned} P_t^r(T) &= \exp(B(T-t)'X_t + A(T-t)) \\ &= \exp\left(B^1(T-t)X_t^1 + B^2(T-t)X_t^2 + A(T-t)\right). \end{aligned}$$

For the factor loadings in the zero-coupon bond prices, A and B are the solutions to

$$\begin{aligned}
\frac{dB^1(t, T)}{dt} &= 1 + \kappa_2 B^1(t, T) - \frac{1}{2} \left(\sigma_{21} B^2(t, T) + \sigma_{11} B^1(t, T) \right)^2, \\
\frac{dB^2(t, T)}{dt} &= 1 + \kappa_1 B^2(t, T), \\
\frac{dA(t, T)}{dt} &= -\kappa_1 \theta_1 B^1(t, T) - \kappa_2 \theta_2 B^2(t, T) - \frac{1}{2} \sigma_{22}^2 B^2(t, T)^2.
\end{aligned}$$

The solutions to $B(t, T)$ are formulated as

$$\begin{aligned}
B^1(t, T) &= B^1(T - t) = -\frac{1}{\kappa_1} \left[1 - e^{-\kappa_1(T-t)} \right], \\
B^2(t, T) &= B^2(T - t) = -\frac{(2 - \sigma_{21}^2 B^1(t, T)^2)(\exp\{(T - t)h\} - 1)}{2h + (h + \kappa_2 - \sigma_{21}\sigma_2 B^1(t, T))(\exp\{(T - t)h\} - 1)}, \\
h &= \sqrt{2\sigma_{21}\kappa_2\sigma_2 B^1(T - t) - \kappa_2^2 - 2\sigma_2^2}.
\end{aligned}$$

and the solution to $A(t, T)$ can be obtained by integration.

1.C.2 Introduce correlation by correlated Brownian Motions

Assume W_t^1, W_t^2 are independent Brownian motions, and the short rate $r_t = X_t^1 + X_t^2$.

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \end{pmatrix} = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} \left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_{11}\sqrt{X_t^1} & 0 \\ \rho\sigma_{22} & \sqrt{1 - \rho^2}\sigma_{22} \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix},$$

We can re-write it as

$$\begin{aligned}
dX_t^1 &= \kappa_1(\theta_1 - X_t^1)dt + \sigma_{11}\sqrt{X_t^1}dW_t^1, \\
dX_t^2 &= \kappa_2(\theta_2 - X_t^2)dt + \rho\sigma_{22}dW_t^1 + \sqrt{1 - \rho^2}\sigma_{22}dW_t^2 \\
&= \kappa_2(\theta_2 - X_t^2)dt + \sigma_{22}dW_t^3,
\end{aligned}$$

where W_3 is a Brownian motion satisfying $\text{corr}(dW_t^1, dW_t^3) = \rho dt$.

Here, ρ introduce a correlation between the CIR process and the Vasicek process by correlated Brownian Motions. If we set $\rho = 0$, this framework reduces to the case where we have two independent processes with X_t^1 being a CIR process and X_t^2 being a Vasicek process. When $\rho \neq 0$, this model is no longer affine since

$$\Sigma\Sigma' = \begin{pmatrix} \sigma_{11}\sqrt{X_t^1} & 0 \\ \rho\sigma_{22} & \sqrt{1 - \rho^2}\sigma_{22} \end{pmatrix} \begin{pmatrix} \sigma_{11}\sqrt{X_t^1} & \rho\sigma_{22} \\ 0 & \sqrt{1 - \rho^2}\sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_{11}^2 X_t^1 & \rho\sigma_{11}\sigma_{22}\sqrt{X_t^1} \\ \rho\sigma_{11}\sigma_{22}\sqrt{X_t^1} & \sigma_{22}^2 \end{pmatrix},$$

1.D Proof of Proposition 1

Point (a) is obtained by combining Equation (1.11) with Equation (1.12). For point (b), we have the dynamics of the zero-coupon bond price for $z \in \{x, y\}$

$$\frac{dP_t^z(T)}{P_t^z(T)} = z_t dt + B^z(T-t)\sigma_t^z dW_t^z, \quad 0 \leq t \leq T.$$

Applying Ito's quotient rule, we have for $0 \leq t \leq T < S$,

$$\begin{aligned} \frac{d(P_t^z(T)/P_t^z(S))}{P_t^z(T)/P_t^z(S)} &= \frac{dP_t^z(T)}{P_t^z(T)} - \frac{dP_t^z(S)}{P_t^z(S)} - \frac{dP_t^z(T)}{P_t^z(T)} \frac{dP_t^z(S)}{P_t^z(S)} + \left(\frac{dP_t^z(S)}{P_t^z(S)} \right)^2 \\ &= \sigma_t^z (B^z(T-t) - B^z(S-t)) (dW_t^z - \sigma_t^z B^z(S-t)dt). \end{aligned}$$

Combining this expression with (1.12), the dynamics of the forward rate become

$$\begin{aligned} dF_t^z(T, S) &= \frac{1}{\Delta} d(P_t^z(T)/P_t^z(S)) \\ &= \left(\frac{1}{\Delta} + F_t^z(T, S) \right) \sigma_t^z ((B^z(S-t) - B^z(T-t)) (-\sigma_t^z B^z(S-t)dt + dW_t^z)) \\ &= \sigma^z(t, F_t^z(T, S)) (-\zeta_t^{z,S} dt + dW_t^z) \\ &= \sigma^z(t, F_t^z(T, S)) dW_t^{z,S}, \end{aligned}$$

where

$$\sigma^z(t, v) := \sigma_t^z \left[\frac{1}{\Delta} + v \right] ((B^z(S-t) - B^z(T-t))).$$

Using Equation (1.11) and the independence between P^x, P^y ,

$$\frac{P_t^r(S)}{M_t^r} = \frac{P_t^x(S)}{M_t^x} \frac{P_t^y(S)}{M_t^y} = P_0^r(S) \mathcal{E} \left(\int_0^t \sigma_s^x dW_s^x + \int_0^t \sigma_s^y dW_s^y \right),$$

where $\mathcal{E}$ the Doléans-Dade exponential. Hence,

$$\mathbb{E} \left[\frac{dQ^S}{dQ} \middle| \mathcal{F}_t \right] = \frac{P_t^r(S)}{P_0^r(S)M_t^r} = \mathcal{E} \left(\int_0^t \zeta_s^{x,S} dW_s^x + \int_0^t \zeta_s^{y,S} dW_s^y \right).$$

Applying the Girsanov theorem, the processes $W^{x,S}$ and $W^{y,S}$, defined as $W_t^{z,S} := W_t^z - \int_0^t \zeta_s^{z,S} ds$, $z \in \{x, y\}$, are independent $\mathbb{Q}^S$ -Brownian motions. Hence, $F^z(T, S)$ is a $\mathbb{Q}^S$ -martingale for $z \in \{x, y\}$. Since $\sigma_t^y = \sigma_t^{y^\phi}$ and $B^y = B^{y^\phi}$, $F^{y^\phi}(T, S)$ is also driftless under $\mathbb{Q}^S$. Finally, the martingale property of $F^r(T, S)$ is obtained by applying Ito's product rule to Equation (1.13) together with the independence between $W^{x,S}$ and $W^{y,S}$:

$$dF_t^r(T, S) = (1 + \Delta F_t^y(T, S))dF_t^x(T, S) + (1 + \Delta F_t^x(T, S))dF_t^y(T, S) + \underbrace{\Delta d\langle F^x(T, S), F^y(T, S) \rangle_t}_{=0}.$$

This shows that $F^r(T, S)$ is a sum of two $\mathbb{Q}^S$ -martingales and thus itself is a martingale under

$\mathbb{Q}^S$, concluding the proof of Point (b).

The expression of the diffusion coefficients for x, y and y^ϕ are provided in Equation (1.29). For the Vasicek model, substituting $\sigma_t^x = \sigma_1$ leads to the expression for $\sigma^x(t, F_t^y(T, S))$. For the CIR model, $\sigma_t^y = \sigma_2 \sqrt{y_t}$, which can be written as a function of $F_t^y(T, S)$ using Equation (1.4) and Equation (1.12). Combining these elements leads to the expression for $\sigma^y(t, F_t^y(T, S))$. Finally, $\sigma^{y^\phi}(t, F_t^{y^\phi}(T, S))$ can be derived using $\sigma_t^{y^\phi} = \sigma_t^y$, $B^{y^\phi} = B^y$ and $A^{y^\phi}(\tau) = e^{-\phi\tau} A^y(\tau)$, concluding the proof of point (c).

1.E Calibration results of implied volatility curve: 29 January 2021

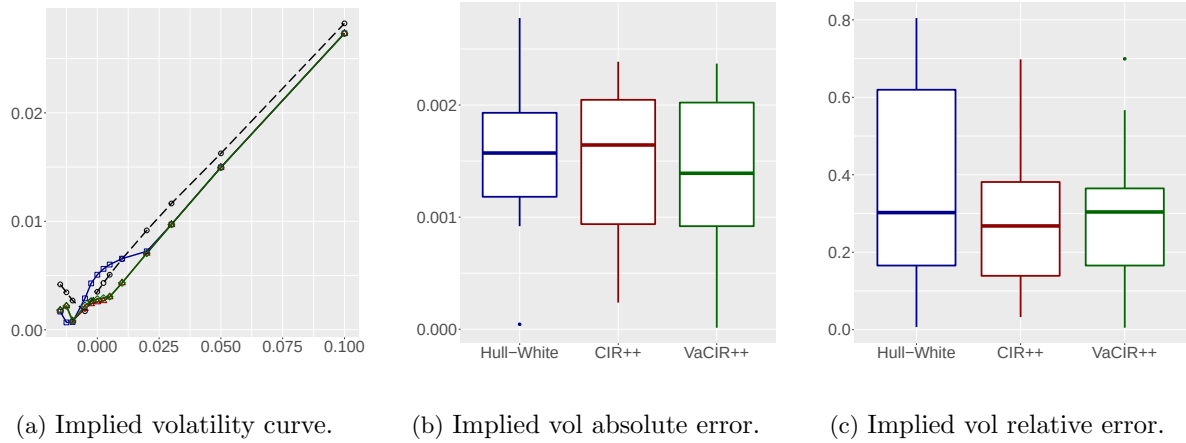


Figure 1.10: Implied volatility curves with associated absolute and relative volatility errors on 29 January 2021 where the models are calibrated on the criteria of implied volatility error minimization. Panel (a) contains the implied volatility curves with respect to different strike rates. Panel (b) and Panel (c) provide respectively the box plots of the associated absolute errors and relative errors for the three candidate models.

We take the calibration results on 29 January 2021 to illustrate the performance of the three frameworks where the CIR++ model has superior performance compared to the Hull-White model. The implied volatility curves with the corresponding errors are exhibited in [Figure 1.10](#). We note that although the CIR++ model has a slightly higher median absolute error compared to the Hull-White model, it has an overall smaller discrepancy from the market data, especially evaluated in terms of relative volatility errors. Moreover, the VaCIR++ model further reduces the absolute pricing error and its performance is closer to the CIR++ framework.

1.F The G2++ model

We also consider the G2++ model as a benchmark. The model is specified as follows

$$r_t = x_t + y_t + \phi^r(t), \quad r_0 \in \mathbb{R}, \quad (1.29)$$

$$dx_t = -ax_t dt + \sigma dW_t^x, \quad x_0 = 0, \quad (1.30)$$

$$dy_t = -by_t dt + \eta dW_t^y, \quad y_0 = 0. \quad (1.31)$$

where a, b, σ, η are positive constants, and $dW_t^x dW_t^y = \rho dt$ with $-1 \leq \rho \leq 1$. $\phi^r(t)$ is the deterministic shift function under the G2++ model with $\phi^r(0) = r_0$ which allows the model to perfectly fit the term structure of interest rates.¹⁹ The expression for the shift function is given by

$$\phi^r(t) = f^M(0, t) + \frac{\sigma^2}{2a^2} (1 - e^{-at})^2 + \frac{\eta^2}{2b^2} (1 - e^{-bt})^2 + \rho \frac{\sigma\eta}{ab} (1 - e^{-at}) (1 - e^{-bt}). \quad (1.32)$$

We plot in [Figure 1.11](#) the density curves obtained by calibrating the G2++ and the VaCIR++ models (solid lines) to the data on the first available date in our sample. [Figure 1.11a](#) shows that the calibrated density under the G2++ model has a right-shifted distribution and is more symmetric, compared to the market-implied one. Furthermore, it provides a closer fit to the market-implied density on the right-hand side but deviates significantly on the left-hand side. In addition, the calibrated density obtained from the VaCIR++ model is displayed in [Figure 1.11b](#), and we conclude that the combined model matches well both tails and captures the overall market-implied density curve better than the G2++ model. The superior performance of the VaCIR++ is further confirmed by the significantly lower JS divergence value compared to the G2++ model, see the evolution of the JS divergence over time in [Figure 1.12](#).

In addition, we observe that the performance of the G2++ and Hull-White models in terms of matching the market-implied forward rate density are essentially indistinguishable. See the blue marks (Hull-White model) in [Figure 1.12](#) which are almost perfectly aligned with the yellow curve (G2++ model). In fact, the estimated parameter ρ under the G2++ model is extremely close to minus one over the whole sample period. In this case, the G2++ model degenerates into a one factor short-rate process. This observation is consistent with the results of the empirical application presented in Chapter 4 of [Brigo and Mercurio \(2007\)](#). Furthermore, we only consider

¹⁹The interested reader is referred to Chapter 4 of [Brigo and Mercurio \(2007\)](#) for more details on the G2++ model.

cap products in our application whose prices do not depend on the correlation between forward rates. We thus conclude that the two-factor Gaussian model does not improve over the one-factor Gaussian model in terms of matching performance for the market forward distribution implied by caplet products.²⁰

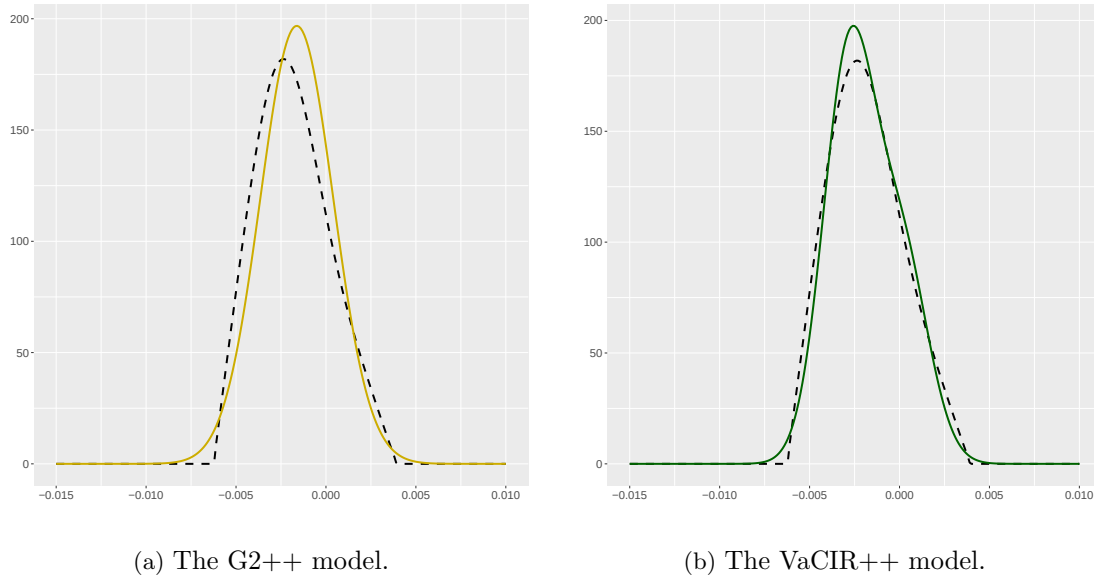


Figure 1.11: Density curves of the forward rate $L(6M, 1Y)$ on 31 March 2016 where models are calibrated on the criteria of JS divergence minimization. Panel (a) and panel (b) exhibit the market-implied forward density curve and the modelled forward density curves (dashed lines) and under the G2++ and the VaCIR++ models, respectively (solid lines).

²⁰Although the two-factor G2++ model has limited contribution in our application compared to the one-factor Hull-White model, it might lead to improved performance in empirical applications where the correlation between forward rates plays an essential role in the calibration; e.g. for swaption contracts.

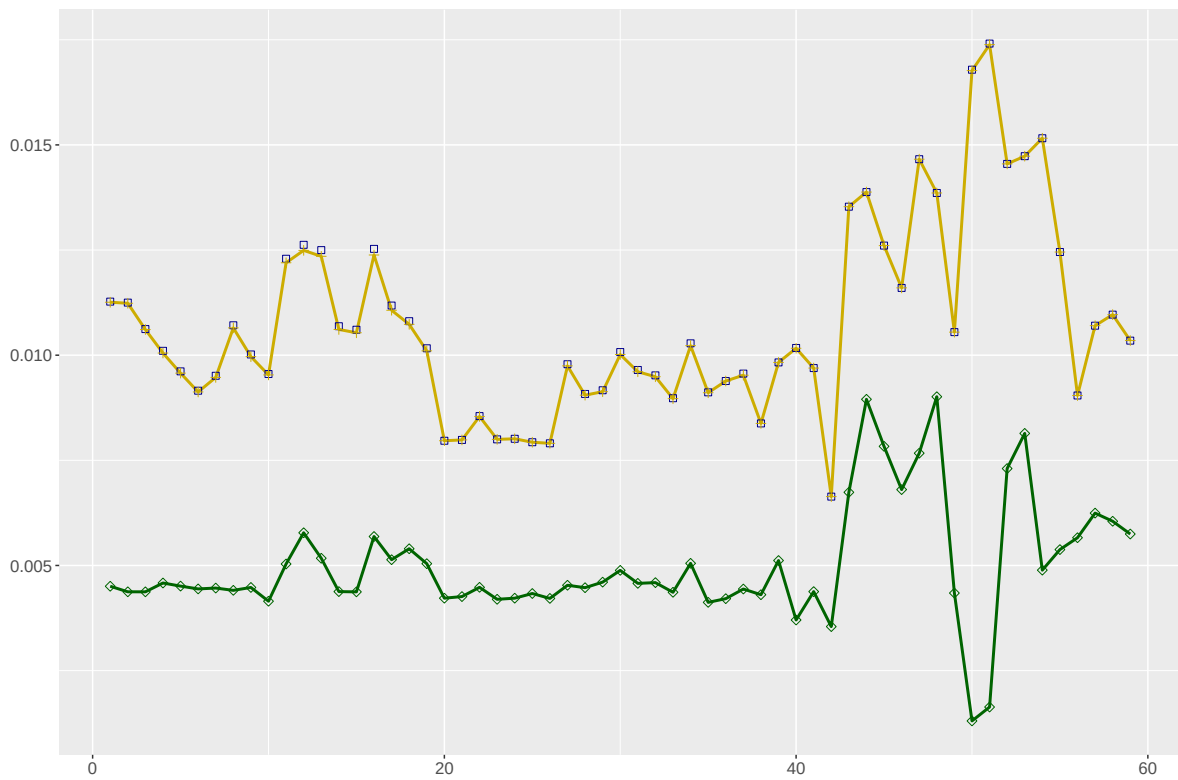


Figure 1.12: Evolution over time of the JS divergence resulting from the density matching calibration under the Hull-White ($\square$, blue), the G2++ (+, yellow) and VaCIR++ ($\diamond$, green) models.

Chapter 2

A dynamic conditional score model for the log correlation matrix¹

2.1 Introduction

Correlation matrices are pervasive in many applied sciences such as finance and econometrics. Reliable models for the dynamics of correlations between different assets or underlying risk factors are essential for asset pricing, portfolio selection and risk management. Nonetheless, correlation matrices are notoriously difficult to model in high dimensions, for at least two reasons: the positivity constraint and the balance between flexibility and computational feasibility. These challenges are well documented and addressed in the framework of multivariate GARCH models, see e.g. [Bauwens et al. \(2006\)](#), [Bauwens et al. \(2012\)](#) and [Francq and Zakoian \(2019\)](#).

This paper proposes a new model for the joint modelling of volatilities and correlations, which both are driven by the respective conditional scores of the likelihood, and hence fits into the general modelling framework of *dynamic conditional score* or, in an alternative terminology, *generalized autoregressive score* models, as introduced by the seminal works of [Creal et al. \(2011\)](#), [Creal et al. \(2013\)](#) and [Harvey \(2013\)](#). Score driven correlation models have been proposed recently in the context of modelling realized variances and covariances. For example, [Gorgi et al. \(2018\)](#) use the Wishart distribution, and [Opschoor et al. \(2018\)](#) and [Vassallo et al. \(2021\)](#)

¹This chapter corresponds to Hafner, C.M. and L. Wang (2021): “A dynamic conditional score model for the log correlation matrix”, forthcoming at *Journal of Econometrics*, doi: <https://doi.org/10.1016/j.jeconom.2021.09.004>.

the more flexible matrix-F distribution.

Our aim is to allow for fat tails in the conditional distributions, and in particular, to generalize the popular Beta-t-EGARCH model proposed in [Harvey \(2013\)](#) (Section 4.1), which uses the student-t distribution, to the multivariate case. The conditional covariance matrix can be decomposed into volatility and correlation components, which facilitates estimation since it can be split into two parts as in the DCC model of [Engle \(2002\)](#) and in [Vassallo et al. \(2021\)](#).

For the correlation part, we specify the dynamics for the matrix logarithm of the correlation matrix to guarantee that positive definiteness is preserved in arbitrary dimensions. This approach is analogous to the Beta-t-EGARCH model in which the log volatility process is modelled to ensure positivity. In the bivariate case, the matrix log reduces to the Fisher transformation of correlations which has been adopted e.g. by [Harvey \(2013\)](#) (Section 7.3.1) and [Hansen et al. \(2014\)](#). The idea of modelling the matrix log of a correlation or covariance matrix is not new. For example, [Kawakatsu \(2006\)](#) proposed a multivariate version of the EGARCH model of [Nelson \(1991\)](#), but with *ad hoc* rather than score-driven dynamics. In a static framework, [Hafner et al. \(2020\)](#) impose a multiplicative structure for the correlation matrix, which becomes additive under the matrix log transformation, and hence easier to estimate using minimum distance type estimators.

Modelling the log correlation matrix circumvents the need to impose parameter restrictions to ensure positive definiteness of the resulting correlation matrix and thus allows the implementation of an unconstrained model for the vectorized log correlation parameter. Moreover, block structures are preserved under the matrix log transformation, as shown by [Archakov and Hansen \(2020\)](#), so that factor structures may be imposed for the log correlation parameter to allow for estimation in high dimensions. This idea is used by [Archakov et al. \(2020\)](#) to estimate a multivariate realized GARCH model.

The parameter characterizing the dynamics of the log correlation matrix will be driven by the score of the likelihood with respect to this parameter. The distribution characterizing the dependence part of the model is assumed to be a student-t copula. If all conditional marginal distributions happen to be univariate student-t with the same degrees-of-freedom, equal to the parameter of the t-copula, then we have a multivariate student-t distribution. However, we do not impose this restriction, so we can have what is commonly called a Meta-t distribution, see [Demarta and McNeil \(2005\)](#), i.e. a student-t copula combined with arbitrary student-t

marginals.

There is of course an identification issue because for an $(N \times N)$ correlation matrix, only $N(N - 1)/2$ elements are freely determinable. Consequently, for the matrix log of a correlation matrix, which is symmetric by definition, the diagonal elements are implied by the off-diagonal elements, although no analytical form is available in higher dimensions. [Archakov and Hansen \(2021\)](#) propose an algorithm that reconstructs the original correlation matrix based on the lower-triangular part, excluding the diagonal, of the log correlation matrix. We follow the same approach here by endowing this $N(N - 1)/2$ -dimensional parameter with score-driven dynamics.

We propose two approaches for estimation, the first is a one-step estimator where all parts of the model are estimated simultaneously, which ensures statistical efficiency but may not be computationally feasible in high dimensions. In this case, we assume equality of the conditional marginal distributions, so that the conditional distribution is indeed a multivariate student-t with only one degree-of-freedom parameter. The second is a two-step estimator where the volatilities are estimated in a first step, and the correlations in a second step. The marginal student-t distributions are allowed to have different degrees-of-freedom parameters. This estimator, although statistically inefficient, is computationally simple and feasible even in higher dimensions. We derive the asymptotic estimation theory for both estimators.

In an empirical application to a set of six financial, commodity and cryptocurrency indices, we show that the proposed model fits well the data using two-step estimation and a variety of diagnostics and specification tests. We conclude that parameter restrictions across the components of the log correlation parameter may be employed to reduce the dimension of the parameter space and are supported by classical information criteria. Furthermore, we conduct a performance comparison of the proposed log correlation model with alternative parameterizations for both in-sample estimation and out-of-sample forecasting. We show that our model provides the best in-sample model fit and has similar out-of-sample performance compared with other score-driven models.

The remainder of the paper is organized as follows. The following section recalls the definition and the structure of the matrix log of a correlation matrix. Section [2.3](#) introduces the model and proposes a one-step and two-step maximum likelihood estimator, including their corresponding asymptotic estimation theories. Section [2.4](#) presents a detailed empirical application of the model, and Section [2.5](#) concludes. Theoretical material including proofs of the theorems is

collected in Appendix 2.A, and some tables and figures of the empirical example are shown in Appendix 2.B. Appendix 2.C shows that our model reduces to the one of Harvey (2013), Section 7.3.1, in the bivariate Gaussian case. Appendix 2.D includes additional derivations on the Lyapounov exponent of the stochastic recurrence equation, and Appendix 2.E gives analytical expressions for the information matrix. We provide a Monte Carlo study in Appendix 2.F to analyze the finite sample performance of the one-step and two-step estimation procedures, and additional tables and figures are collected in Appendix 2.G.

2.2 Matrix logarithm of a correlation matrix

The matrix logarithm of a correlation matrix Θ is defined as $\log \Theta = \Gamma(\log \Lambda)\Gamma^\top$, where Γ is the matrix containing the eigenvectors of Θ in its columns, and Λ is a diagonal matrix containing the corresponding eigenvalues on the diagonal, which are all strictly positive. The diagonal matrix $\log \Lambda$ replaces the eigenvalues on the diagonal by their logarithm. The matrix $\log \Theta$ has a complicated structure, its diagonal elements can take any non-positive values (see Lemma 2 of Archakov and Hansen (2021)) and its off-diagonals can take any values. As an illustration, suppose that

$$\Theta = \begin{pmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0.2 \\ 0.5 & 0.2 & 1 \end{pmatrix}, \quad \text{then} \quad \log \Theta = \begin{pmatrix} -0.75 & 1.18 & 0.64 \\ 1.18 & -0.55 & -0.07 \\ 0.64 & -0.07 & -0.17 \end{pmatrix}.$$

There are $N(N+1)/2$ parameters in $\log \Theta$; we call these *log parameters*. On the other hand, Θ has only $N(N-1)/2$ original parameters. For each Θ , its $N(N-1)/2$ original parameters completely identify its $N(N+1)/2$ log parameters. In other words, there exists a function $f : \mathbb{R}^{N(N-1)/2} \rightarrow \mathbb{R}^{N(N+1)/2}$ which maps the original parameters to the log parameters. However, when $N > 4$, f does *not* have a closed form because when $N > 4$ the continuous functions which map elements of a matrix to its eigenvalues have no closed form. When $N = 2$, we can solve f by hand (see Example 1).

Example 1. *Suppose*

$$\Theta = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The eigenvalues of Θ are $1 + \rho$ and $1 - \rho$, respectively. The corresponding eigenvectors are $(1, 1)^\top/\sqrt{2}$ and $(1, -1)^\top/\sqrt{2}$, respectively. Therefore

$$\begin{aligned} \log \Theta &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \log(1 + \rho) & 0 \\ 0 & \log(1 - \rho) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{2} \\ &= \begin{pmatrix} \frac{1}{2} \log(1 - \rho^2) & \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right) \\ \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right) & \frac{1}{2} \log(1 - \rho^2) \end{pmatrix}. \end{aligned}$$

Thus

$$f(\rho) = \left(\frac{1}{2} \log(1 - \rho^2), \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{2} \log(1 - \rho^2) \right)^\top,$$

and we see that ρ generates two distinct entries for $\log \Theta$. The off-diagonal element $\frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right)$ is the well-known Fisher's z -transformation of ρ . We also see that $\log \Theta$ is not only symmetric about the diagonal, but also symmetric about the cross-diagonal (from the upper right to the lower left).

The idea of [Archakov and Hansen \(2021\)](#) is to identify a subset of unrestricted log parameters and then fill in the remainder afterwards. They propose the following procedure. Let θ be the vector containing only the lower triangular portion of $\log \Theta$, excluding the diagonal. Then, the diagonal elements of $\log \Theta$ are uniquely determined by some function $\phi : \mathbb{R}^{N(N-1)/2} \rightarrow \mathbb{R}^N$, which can be obtained numerically (in the case $N = 2$ there is a closed form, see [Example 1](#)). [Archakov and Hansen \(2021\)](#) gave a concrete algorithm to do this and established its validity. We will use this idea in our model, starting from a dynamic parameter θ_t , to first reconstruct the full log correlation matrix $\log \Theta_t$, and then by matrix exponentiation the original correlation matrix Θ_t .

2.3 The model and its estimation

We introduce a model that combines conditional heteroskedasticity, dynamic conditional correlations, and conditional leptokurtosis in a dynamic score-driven framework. Consider the return vector for N assets, $y_t = (y_{1t}, \dots, y_{Nt})^\top$. To simplify the presentation and discussion of the model, we will first assume that the conditional distribution of y_t given the past is a multivariate student-t distribution, but later allow for marginal distributions with different degrees of freedom parameters.

We further assume that the parameter vector ϑ can be separated into a vector τ that characterizes the marginal distributions and the volatility part, a vector δ which characterizes the correlation part, and the degrees-of-freedom parameter η , and hence, $\vartheta = (\tau^\top, \delta^\top, \eta)^\top$.

The core of the model is the process generating the conditional correlation matrix Θ_t . Before presenting this part of the model, we introduce the model framework for y_t including the marginal dynamics and the distribution conditional on the information set $\mathcal{F}_{t-1}$,

$$y_t | \mathcal{F}_{t-1} \sim F_{\eta, H_t}, \quad \eta > 2 \quad (2.1)$$

$$H_t(\vartheta) = D_t(\tau) \Theta_t(\delta) D_t(\tau) \quad (2.2)$$

$$D_t(\tau) = \text{diag}((\exp(h_{1t}), \dots, \exp(h_{Nt}))^\top) \quad (2.3)$$

$$h_{it}(\tau_i) = \omega_i^v + \phi_i^v h_{i,t-1} + \kappa_i^v u_{i,t-1}^v, \quad i = 1, \dots, N \quad (2.4)$$

$$u_{it}^v = \frac{(\eta + 1)y_{it}^2}{(\eta - 2)\exp(2h_{it}) + y_{it}^2} - 1, \quad i = 1, \dots, N \quad (2.5)$$

where F_{η, H_t} is the standardized multivariate student-t distribution with η d.o.f., $\eta > 2$, and dispersion matrix H_t , i.e. its log density is given by

$$f_{\eta, H_t}(y_t) = k(\eta) - \frac{1}{2} \left\{ \log |H_t| + (\eta + N) \log \left(1 + \frac{1}{\eta - 2} y_t^\top H_t^{-1} y_t \right) \right\}.$$

where $k(\eta) = \log \Gamma\left(\frac{\eta+N}{2}\right) - \log \Gamma\left(\frac{\eta}{2}\right) - \frac{N}{2} \log((\eta - 2)\pi)$. The structure of the model implies that the conditional mean of y_t is zero, the conditional variances are given by $\exp(2h_{it})$, and the conditional correlation matrix is given by Θ_t . Each one of D_t , Θ_t and H_t are assumed to be measurable functions of the information set $\mathcal{F}_{t-1}$.

Each asset's volatility is assumed to follow a univariate Beta-t-EGARCH process, where the score with respect to log volatility is given by Equation (2.5). The only difference with respect to the standard formulation of this model in [Harvey \(2013\)](#) is that we do not consider the classical but the standardized student-t distribution, but the score is the same up to a scaling factor that depends on the degrees-of-freedom. It is straightforward to generalize this model to allow for volatility spillover, perhaps following some test for Granger causality in volatility. To keep the notation simple and estimation feasible in high dimensions, we refrain from doing so here as it is not directly relevant for our main focus, the modelling of the correlation part.

We denote by $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})^\top$ the vector of standardized residuals obtained after estimation of the volatility model, i.e., $\varepsilon_t = D_t(\tau)^{-1} y_t$, so that the components of ε_t have conditional mean zero and conditional variance unity. However, the components of ε_t are conditionally

correlated, and the conditional correlation matrix is given by $\Theta_t = \mathbb{E}[\varepsilon_t \varepsilon_t^\top | \mathcal{F}_{t-1}]$.

Now consider the following score-driven model for the conditional correlation matrix $\Theta_t(\delta)$. As in [Archakov and Hansen \(2021\)](#) we model the lower triangular part of $\log \Theta_t(\delta)$, excluding the diagonal, i.e., $\theta_t(\delta) := \text{vecl}(\log \Theta_t(\delta))$, and then fill in the diagonal elements to ensure that Θ_t is a correlation matrix. The model is a first order dynamic conditional score model for the log correlation parameter, i.e.

$$\theta_t(\delta) = \omega + \Phi \theta_{t-1}(\delta) + K u_{t-1} \quad (2.6)$$

where u_t is the score of the likelihood function with respect to θ_t , and $\Phi = \text{diag}(\phi_1, \dots, \phi_{N^*})$ and $K = \text{diag}(\kappa_1, \dots, \kappa_{N^*})$ are $N^* \times N^*$ diagonal parameter matrices, where $N^* = N(N-1)/2$. We restrict Φ and K to be diagonal, as otherwise the number of parameters would be of order $O(N^4)$, which is infeasible in high dimensions. However, similar to the volatility case, one might consider “log correlation spillover”, and tests thereof, but interpretation is not obvious since the link between the log parameters and correlations is not trivial.

The specific form of the score vector u_t depends of course on the conditional distribution, which will be discussed in detail in the following in the context of maximum likelihood estimation. We distinguish between a multivariate student-t distribution, which is used for one-step estimation, and a Meta-t distribution which is used for two-step estimation. The score vectors in both cases are very similar but not exactly the same. Furthermore, one may use standardized versions of the score vector, by pre-multiplying u_t by the inverse of the (square root of the) information matrix, see e.g. the model of [Harvey \(2013\)](#), Section 7.3.1., which is motivated in particular by cases where the standardized score is an i.i.d. random vector. However, this will generally not be the case in our setting, so we will not pursue this strategy here to keep things simple.

Although one of the main advantages of our model specification lies in the separation of volatility and correlation parts, just as in [Engle \(2002\)](#) and related works, it would be more efficient from a statistical viewpoint to do the estimation of all model parameters in one step. In the following, we therefore present both approaches, one-step and two-step estimation.

2.3.1 One-step estimation

For one-step estimation we assume for simplicity that $\eta_1 = \dots = \eta_N = \eta$, so that the conditional joint distribution of y_t is a multivariate student-t. The log likelihood function for a sample of T observations as a function of the parameter vector $\vartheta = (\tau^\top, \delta^\top, \eta)^\top$, $\delta := (\omega^\top, \text{diag}(\Phi)^\top, \text{diag}(K)^\top)^\top$ takes the form

$$L(\vartheta) = \sum_{t=1}^T l_t(\vartheta) = T \log \Gamma\left(\frac{\eta + N}{2}\right) - T \log \Gamma\left(\frac{\eta}{2}\right) - \frac{TN}{2} \log((\eta - 2)\pi) \\ - \frac{1}{2} \sum_{t=1}^T \left\{ 2 \log |D_t(\tau)| + \log |\Theta_t(\delta)| + (\eta + N) \log \left(1 + \frac{1}{\eta - 2} \varepsilon_t^\top \Theta_t(\delta)^{-1} \varepsilon_t \right) \right\} \quad (2.7)$$

The spectral decomposition of the log correlation matrix is denoted by $\log \Theta_t = \Gamma_t \Lambda_t \Gamma_t^\top$, where Γ_t is the $(N \times N)$ matrix whose columns are the eigenvectors of $\log \Theta$, and Λ_t is the $(N \times N)$ diagonal matrix containing the corresponding eigenvalues $\lambda_{it}, i = 1, \dots, N$ on the diagonal. Define the matrix

$$A_t := (\Gamma_t \otimes \Gamma_t) \Xi_t (\Gamma_t \otimes \Gamma_t)^\top, \quad (2.8)$$

where Ξ_t is the $N^2 \times N^2$ matrix with elements given by

$$\xi_{ij} = \Xi_{(i-1)N+j, (i-1)N+j} = \begin{cases} \exp(\lambda_{it}), & \text{if } \lambda_{it} = \lambda_{jt} \\ \frac{\exp(\lambda_{it}) - \exp(\lambda_{jt})}{\lambda_{it} - \lambda_{jt}}, & \text{if } \lambda_{it} \neq \lambda_{jt} \end{cases}$$

for $i, j = 1, \dots, N$. Furthermore, define the following elimination matrices: E_d ($N \times N^2$), E_l ($(N(N-1)/2 \times N^2)$), and E_u ($(N(N-1)/2 \times N^2)$) such that $\text{vecl}(Q) = E_l \text{vec}(Q)$, $\text{vecl}(Q^\top) = E_u \text{vec}(Q)$ and $\text{diag}(Q) = E_d \text{vec}(Q)$ for any square matrix Q , and $E_s := E_l + E_u$. Finally, define the permutation matrix P as $P(\text{diag}(Q)^\top, \text{vecl}(Q)^\top)^\top = \text{vec}(Q)$ for any symmetric conformable matrix Q . With this notation, the following theorem provides the analytic form of the score with respect to the log correlation parameter.

Theorem 1. *The score with respect to the log correlation parameter θ_t is given by*

$$u_t = \frac{\partial l_t}{\partial \theta_t} = \frac{1}{2} M_t \text{vec} \left(\frac{\eta + N}{\eta - 2 + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t} \Theta_t^{-1} \varepsilon_t \varepsilon_t^\top \Theta_t^{-1} - \Theta_t^{-1} \right) \quad (2.9)$$

where

$$M_t := \begin{pmatrix} -(E_d A_t E_d^\top)^{-1} E_d A_t E_s^\top \\ I_{N(N-1)/2} \end{pmatrix}^\top P^\top A_t \quad (2.10)$$

Remark 1. In the limiting case $\eta \rightarrow \infty$, we obtain the Gaussian distribution as a special case, for which we let $\vartheta = (\tau^\top, \delta^\top)^\top$, and the log likelihood function takes the form

$$L(\vartheta) = -\frac{TN}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \left\{ 2 \log |D_t(\tau)| + \log |\Theta_t(\delta)| + \varepsilon_t^\top \Theta_t^{-1}(\delta) \varepsilon_t \right\} \quad (2.11)$$

and the score reduces to

$$u_t = \frac{\partial l_t}{\partial \theta_t} = \frac{1}{2} \begin{pmatrix} -(E_d A_t E_d^\top)^{-1} E_d A_t E_s^\top \\ I_{N(N-1)/2} \end{pmatrix}^\top P^\top A_t \text{vec}(\Theta_t^{-1} \varepsilon_t \varepsilon_t^\top \Theta_t^{-1} - \Theta_t^{-1}) \quad (2.12)$$

Remark 2. Note that the score can be written as $u_t = g(\theta_t) Z_t$ where $g(\theta_t) = \frac{1}{2} M_t(\Theta_t^{-1/2} \otimes \Theta_t^{-1/2})$, and $Z_t = \text{vec}\left(\frac{\eta+N}{\eta-2+\xi_t^\top \xi_t} \xi_t \xi_t^\top - I_N\right)$ is an i.i.d. random vector. The score vector u_t is a martingale difference sequence, which can be shown using results of [Fiorentini et al. \(2003\)](#). However, it is not i.i.d. in general, although it is i.i.d. in the univariate case because, in that case, $g(\theta_t) = 1/2$.

Remark 3. The process θ_t has an SRE representation as

$$\theta_t = \psi_t(\theta_{t-1}) \quad (2.13)$$

where $\psi_t(s) = \omega + \Phi s + K g(s) Z_t$. Appendix 2.A.1 shows that, under our assumptions, this is a Lipschitz map with maximum Lipschitz constant $\Lambda(\psi) < \infty$ a.s. We derive an analytical expression for $\Lambda(\psi)$ in the bivariate case, while in higher dimensions, we show that there exists an upper bound. See Appendix 2.A.1 and Appendix 2.D for details.

The volatility part of the likelihood, i.e., $D_t(\tau)$ in Equation (2.7), is straightforward to calculate and exactly the same as in univariate Beta-t-EGARCH models. For the correlation part we use the following algorithm to compute $\Theta_t(\delta)$ and the score u_t , conditional on the volatility parameter τ :

1. Initialization: Set $u_0 = 0$ and $\theta_0 = (I_{N(N-1)/2} - \Phi)^{-1} \omega$.
2. For $t = 1, \dots, T$:
 - (a) Calculate θ_t as a function of past observations according to Equation (2.6).
 - (b) Reconstruct $\text{diag}(\log \Theta_t)$ using the algorithm of [Archakov and Hansen \(2021\)](#).
 - (c) Calculate the matrix exponential of $\log \Theta_t$ to obtain Θ_t .
 - (d) Calculate the score u_t according to Equation (2.9).

3. Calculate the log likelihood according to Equation (2.7).

The MLE is then defined as the maximizer of the log likelihood function over some parameter space Ω ,

$$\hat{\vartheta} = \arg \max_{\vartheta \in \Omega} L(\vartheta). \quad (2.14)$$

No analytical solution is available, but standard numerical optimization routines can be employed. For the estimation theory we use the following assumptions.

- (A1) The conditional distribution of y_t is a standardized multivariate student-t distribution with η degrees of freedom, $\eta > 2$ and dispersion matrix H_t .
- (A2) The parameter vector $\vartheta \in \Omega$, where Ω is a compact subset of $\mathbb{R}^J$, $J = \dim(\vartheta)$, such that $|\phi_i^v| < 1, i = 1, \dots, N$, $|\phi_j| < 1, j = 1, \dots, N^*$, $\kappa_i^v \neq 0, i = 1, \dots, N$, and $\kappa_j \neq 0, j = 1, \dots, N^*$.
- (A3) The true parameter ϑ_0 lies in the interior of Ω .
- (A4) $|(\phi_i^v)^2 + 2\kappa_i^v \phi_i^v \mathbb{E}(\partial u_{it}^v / \partial h_{it})| + (\kappa_i^v)^2 \mathbb{E}[(\partial u_{it}^v / \partial h_{it})^2] < 1, \quad i = 1, \dots, N$
- (A5) $|\phi_j^2 + 2\kappa_j \phi_j \mathbb{E}(\partial u_{jt} / \partial \theta_{jt})| + \kappa_j^2 \mathbb{E}[(\partial u_{jt} / \partial \theta_{jt})^2] < 1, \quad j = 1, \dots, N^*$
- (A6) The smallest eigenvalue of Θ_t is bounded away from zero by a positive constant, $\lambda_{\min}(\Theta_t) \geq \epsilon > 0$, a.s. We call $\mathcal{T}_{\epsilon, N}$ the set of log-parameters θ_t for which this assumption holds, i.e. we assume $\theta_t \in \mathcal{T}_{\epsilon, N} \subset \mathbb{R}^{N(N-1)/2}$ a.s.
- (A7) $\mathbb{E}[\log^+(\Lambda(\psi_t))] < \infty$ and $\mathbb{E}[\log(\Lambda(\psi_t))] < 0$.

Assumption (A2) is standard and necessary for identifiability and stationarity of the volatility and log correlation processes. (A3) is also standard and is needed for asymptotic normality. (A4) and (A5) ensure that the scores with respect to log volatility and log correlation, respectively, and their first derivatives, have finite second moments. They are also standard in this literature, see e.g. the conditions of Theorem 1 of [Harvey \(2013\)](#). Assumption (A6) ensures that the stochastic recurrence equation is Lipschitz continuous, as shown in Appendix 2.D. It is standard in the literature on modelling high-dimensional covariance matrices, see e.g. [Fan et al. \(2011\)](#) and [Chang et al. \(2018\)](#). Assumption (A7) is a contraction condition which ensures that the process θ_t is a stationary and ergodic sequence. For given parameters and lower bound ϵ for

the smallest eigenvalue of Θ_t , it can be verified by numerical simulation, see Appendix 2.D for details.

Let $\vartheta = (\tau^\top, \delta^\top, \eta)^\top$, and decompose the information matrix as

$$I = \begin{pmatrix} I_{\tau\tau} & I_{\tau\delta} & I_{\tau\eta} \\ I_{\tau\delta}^\top & I_{\delta\delta} & I_{\delta\eta} \\ I_{\tau\eta}^\top & I_{\delta\eta}^\top & I_{\eta\eta} \end{pmatrix} \quad (2.15)$$

We then have the following result.

Theorem 2. *Under Assumptions (A1)-(A7), $\hat{\vartheta} \rightarrow_p \vartheta$ and the asymptotic distribution of the estimator obtained in Equation (2.14) is given by*

$$\sqrt{T}(\hat{\vartheta} - \vartheta) \xrightarrow{\mathcal{L}} N(0, I^{-1}) \quad (2.16)$$

where I is given in Equation (2.15) with

$$\begin{aligned} I_{\delta\delta} &= \mathbb{E} \left[\frac{\partial \theta_t^\top}{\partial \delta} M_t \left\{ \frac{N + \eta}{2(N + \eta + 2)} (\Theta_t^{-1} \otimes \Theta_t^{-1}) - \frac{1}{2(N + \eta + 2)} \text{vec}(\Theta_t^{-1}) \text{vec}(\Theta_t^{-1})^\top \right\} M_t^\top \frac{\partial \theta_t}{\partial \delta^\top} \right] \\ I_{\eta\eta} &= \frac{\eta^4}{4} \left[\psi' \left(\frac{\eta}{2} \right) - \psi' \left(\frac{N + \eta}{2} \right) \right] - \frac{N\eta^4[\eta^2 + N(\eta - 4) - 8]}{2(\eta - 2)^2(N + \eta)(N + \eta + 2)} \\ I_{\delta\eta} &= -\frac{(N + 2)\eta^2}{(\eta - 2)(N + \eta)(N + \eta + 2)} \mathbb{E} \frac{\partial \theta_t^\top}{\partial \delta} M_t \text{vec}(\Theta_t^{-1}) \\ I_{\tau\tau} &= \mathbb{E} \left[\frac{\partial h_t^\top}{\partial \tau} \left\{ \frac{N + \eta}{2(N + \eta + 2)} (H_t^{-1} \otimes H_t^{-1}) - \frac{1}{2(N + \eta + 2)} \text{vec}(H_t^{-1}) \text{vec}(H_t^{-1})^\top \right\} \frac{\partial h_t}{\partial \tau^\top} \right] \\ I_{\tau\delta} &= \mathbb{E} \left[\frac{\partial h_t^\top}{\partial \tau} \left\{ \frac{N + \eta}{2(N + \eta + 2)} (H_t^{-1} \otimes H_t^{-1}) - \frac{1}{2(N + \eta + 2)} \text{vec}(H_t^{-1}) \text{vec}(H_t^{-1})^\top \right\} (D_t \otimes D_t) M_t^\top \frac{\partial \theta_t}{\partial \delta^\top} \right] \\ I_{\tau\eta} &= -\frac{(N + 2)\eta^2}{(\eta - 2)(N + \eta)(N + \eta + 2)} \mathbb{E} \frac{\partial \text{vec}^\top(H_t)}{\partial \tau} \text{vec}(H_t^{-1}) \end{aligned}$$

The information matrix further depends on the derivative of the log correlation parameter θ_t with respect to δ , which is given by

$$\frac{\partial \theta_t}{\partial \delta^\top} = z_{t-1} + x_{t-1} \frac{\partial \theta_{t-1}}{\partial \delta^\top}, \quad (2.17)$$

where

$$z_t = \begin{pmatrix} I_{N^*} & \text{diag}(\theta_{1t}, \dots, \theta_{N^*t}) & \text{diag}(u_{1t}, \dots, u_{N^*t}) \end{pmatrix}, \quad x_t = \Phi + K \frac{\partial u_t}{\partial \theta_t^\top}. \quad (2.18)$$

The information matrix can be estimated consistently by evaluating it at $\hat{\vartheta}$ and replacing the expectation operators by sample averages.

2.3.2 Two-step estimation

For two-step estimation we allow for different d.o.f. parameters of the marginal distributions, $\eta_1, \dots, \eta_N$. Thus we now assume a conditional Meta-t distribution with degrees of freedom parameter η and scale matrix Θ_t , meaning that the conditional dependence structure of y_t is determined by a student-t copula with scale matrix Θ_t , see e.g. [Demarta and McNeil \(2005\)](#) for details on Meta-t distributions and the student-t copula.

As before, the conditional marginal distributions are given by score-driven volatility processes and student-t distributions, but now with potentially different degrees of freedom η_i . If all d.o.f. parameters happen to be the same and equal to η , then this model reduces to the one considered for one-step estimation, and in this case two-step estimation would be less efficient than one-step estimation, although it might be the only feasible estimation procedure computationally, at least in high dimensions.

Formally, for the theory we replace Assumption (A1) by the following.

(A1') The conditional distribution of y_t is given by

$$F_t(y_{1t}, \dots, y_{Nt}) = C_{\eta, \Theta_t}(F_{\eta_1, h_{1t}}(y_{1t}), \dots, F_{\eta_N, h_{Nt}}(y_{Nt}))$$

with $\eta > 2$ and $\eta_i > 2$, $i = 1, \dots, N$, where $F_{\eta_i, h_{it}}$ are the standardized student-t marginal cdf's with log scale h_{it} , and C_{η, Θ_t} is the cdf of a standardized student-t copula with dispersion matrix Θ_t .

Note that the copula is invariant to standardizations of the marginal distributions, so that the student-t copula with scale matrix H_t is the same as the student-t copula with scale matrix Θ_t . Note also that if all marginal degrees of freedom are the same and equal to the copula parameter η , then Assumption (A1') collapses to (A1).

We augment the vector of parameters characterizing the marginal distributions as $\tau_i = (\omega_i^v, \phi_i^v, \kappa_i^v, \eta_i)^\top$, $i = 1, \dots, N$, and $\tau = (\tau_1^\top, \dots, \tau_N^\top)^\top$. The copula density has the form

$$c_{\eta, \Theta_t}(u) = \frac{f_{\eta, \Theta_t}(t_\eta^{-1}(u_1), \dots, t_\eta^{-1}(u_N))}{\prod_{i=1}^N g_\eta(t_\eta^{-1}(u_i))}$$

see e.g. [Demarta and McNeil \(2005\)](#), where $f_{\eta, \Theta_t}(\cdot)$ is the joint pdf of a multivariate standardized student-t distribution with scale matrix Θ_t , and $t_\eta(\cdot)$ and $g_\eta(\cdot)$ are standardized univariate

student-t cdf and pdf, respectively. The conditional joint density of y_t is then given by

$$f_t(y_t) = c_{\eta, \Theta_t}(F_{\eta_1, h_{1t}}(y_{1t}), \dots, F_{\eta_N, h_{Nt}}(y_{Nt})) \prod_{i=1}^N g_{\eta_i, h_{it}}(y_{it})$$

which is used to construct the likelihood function.

The log likelihood function for a sample of T observations as a function of the parameter vector $\vartheta = (\tau^\top, \delta^\top, \eta)^\top$, $\delta := (\omega^\top, \phi^\top, \kappa^\top)^\top$ takes the form

$$\begin{aligned} L(\vartheta) &= L^V(\tau) + L^C(\vartheta) \\ L^V(\tau) &= \sum_{i=1}^N l_i^V(\tau) \\ l_i^V(\tau) &= c(\eta_i, 1) - \frac{1}{2} \sum_{t=1}^T \left\{ 2h_{it} + (\eta_i + 1) \log \left(1 + \frac{y_{it}^2 \exp(-2h_{it})}{\eta_i - 2} \right) \right\} \\ L^C(\vartheta) &= c(\eta, N) - Nc(\eta, 1) - \frac{1}{2} \sum_{t=1}^T l_t^C(\vartheta) \\ l_t^C(\vartheta) &= \log |\tilde{\Theta}_t(\delta)| + (\eta + N) \log \left(1 + \frac{\tilde{\varepsilon}_t^\top \tilde{\Theta}_t(\delta)^{-1} \tilde{\varepsilon}_t}{\eta - 2} \right) - \sum_{i=1}^N (\eta + 1) \log \left(1 + \frac{\tilde{\varepsilon}_{it}^2}{\eta - 2} \right) \\ c(\eta, N) &= T \log \Gamma \left(\frac{\eta + N}{2} \right) - T \log \Gamma \left(\frac{\eta}{2} \right) - \frac{TN}{2} \log((\eta - 2)\pi) \end{aligned}$$

where $\tilde{\varepsilon}_{it} := s_\eta(\varepsilon_{it}, \eta_i) = t_\eta^{-1}(t_{\eta_i}(\varepsilon_{it}))$ is a mapping of the quantiles of a standardized student-t distribution with η_i d.o.f. into the quantiles of a standardized student-t with η d.o.f. The mapping function $s_\eta(\cdot)$ is strictly monotone and its inverse is given by $s_\eta^{-1}(\cdot, \eta_i) = t_{\eta_i}^{-1}(t_\eta(\cdot))$. When applied to a vector, we define $s_\eta(\cdot)$ as $s_\eta(\varepsilon_t, \underline{\eta}) = (s_\eta(\varepsilon_{1t}, \eta_1), \dots, s_\eta(\varepsilon_{Nt}, \eta_N))^\top$, where $\underline{\eta} = (\eta_1, \dots, \eta_N)^\top$, and similarly for $s_\eta^{-1}(\varepsilon_t, \underline{\eta})$.

Note that the likelihood reduces to that of the one-step estimation when imposing the restriction $\eta_1 = \dots = \eta_N = \eta$.

The matrix $\tilde{\Theta}_t(\delta)$ is such that $\text{vecl}(\log \tilde{\Theta}_t(\delta)) =: \tilde{\theta}_t(\delta)$ follows the same process specified in Equation (2.6) as above for the one-step method, but in the score u_t replacing $\Theta_t(\delta)$ by $\tilde{\Theta}_t(\delta)$ and ε_t by $\tilde{\varepsilon}_t$, i.e.

$$\begin{aligned} \tilde{\theta}_t(\delta) &= \omega + \Phi \tilde{\theta}_{t-1}(\delta) + K \tilde{u}_{t-1} \\ \tilde{u}_t &= \frac{1}{2} \tilde{M}_t \text{vec} \left(\frac{\eta + N}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} - \tilde{\Theta}_t^{-1} \right) \end{aligned}$$

and where $\tilde{M}_t$ is constructed as M_t above but replacing Θ_t by $\tilde{\Theta}_t(\delta)$.

Now, due to the nonlinear transformation $s_\eta(\cdot)$, $\tilde{\Theta}_t(\delta)$ is not exactly equal to the conditional

correlation matrix $\Theta_t(\delta)$. It can however be reconstructed easily by numerical methods. For example, for a given $\tilde{\Theta}_t$, generate a large number K of r.v. X with standardized multivariate student-t distribution with η d.o.f. and correlation matrix $\tilde{\Theta}_t$. Then, obtain the sample correlation matrix of the transformed data as

$$\Theta_t = \frac{1}{K} \sum_{k=1}^K s_{\eta}^{-1}(X_k, \underline{\eta})(s_{\eta}^{-1}(X_k, \underline{\eta}))^{\top}$$

The diagonal elements of Θ_t should be very close to one if K is sufficiently large. To illustrate this mapping of correlations, we plot in [Figure 2.1](#) the correlation ρ as a function of $\tilde{\rho}$ for the case $\eta = 7$, $\eta_1 = 4$ and $\eta_2 = 10$, obtained by simulating $K = 10^7$ bivariate student-t random variables. If all degrees-of-freedom parameters were the same, then this would be the identity function. Here this function is different, but very close to the identity function, with the difference shown on the right of the [Figure 2.1](#). Hence, the correlations change after transformation but only slightly so. In our empirical application of Section 4, where the marginal degrees of freedom are of a similar range as in the numerical example, we therefore decided to skip this rescaling and set $\Theta_t = \tilde{\Theta}_t$, because it practically makes no difference and alleviates the computational burden.

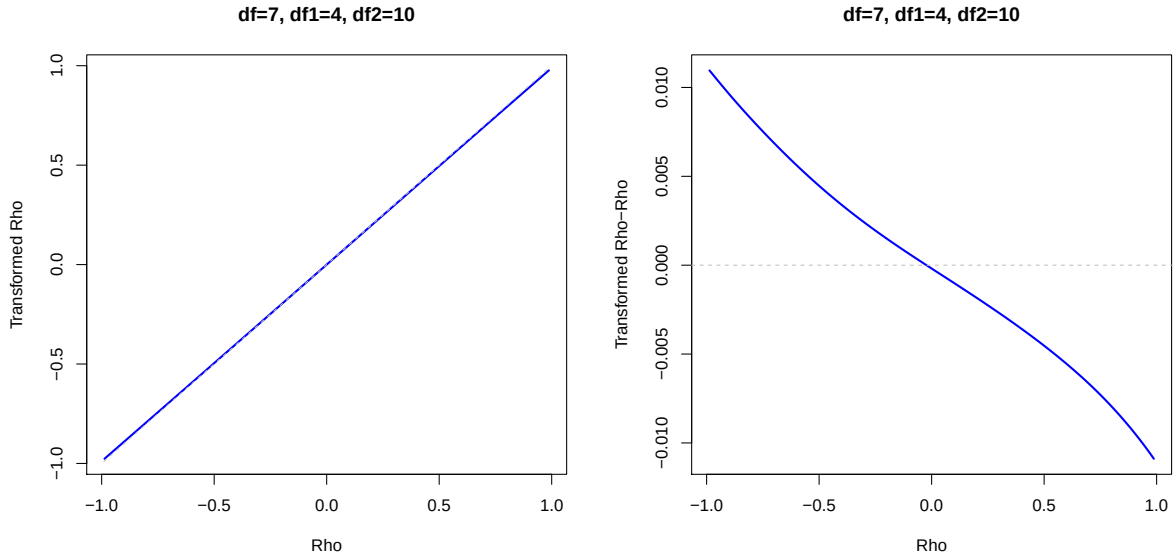


Figure 2.1: The left panel shows the correlation ρ as a function of $\tilde{\rho}$, where $\tilde{\rho}$ is the correlation parameter of a bivariate standardized student-t r.v. X with $\eta = 7$ d.o.f., and ρ is the correlation of transformed r.v. $s_{\eta}^{-1}(X_1, \eta_1)$ and $s_{\eta}^{-1}(X_2, \eta_2)$ with $\eta_1 = 4$ and $\eta_2 = 10$. The right panel shows the difference of this function from the identity function.

The sum of the marginal log likelihood functions, $L^V(\tau)$, only depends on the marginal parameters, not on the copula parameters. It is in fact equal to the sum of the univariate volatility log likelihoods that are equal to the Beta-t-EGARCH model log likelihoods. In the first step, the inference function for margins approach proposed by Joe (1997) maximizes $L^V(\tau)$ with respect to τ to obtain $\hat{\tau}$, which is then used in a second step to maximize $L^C(\hat{\tau}^\top, \delta, \eta)^\top$ with respect to δ and η , to give $\hat{\delta}$ and $\hat{\eta}$. Then, $\hat{\vartheta} = (\hat{\tau}^\top, \hat{d}^\top)^\top$, where $d = (\delta^\top, \eta)^\top$ is the parameter characterizing the copula, and $\hat{d} = (\hat{\delta}^\top, \hat{\eta})^\top$.

We now have the following result.

Theorem 3. *Under Assumptions (A1')-(A7), $\hat{\vartheta} \rightarrow_p \vartheta$ and*

$$\sqrt{T}(\hat{\vartheta} - \vartheta) \rightarrow_d N(0, (-F)^{-1}G(-F)^{-1})$$

where

$$-F = \begin{pmatrix} \mathcal{J}_{11} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathcal{J}_{NN} & 0 \\ \mathcal{I}_{d1} & \cdots & \mathcal{I}_{dN} & \mathcal{I}_{dd} \end{pmatrix}, \quad G = \begin{pmatrix} \mathcal{J}_{11} & \cdots & \mathcal{J}_{1N} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{J}_{N1} & \cdots & \mathcal{J}_{NN} & 0 \\ 0 & \cdots & 0 & \mathcal{I}_{dd} \end{pmatrix}$$

and where

$$\begin{aligned} \mathcal{J}_{ij} &= \mathbb{E} \left[\frac{\partial l_i^V(\tau_i)}{\partial \tau_j} \frac{\partial l_j^V(\tau)}{\partial \tau^\top} \right], \quad i, j = 1, \dots, N \\ \mathcal{I}_{dd} &= -\mathbb{E} \left[\frac{\partial^2 l(\vartheta)}{\partial d \partial d^\top} \right], \end{aligned} \tag{2.19}$$

$$\mathcal{I}_{di} = -\mathbb{E} \left[\frac{\partial^2 l(\vartheta)}{\partial d \partial \tau_i^\top} \right], \quad i = 1, \dots, N \tag{2.20}$$

Note that the matrices $\mathcal{J}_{ii}, i = 1, \dots, N$, are the information matrices of the marginal likelihoods, so well-known results of univariate Beta-t-EGARCH models can be used for their calculation, see e.g. Proposition 20 of Harvey (2013).

Analytical expressions for the components of $\mathcal{I}$ in Equation (2.19) and Equation (2.20) are given in Appendix 2.E. They can be used, as for one-step estimation, to estimate consistently the asymptotic covariance matrix by evaluating it at $\hat{\vartheta}$ and replacing the expectation operators by sample averages. We will use this two-step estimation approach in our empirical application, to which we turn in the following.

2.4 Empirical application

Modelling the day-to-day changes in the correlation structure of financial assets plays an important role in the financial decision making process. We take the standpoint of a representative investor who, in a context of near-zero interest rates, considers not only classical assets such as stocks and bonds, but also commodities (oil), safe-haven investments (gold), and alternative investments (cryptocurrencies). For each market we take a representative index and then model the set of market indices using our multivariate framework. The rest of this section is organized as follows: Section 2.4.1 presents the data description and sample coverage for our in-sample and out-of-sample analyses; Section 2.4.2 covers the results for the estimation of the conditional volatility dynamics; Section 2.4.3 covers the results for the estimation of the conditional correlation dynamics. We conduct an in-depth benchmarking of our proposed model with alternative approaches introduced in the literature both for the in-sample and out-of-sample exercises.

2.4.1 Data description

We use historical data on daily log-returns for six representative indices of different asset classes, including both traditional financial assets (i.e. stocks and bonds) as well as commodities and alternative investments (i.e. oil, gold and cryptocurrencies). The sample period spans from August 1, 2014 until December 11, 2020, with the starting date determined by the first available data point of the CRIX cryptocurrency index, see [Trimborn and Härdle \(2018\)](#) for details on the construction of CRIX. The symbols and description for the indices are shown in [Table 2.6](#).

The time series of prices and returns are shown in [Figure 2.2](#) and [Figure 2.6](#), respectively. We divide the whole sample series into the in-sample period from 2014 to the end of 2018 and the out-of-sample period from 2019 to 2020. We notice that the CRIX index exhibited bubble-like features in 2017 and possibly end of 2020. [Table 2.7](#) presents the main descriptive statistics for our historical return series. Alternative investments generally have larger mean returns than stocks or bonds, with higher associated volatilities. The equity index EMU has much larger skewness and kurtosis than the SPX index, which are primarily driven by a single abnormal return in 2016 and by more negative returns during the period of the COVID-19 crisis. Equity and bond returns are generally negatively skewed, corresponding to a long tail in negative returns, whereas the oil index is positively skewed. In addition, the return series for

the gold index exhibits approximate symmetry and the CRIX index has a significantly negative skewness, which may be explained by the market downturn of 2018. Furthermore, normality is clearly rejected for all assets by classical tests at conventional significance levels.

We report the sample correlation matrix of historical returns for the six market indices in [Table 2.8](#). For financial indices, SPX and EMU are highly correlated with a Pearson correlation coefficient of 0.58. The negative or low coefficients between pairs of SPX-SBW (-0.14) and SPX-GOLD (0.08) confirm the common view that bonds and gold often serve as an effective hedge against stock market risk. This characteristic also results in the fact that the correlation between gold and bonds is relatively high with a coefficient of 0.55. Oil is more closely correlated with the equity indices (0.28 with SPX and 0.25 with EMU), while correlation with other assets is negligible.

2.4.2 Volatility estimation results

We follow the two-step estimation procedure outlined in [Section 2.3.2](#), where the first step consists in fitting volatility models to the return series. For each series, we fit a Beta-t-EGARCH(1,1) model and the model is estimated by maximum likelihood under the assumption of a conditional student-t distribution. In order to improve the convergence of the optimization algorithm, we estimate the inverse of the degrees of freedom parameter, i.e. $\bar{\eta} = 1/\eta$. The results are shown in [Table 2.9](#). All parameter estimates are significantly different from zero using estimated asymptotic standard errors. Moreover, the estimates for the autoregressive coefficient ϕ_i indicate high persistence in volatility for all time series and thus confirm the presence of volatility clustering. Note that the CRIX index has the lowest estimated d.o.f. $\eta_i = 1/\bar{\eta}_i$ which corresponds to its observed high kurtosis and thick tails. In [Figure 2.7](#), we plot the estimated volatility series estimated using the Beta-t-EGARCH model. We observe high persistence in the estimated volatility series, which is consistent with the estimates for the autoregressive coefficient being close to unity.

We now focus on assessing the validity of the assumptions for the Beta-t-EGARCH model in terms of the residuals distribution and in terms of the score series. Firstly, we test whether the score series is a martingale difference sequence (MDS) using a Box-Pierce test for autocorrelation. The p-values are displayed in [Table 2.10](#). Since the p-values are larger than 5% for all six indices, we fail to reject the null hypothesis that the score series have zero autocorrelation up

to a lag of 30. This result provides support to the validity of the martingale difference property. Furthermore, we apply the Box-Pierce test to the standardized residuals and their squares. The results, also reported in [Table 2.10](#), indicate that significant autocorrelation exists only for the residuals of the CRIX index, whereas the squared residuals for all assets are not serially correlated at the 5% significance level.

The descriptive statistics of standardized residuals are summarized in [Table 2.11](#). As expected, the sample means are close to zero and the standard deviations close to one. The only exception is for the CRIX index whose residuals have a standard deviation of 0.78. Again, this may be caused by some remaining serial correlation in the residuals and/or outlier observations. There is remaining excess kurtosis in the standardized residual series, which implies that a conditionally leptokurtic distribution might provide a better fit to the data than the standard normal distribution.

Finally, we use the probability integral transformation (PIT) to check how well the assumed conditional distribution fits the data. In [Figure 2.8](#), we display the histograms of PITs of residual series, assuming respectively a student-t distribution in the left-hand column and a normal distribution in the right-hand column. Most of the histograms for the normal distribution have a hump in the middle and pronounced wings in the tails, which corresponds to our finding of leptokurtic standardized residuals in [Table 2.11](#). Although not perfect, the histograms of the PITs for the student-t distribution are much closer to a uniform distribution than the Gaussian PITs. Introducing skewness may be beneficial in improving the fit of the conditional distribution of the data. However, we did not pursue this avenue in our analysis and leave it as a potential extension for future research.

In the next section, we move on to the second step of the two-step estimation and focus on the correlation part of the model, taking the standardized residual series of the univariate volatility models as inputs for the dynamic correlation model.

2.4.3 Results for the conditional correlations

We now turn to the results for the second step estimation regarding the conditional correlation matrix. First, we want to test the reliability of our assumption of diagonal K and Φ matrices. Estimating a model with six assets and full matrices K and Φ is computationally infeasible as it would entail the estimation of 466 parameters. Instead, we chose five subsets composed of three

out of six assets with relatively high correlations and perform likelihood-ratio tests to assess the goodness of fit for the diagonal case (H_0) with respect to these alternative specifications. We fail to reject the null hypothesis H_0 at the 5% significance level for the five specifications considered. This result corroborates our choice of diagonal parameter matrices for the analysis.²

The estimated parameters for the models with diagonal K and Φ are reported in [Table 2.12](#). All estimated κ_i coefficients are positive and close to zero, while most of the estimated ϕ_i are close to one. Similar to the volatility case, the correlation series are characterized by stationary dynamics but high persistence. Furthermore, we notice that the inverse of the degrees of freedom parameter is approximately 0.029. This result indicates that the t-copula modelling the residual series after de-volatilization is close to, but significantly different from, the Gaussian copula.

Although diagonality constraints of the parameter matrices highly reduce the complexity of the model, we still have 46 parameters to estimate. To further simplify the model, we consider five additional model restrictions. We notice from the results in [Table 2.12](#) that the parameter estimates for ω , K , and Φ have quite similar values. This observation motivates alternative specifications in which parameters are restricted to be equal. [Table 2.13](#) provides a summary of the different combinations of restrictions considered with the corresponding number of parameters. In the interest of space, we do not provide the estimates of the restricted models and only provide a summary of model performance measures in the following paragraph.

In terms of model diagnostics, we first apply Portmanteau tests to the correlation score series, the correlation residual series, and the outer product of the correlation residual series. The results are reported in [Table 2.1](#). For all the log correlation series, there is no evidence against the martingale difference property for the score series at conventional significance levels. The p-values of the test for the correlation residual series indicate that we cannot reject the null hypothesis of no serial autocorrelation at the 1% significance level, whereas we do not find evidence supporting the presence of serial autocorrelation in the outer product of the correlation residual series at all conventional significance levels. Furthermore, we perform likelihood-ratio tests of the restricted models with respect to the diagonal one. It can be observed that we reject the null hypothesis of equal likelihood between the restricted models and Model 1 at the 5% significance level. We reach the same conclusion when considering the 1% significance level except in the case of Model 3. We additionally report the Akaike and Bayesian information criteria (AIC and

²Results for the specification analysis are available from the authors upon request.

BIC), to quantify the model performance in terms of goodness-of-fit and model complexity. We conclude that model 2 with 18 parameters should be selected according to the minimum BIC and AIC criteria. In this case, the evidence is in favor of restricting the diagonal elements of K and those of Φ to be equal, which reduces the parameter dimension substantially, while allowing variation for ω across different series. This result also suggests that the dynamics of the log correlation series θ_i can be assumed to be the same across different components, and that the only difference occurs in the level of the series. In higher dimensions, we conjecture that similar parameter restrictions are reasonable and allow us to obtain computationally feasible results in cases of several hundreds of assets. This actually corresponds to a similar observation made by [Caporin and McAleer \(2012\)](#) for the classes of BEKK and DCC models, for which the “scalar” versions of the models are preferred in high dimensions to avoid the curse of dimensionality.

2.4.3.1 In-sample performance

In this part, we conduct an empirical study for model performance comparison considering the GARCH-cDCC, the GAS class of models proposed by [Creal et al. \(2011\)](#) and our model. We categorize the 8 models into two classes. The Meta-t models on the one hand are estimated using a two-step estimation procedure. In this case, the volatilities are all estimated by a Beta-t-EGARCH model assuming different degrees of freedom for each asset. The dynamic correlations are modelled with different specifications, namely, our log correlation matrix model (Meta-t-log-corr), the cDCC model (Meta-t-cDCC), the GAS model introduced in [Creal et al. \(2011\)](#) (Meta-t-GAS), and its variation using hyperspherical coordinates decomposition (Meta-t-GAS-h). The second group of models, on the other hand, contains the multivariate t models estimated by a one-step procedure. In particular, the t-cDCC, t-GAS, t-GAS with log-volatility (t-GAS-l) specification are estimated using variance and correlation targeting techniques. We also consider the t-GAS model with a hyperspherical coordinates decomposition (t-GAS-hl) for comparison.

Note that all models based on the Meta-t distribution face the problem of rescaling of Θ_t that is discussed in Section 3.2, because for a Meta-t distribution, $\tilde{\Theta}_t$ is not exactly equal to the correlation matrix. However, as we have shown in our example with marginal d.o.f. equal to 4 and 10, and a copula d.o.f. of 7, the approximation is very accurate. The range of marginal d.o.f. is similar for our data. We therefore decided to skip the rescaling and set $\Theta_t = \tilde{\Theta}_t$, which

practically makes no difference but simplifies the computation of all considered Meta-t models.

The parameters estimated from the group of Meta-t models are reported in [Table 2.14](#).³ It can be observed that the autoregressive parameter values Φ for all models are close to unity indicating high persistence in the correlation series. Moreover, the inverse of the degrees of freedom parameter is close to zero (around 0.033) but significantly different from zero. [Table 2.15](#) presents the parameter estimation of the group of multivariate t models. For each model, the parameters α_i and β_i with $i = 1, 2, \dots, 6$ are the volatility parameters while α_7 and β_7 are the correlation parameters with different specifications. Similarly, all the autoregressive parameters are close to one such that the estimated volatility series and correlation series are highly persistent.

The Meta-t log correlation specification has the best model fit in terms of log likelihood values, reported in [Table 2.2](#). The differences between the likelihood values of the four Meta-t variations are quite small and would not be significant using likelihood ratio tests, if the models were nested. However, the difference between the Meta-t models and the multivariate student-t models are quite substantial. Therefore, the added flexibility in the Meta-t specification allowing for different degrees of freedom for individual assets outperforms the specification using a multivariate t distribution with a single degrees of freedom parameter by a larger log likelihood of around 80 to 100. Furthermore, since the only difference among the four Meta-t models comes from the parameterization of the correlation matrix, the results indicate that the different specifications of the correlation dynamics after the devolatilization appear to make a smaller difference. This observation is consistent with the empirical results in [Creal et al. \(2011\)](#). Based on the third and fourth rows in [Table 2.2](#), the Meta-t log correlation model would be selected using the classical model selection criteria AIC and BIC. The superior performance of the Meta-t models is confirmed by the AIC and BIC results although the multivariate t models have fewer parameters.

[Figure 2.3](#) plots the estimated correlation series for four selected pairs of asset indices, namely, SPX–SBW, SPX–EMU, SPX–OIL, and SBW–GOLD. We compare the results for our Meta-t-log-corr model and the best performing multivariate t model, i.e. the t-GAS-hl. While the overall level of estimated correlations is about the same for both models, the Meta-t-log-corr model is more volatile and less persistent over time. This is consistent with the fact that the

³To save space, we only report the K , Φ and $\bar{\eta}$ parameters and leave out the ω vectors.

Meta-t-log-corr model has a slightly lower estimated autoregressive parameter than the t-GAS-hl model.

2.4.3.2 Out-of-sample performance

To verify the short-term forecasting ability of the dynamic log correlation model, we compare the performance of different approaches in a Global Minimum Variance (GMV) portfolio allocation exercise. To this end, we forecast the 1-day ahead covariance matrices using the better performing models, namely the group of Meta-t models. Our out-of-sample exercise is conducted over a period ranging from the start of 2019 to the end of 2020. Market conditions are rather stable for the 2019 period and we thus only re-estimate the models twice over the year. The year 2020, on the other hand, featured extremely volatile market conditions which culminated at the onset of the COVID-19 crisis in March. We address the potential impact of these exceptional market conditions on our out-of-sample results by re-estimating the models every two months to incorporate new information on these rapidly evolving financial conditions. The 1-day ahead volatility is forecast using the Beta-t-EGARCH model and the correlation matrices are forecast respectively from the Meta-t-log-corr model, the Meta-t-cDCC model, the Meta-t-GAS model, and the Meta-t-GAS-h model. Based on the daily forecasts, we construct the conditional covariance matrix and use it as an input to compute the optimal GMV portfolio allocation for each day.

We present in [Table 2.3](#) the out-of-sample performance results in terms of the annualized portfolio standard deviation. We observe that the Meta-t-log-corr model has similar portfolio volatility to the Meta-t-GAS and Meta-t-GAS-h models. Furthermore, all three score-driven models exhibit better performances than the Meta-t-cDCC model. This result highlights the fact that score driven models are more robust to outliers and, as a consequence, exhibit superior performance under extreme market conditions. Furthermore, we can draw similar conclusions for model comparisons in terms of the Sharpe ratio of the considered portfolios. More specifically, the Sharpe ratios of the dynamic conditional score models are close to each other and higher than the corresponding Sharpe ratio for the cDCC model specification.

2.5 Conclusions

The proposed score-driven correlation model ensures positivity of the conditional correlation matrix while offering high flexibility in the modelling of the conditional distribution and dynamics. In the empirical application we have shown that the model fits well a set of financial and alternative assets, passing various diagnostic tests for the distribution and dynamics. We also show that our model performs well compared to alternative correlation parameterizations in terms of in-sample fit. Our out-of-sample analysis highlights the superior performance of score driven models compared to the cDCC model in terms of correlation forecasting applied to minimum variance portfolio selection.

Many generalizations are possible, for example using generalized student-t distributions as in [Harvey and Lange \(2017\)](#), and distributions allowing for skewness as in [Bauwens and Laurent \(2005\)](#) or [Harvey and Sucarrat \(2014\)](#). Rather than models based on the student-t distribution, one may also generalize the Gamma-GED-EGARCH to the multivariate case. Future work may also address specification tests in high dimensions, for example clustering of parameters characterizing the dynamics of the log correlation matrix. We have discussed two extreme cases in the empirical example, the unrestricted case and the one setting all parameters equal, but developing strategies to specify intermediate cases such as the factor structure used in [Archakov et al. \(2020\)](#) would seem reasonable as well.

2.A Theoretical material

2.A.1 Lyapounov exponent of the SRE in the bivariate case

In the bivariate case, the SRE is given by Equation (2.13) where $\psi_t(s) = \omega + \phi s + \kappa g(s)Z_t$. Since $g(s)$ is continuously differentiable, we are looking for

$$\Lambda(\psi) = \sup_{s \in \mathbb{R}} |\partial \psi(s) / \partial s|.$$

Straightforward calculations show that $g(s) = \frac{1}{2}(-\rho, 1, 1, -\rho)^\top$, with $\rho = \frac{\exp(2s)-1}{\exp(2s)+1}$, and $Z_t = \text{vec}(w_t \xi_t \xi_t^\top - I_N)$, with $w_t = \frac{\eta + N}{\eta - 2 + \xi_t^\top \xi_t}$. Hence,

$$\frac{\partial \psi(s)}{\partial s} = \phi - \kappa \frac{\exp(2s)}{(\exp(2s) + 1)^2} \left\{ w_t (\xi_{1t}^2 + \xi_{2t}^2) - 2 \right\}$$

and because $\inf_{s \in \mathbb{R}} \frac{\exp(2s)}{(\exp(2s)+1)^2} = 0$ and $\sup_{s \in \mathbb{R}} \frac{\exp(2s)}{(\exp(2s)+1)^2} = 1$, we have

$$\Lambda(\psi) = \max \left(\phi, \phi - \kappa \left[w_t (\xi_{1t}^2 + \xi_{2t}^2) - 2 \right] \right).$$

Using this expression, we can simulate a large number of i.i.d. ξ_t and estimate the Lyapounov exponent $\mathbb{E}[\log \Lambda(\psi)]$ by the sample mean of $\log \Lambda(\psi)$. For a student-t with 5 and 15 degrees of freedom, Figure 2.4 shows the parameter space of (ϕ, κ) for which the Lyapounov exponent is negative. As ϕ is approaching 1, κ is constrained to a shrinking region around zero.

For the general case $N \geq 2$ and $\theta_t \in \mathcal{T}_{N,\epsilon} \subset \mathbb{R}^{N(N-1)/2}$, no simple analytical expression is available for $\Lambda(\psi)$, but Appendix 2.D shows that there exists an upper bound. Furthermore, we can obtain numerical approximations to estimate the Lyapounov exponent for a given set of parameters, see Appendix 2.D for details.

2.A.2 Proof of Theorem 1

First, we have

$$\begin{aligned}
dL(\delta) &= -\frac{1}{2} \sum_{t=1}^T \left\{ d \log |\Theta_t| - \frac{\eta + N}{\eta - 2 + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t} \text{Tr} \left(\varepsilon_t \varepsilon_t^\top d\Theta_t^{-1} \right) \right\} \\
&= -\frac{1}{2} \sum_{t=1}^T \text{Tr} \left\{ \Theta_t^{-1} d\Theta_t - \frac{\eta + N}{\eta - 2 + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t} \varepsilon_t \varepsilon_t^\top \Theta_t^{-1} d\Theta_t \Theta_t^{-1} \right\} \\
&= \frac{1}{2} \sum_{t=1}^T \text{Tr} \left\{ \left(\frac{\eta + N}{\eta - 2 + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t} \Theta_t^{-1} \varepsilon_t \varepsilon_t^\top \Theta_t^{-1} - \Theta_t^{-1} \right) d\Theta_t \right\} \\
&= \frac{1}{2} \sum_{t=1}^T \text{vec} \left(\frac{\eta + N}{\eta - 2 + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t} \Theta_t^{-1} \varepsilon_t \varepsilon_t^\top \Theta_t^{-1} - \Theta_t^{-1} \right)^\top d\text{vec} \Theta_t
\end{aligned}$$

Now, using a result of Linton and McCrorie (1995) for the differential of the matrix exponential, we can write

$$d\text{vec} \Theta_t = A_t d\text{vec} \log \Theta_t,$$

and with the definition of the permutation matrix P ,

$$d\text{vec} \Theta_t = A_t P \begin{pmatrix} d\text{diag}(\log \Theta_t) \\ d\text{vecl}(\log \Theta_t) \end{pmatrix}$$

Now, $\text{vecl}(\log \Theta_t) = \theta_t$, and by Archakov and Hansen (2021), $\text{diag}(\log \Theta_t)$ is a unique function of θ_t , for which no analytical form exists for $N > 4$. However, using the proof of Proposition 3 of Archakov and Hansen (2021), we have an analytical expression for the differential of this function, given by

$$d\text{diag}(\log \Theta_t) = -(E_d A_t E_d^\top)^{-1} E_d A_t E_s^\top d\theta_t$$

which completes the stated result.

The expressions given in Remark 1 for the Gaussian distribution, i.e. for the limiting case $\eta \rightarrow \infty$, can be obtained easily. First, using Stirling's approximation, note that

$$\begin{aligned}
&\lim_{\eta \rightarrow \infty} \left[T \log \left(\Gamma \left(\frac{\eta + N}{2} \right) \right) - T \log \left(\Gamma \left(\frac{\eta}{2} \right) \right) - \frac{TN}{2} \log((\eta - 2)\pi) \right] \\
&= \lim_{\eta \rightarrow \infty} T \log \frac{\Gamma \left(\frac{\eta + N}{2} \right)}{\Gamma \left(\frac{\eta}{2} \right) ((\eta - 2)\pi)^{N/2}} \\
&= T \log (2\pi)^{-N/2} = -\frac{TN}{2} \log(2\pi)
\end{aligned}$$

And for the score function, it can be easily shown that

$$\begin{aligned}
& p \lim_{\eta \rightarrow \infty} -\frac{1}{2} \sum_{t=1}^T \left\{ \log |\Theta_t| + (\eta + N) \log \left(1 + \frac{1}{\eta - 2} \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t \right) \right\} \\
&= p \lim_{\eta \rightarrow \infty} -\frac{1}{2} \sum_{t=1}^T \left\{ \log |\Theta_t| + \frac{\eta + N}{\eta - 2} \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t \right\} \\
&= -\frac{1}{2} \sum_{t=1}^T \left\{ \log |\Theta_t| + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t \right\}
\end{aligned}$$

Thus, we arrive at the expression of the log likelihood function for the multivariate normal distribution. Furthermore, the score can be obtained directly by taking the probability limit of $\frac{\eta + N}{\eta - 2 + \varepsilon_t^\top \Theta_t^{-1} \varepsilon_t}$ as $\eta \rightarrow \infty$, which is equal to 1.

2.A.3 Proof of Theorem 2

We first need to show that, under our conditions, θ_t and its derivatives up to order 3, evaluated at the true parameters, are stationary and ergodic processes. The process θ_t satisfies the stochastic recurrence equation (SRE) in Equation (2.13) which by Theorem 2.8 of [Straumann and Mikosch \(2006\)](#) is stationary and ergodic under the contraction condition of Assumption (A6). The compactness of Ω then implies consistency by Lemma 1 of [Jensen and Rahbek \(2004\)](#).

The process $\partial\theta_t/\partial\delta$ permits an SRE as in Equation (2.17). It is stationary and ergodic if $\mathbb{E}[\log |x_t|] < 0$, see [Straumann and Mikosch \(2006\)](#). This is implied by Assumption (A5), i.e. $\mathbb{E}[x_t^2] < 1$, using Jensen's inequality.

The process $\frac{\partial^2 \theta_t}{\partial \delta \partial \delta^\top}$ also permits an SRE:

$$\frac{\partial^2 \theta_t}{\partial \delta \partial \delta^\top} = C_{t-1} + x_{t-1} \frac{\partial^2 \theta_{t-1}}{\partial \delta \partial \delta^\top}$$

where $C_t = \frac{\partial z_t}{\partial \delta^\top} + \frac{\partial \theta_t}{\partial \delta} \frac{\partial x_t}{\partial \delta^\top}$. Again, stationary and ergodicity are implied by Assumption (A5). The same argument applies to the third derivatives of θ_t .

We then obtain asymptotic normality for the score evaluated at the true parameter, invoking the central limit theorem for martingale difference sequences of [Brown \(1971\)](#), i.e.

$$T^{-1/2} \sum_{t=1}^T \frac{\partial l_t(\vartheta_0)}{\partial \vartheta} \rightarrow_d N(0, I(\vartheta_0)).$$

This requires that $\mathbb{E}[(\partial\theta_{it}/\partial\vartheta_j)^2] < \infty$ for $i = 1, \dots, N^*$ and $j = 1, \dots, \dim(\vartheta)$, which is implied by Assumption (A5), see Theorem 1 of [Harvey \(2013\)](#).

Moreover, we have

$$-\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t(\vartheta_0)}{\partial \vartheta \partial \vartheta^\top} \rightarrow_p I(\vartheta_0)$$

as in Proposition 5 of [Harvey \(2013\)](#). It suffices to check that, for a neighborhood $N(\vartheta_0)$ of the true parameter,

$$\mathbb{E} \sup_{\vartheta \in N(\vartheta_0)} \left\| \frac{\partial \text{vec}}{\partial \vartheta'} \frac{\partial^2 l_t(\vartheta)}{\partial \vartheta \partial \vartheta'} \right\| < \infty. \quad (2.21)$$

For the volatility part this has been shown before, see e.g. Theorem 3 of [Harvey \(2013\)](#), so we concentrate on the correlation part. We have

$$\begin{aligned} \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial^2 l_t(\delta)}{\partial \delta \partial \delta^\top} &= (I_m \otimes u_t^\top \otimes I_m) \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial \theta_t^\top}{\partial \delta} + \left(\frac{\partial \text{vec}^\top}{\partial \delta} \frac{\partial \theta_t^\top}{\partial \delta} \otimes I_m \right) (I_{N^*} \otimes \text{vec}(I_m)) \frac{\partial u_t}{\partial \delta^\top} \\ &+ \left(\frac{\partial u_t^\top}{\partial \delta} \otimes I_m \right) \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial \theta_t^\top}{\partial \delta} + (I_m \otimes \frac{\partial \theta_t^\top}{\partial \delta}) \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial u_t}{\partial \delta^\top} \end{aligned} \quad (2.22)$$

with $m = \dim(\delta)$. The first term on the right hand side of Equation (2.22) depends on the third order derivatives of the correlation parameter with respect to δ . A typical element of this matrix is given by

$$\begin{aligned} \frac{\partial^3 \theta_t(\delta)}{\partial \delta_i \partial \delta_j \partial \delta_k} &= \frac{\partial^2 z_{i,t-1}}{\partial \theta_j \partial \theta_k} + \frac{\partial x_{t-1}}{\partial \delta_j} \frac{\partial^2 \theta_{t-1}}{\partial \delta_i \partial \delta_k} + \frac{\partial^2 x_{t-1}}{\partial \delta_j \partial \delta_k} \frac{\partial \theta_{t-1}}{\partial \delta_i} + \frac{\partial x_{t-1}}{\partial \delta_k} \frac{\partial^2 \theta_{t-1}}{\partial \delta_i \partial \delta_j} + x_{t-1} \frac{\partial^3 \tilde{\theta}_{t-1}(\delta)}{\partial \delta_i \partial \delta_j \partial \delta_k} \\ &= \sum_{j=1}^{\infty} \zeta_{t-j} \prod_{i=1}^{j-1} x_{t-i} \end{aligned} \quad (2.23)$$

where z_{it} denotes the i -th column of z_t , and $\zeta_t = \frac{\partial^2 z_{i,t-1}}{\partial \theta_j \partial \theta_k} + \frac{\partial x_{t-1}}{\partial \delta_j} \frac{\partial^2 \theta_{t-1}}{\partial \delta_i \partial \delta_k} + \frac{\partial^2 x_{t-1}}{\partial \delta_j \partial \delta_k} \frac{\partial \theta_{t-1}}{\partial \delta_i} + \frac{\partial x_{t-1}}{\partial \delta_k} \frac{\partial^2 \theta_{t-1}}{\partial \delta_i \partial \delta_j}$. Each term of ζ_t can be uniformly bounded in a neighborhood of δ_0 and a repeated application of the c_r inequality yields $\mathbb{E} \sup_{\delta \in N(\delta_0)} |\zeta_t| < \infty$. Note that $|x_t|$ is a stationary ergodic sequence of nonnegative random variables with $\mathbb{E} \log |x_t| < 0$, which is implied by Assumption (A5). Thus, by Lemma 2.4 of Straumann and Mikosch (2006), $\prod_{i=1}^{t-1} |x_{t-i}| \rightarrow_{e.a.s.} 0$, meaning that there exists a $\gamma > 1$ s.t. $\gamma^t \prod_{i=1}^{t-1} |x_{t-i}| \rightarrow_{a.s.} 0$. This implies that $\mathbb{E} \prod_{i=1}^{t-1} |x_{t-i}(\bar{\delta}_{t-i})| = O(\rho^t)$, which in turn implies that the infinite sum in Equation (2.23) is convergent and we obtain

$$\mathbb{E} \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^3 \theta_t(\delta)}{\partial \delta_i \partial \delta_j \partial \delta_k} \right| < \infty.$$

The remaining terms on the right hand side of Equation (2.22) can be bounded in a similar way. Together with corresponding results for the volatility part, this implies Equation (2.21).

Finally, tedious but straightforward calculations yield the expression of the information matrix using results of [Fiorentini et al. \(2003\)](#). The details are omitted here but available upon request.

2.A.4 Proof of Theorem 3

The proof of consistency and asymptotic normality is essentially the same as that of Theorem 2, noting that stationarity and ergodicity of the process $\tilde{\theta}_t$ hold under the same conditions as for the process θ_t . The form of the asymptotic covariance matrix of the inference function for margins approach is based on Joe (2005), and the components of $\mathcal{I}$ follow by straightforward calculations, which are again omitted to economize on space, but available upon request.

2.B Tables and figures

Table 2.1: Diagnostics and model selection criteria for multivariate models.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
k	46	18	32	31	4	3
T	845	845	845	845	845	845
score test	0.5014	0.3598	0.3228	0.3910	0.4150	0.3134
ξ_t	0.0653	0.0489	0.0818	0.0810	0.0513	0.0461
$vech(\xi_t \xi_t^\top)$	0.9457	0.9463	0.9526	0.9723	0.9585	0.9539
$\ln \mathcal{L}(\hat{\theta})$	446.59	421.01	432.77	424.91	368.12	361.92
LR		51.17	27.65	43.37	156.95	169.35
d.o.f.		28	14	15	42	43
p-value		0.0048	0.0158	0.0001	0.0000	0.0000
BIC	-583.18	-720.70	-649.88	-640.90	-709.28	-703.62
AIC	-801.19	-806.01	-801.54	-787.82	-728.23	-717.83

Note: k is the number of parameters in the model. T is the number of data points in the sample. Score test, ξ_t and $vech(\xi_t \xi_t^\top)$ are the p-values of a multivariate Portmanteau test applied to the correlation score series, the correlation residual series and the outer product of the correlation residual series. The number of lags is 30 and degrees of freedom are $N_{\text{lag}} \times N^{*2}$ with $N_{\text{lag}} = 30$ and $N^* = 15$. The likelihood-ratio statistic is computed as $LR = -2 \ln \frac{\mathcal{L}_i(\hat{\theta})}{\mathcal{L}_1(\hat{\theta})}$, with $\mathcal{L}_i$ corresponding to the likelihood for model i. The number of degrees of freedom for the LR test is computed as the difference in the number of parameters between Model 1 and Model i. The Akaike and Bayesian information criteria are respectively computed as $AIC = -2 \ln \mathcal{L}(\hat{\theta}) + 2k$, and $BIC = -2 \ln \mathcal{L}(\hat{\theta}) + k \ln T$.

Table 2.2: Modelling comparison for dynamic correlation matrix.

	Meta-t-log-corr	Meta-t-cDCC	Meta-t-GAS	Meta-t-GAS-h
k	42	42	42	42
$\ln \mathcal{L}(\hat{\theta})$	16330.63	16329.34	16327.67	16329.73
AIC	-32577.26	-32574.69	-32571.33	-32575.47
BIC	-32378.21	-32375.64	-32372.28	-32376.41
	t-cDCC	t-GAS	t-GAS-l	t-GAS-hl
k	36	36	36	36
$\ln \mathcal{L}(\hat{\theta})$	16244.69	16230.82	16231.58	16247.45
AIC	-32417.39	-32389.63	-32391.15	-32422.89
BIC	-32246.77	-32219.02	-32220.54	-32252.28

Note: k is the number of parameters. The Akaike and Bayesian information criteria are respectively computed as $AIC = -2 \ln \mathcal{L}(\hat{\theta}) + 2k$, and $BIC = -2 \ln \mathcal{L}(\hat{\theta}) + k \ln T$. The first four models are estimated following a two-step estimation procedure with dynamic volatility estimated using the Beta-t-EGARCH model and the dynamic correlation matrix modelled respectively by the Meta-t-log-corr, Meta-t-cDCC, Meta-t-GAS, and Meta-t-GAS-h models. The last four models are estimated following a one-step estimation procedure with dynamic volatility and correlation matrix modelled respectively by t-cDCC, t-GAS, t-GAS-l and t-GAS-hl.

Table 2.3: Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio.

Meta-t-log-corr	Meta-t-cDCC	Meta-t-GAS	Meta-t-GAS-h
Annualized standard deviation			
5.8203	6.0859	5.8416	5.8112
Sharpe ratio			
0.9403	0.6730	0.9591	0.9409

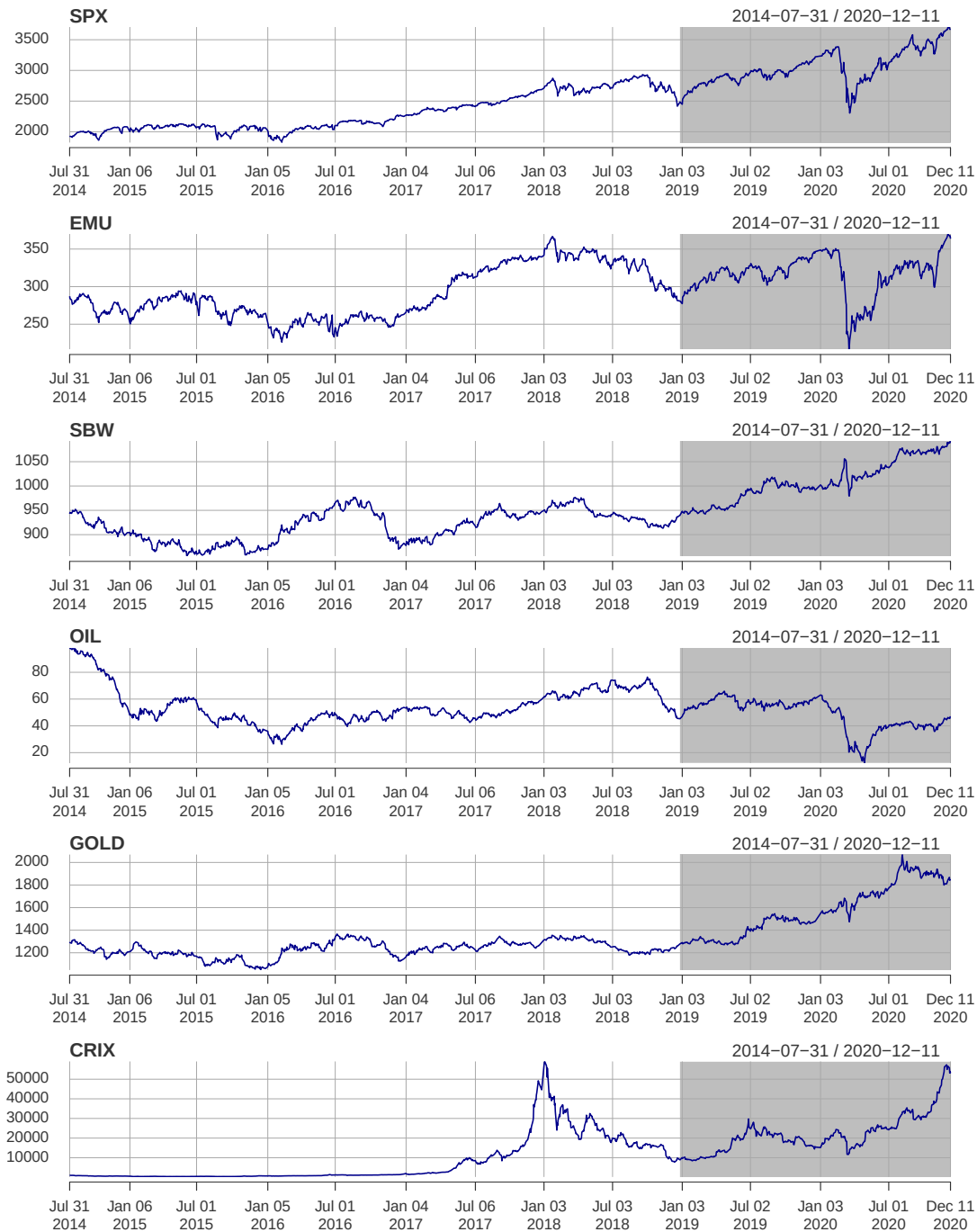


Figure 2.2: Time series of daily asset prices, August 1, 2014 to December 12, 2020, for (from top to bottom) SPX, SBW, EMU, OIL, GOLD, and CRIX. The grey shaded area is the out-of-sample period ranging from the beginning of 2019 to the end of 2020.

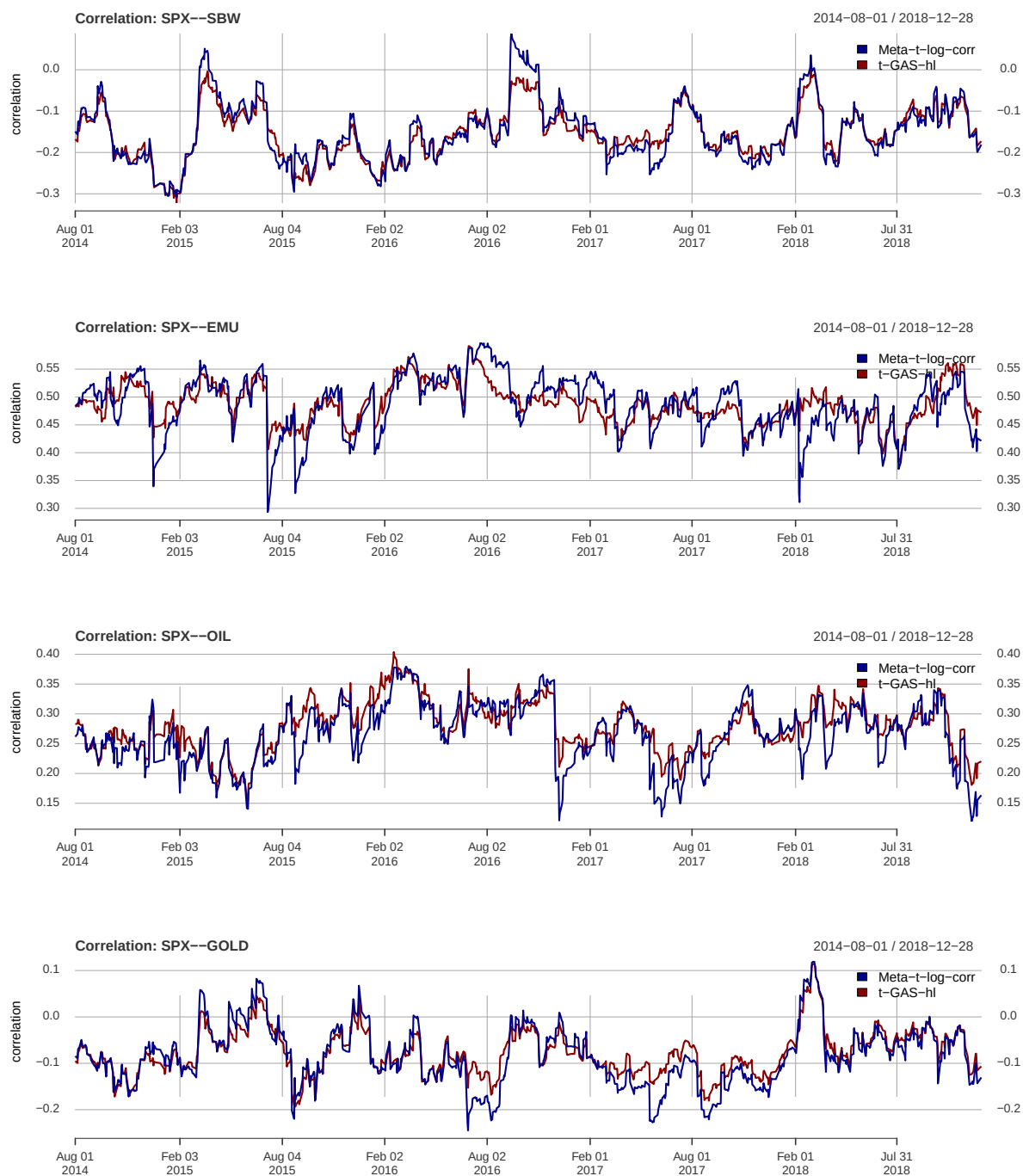


Figure 2.3: Estimated conditional correlation series for four selected pairs (from top to bottom): SPX-SBW, SPX-EMU, SPX-OIL, SPX-GOLD, using two models: the Meta-t-log-corr model and the t-GAS-hl model.

2.C Example: the bivariate Gaussian case

This section shows that our model reduces to that of [Harvey \(2013\)](#) (Section 7.3.1) in the bivariate Gaussian case. It also serves to illustrate the structure of the involved matrix expressions. Let $N = 2$. For Θ and $\log \Theta$, see Example 1 above. We have $A_t = (\Gamma_t \otimes \Gamma_t) \Xi_t (\Gamma_t \otimes \Gamma_t)^\top$ with $\text{diag}(\Xi_t) = (1 + \rho, 2\rho/\log \frac{1+\rho}{1-\rho}, 2\rho/\log \frac{1+\rho}{1-\rho}, 1 - \rho)$ and

$$\Gamma_t \otimes \Gamma_t = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

which gives

$$A_t = \frac{1}{2} \begin{pmatrix} 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho & \rho & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \\ \rho & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho \\ \rho & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho \\ 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho & \rho & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \end{pmatrix}$$

The elimination matrices are $E_l = (0, 1, 0, 0)$, $E_u = (0, 0, 1, 0)$, and

$$E_d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

such that

$$\begin{aligned} E_d A_t E_d^\top &= \frac{1}{2} \begin{pmatrix} 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \\ 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \end{pmatrix}, \quad E_d A_t E_s^\top = \begin{pmatrix} \rho \\ \rho \end{pmatrix} \\ (E_d A_t E_d^\top)^{-1} &= \frac{\log \frac{1+\rho}{1-\rho}}{4\rho} \begin{pmatrix} 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} - 1 \\ \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} - 1 & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \end{pmatrix} \\ -(E_d A_t E_d^\top)^{-1} E_d A_t E_s^\top &= -\frac{\log \frac{1+\rho}{1-\rho}}{4\rho} \begin{pmatrix} 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} - 1 \\ \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} - 1 & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \end{pmatrix} \begin{pmatrix} \rho \\ \rho \end{pmatrix} = \begin{pmatrix} -\rho \\ -\rho \end{pmatrix} \end{aligned}$$

Note that this can be obtained directly from Example 1, where the diagonal elements of $\log \Theta$ are $\frac{1}{2} \log(1 - \rho^2)$, and the off-diagonal element is $\gamma := \frac{1}{2} \log \frac{1+\rho}{1-\rho}$, or $\rho = \frac{\exp(2\gamma)-1}{\exp(2\gamma)+1}$. Hence,

$$\frac{\partial}{\partial \rho} \frac{1}{2} \log(1 - \rho^2) \frac{\partial \rho}{\partial \gamma} = \frac{-\rho}{1 - \rho^2} \frac{4 \exp(2\gamma)}{(\exp(2\gamma) + 1)^2} = \frac{-2\rho}{1 - \rho^2} (1 - \rho^2) = -2\rho.$$

Together with the factor $1/2$ in u_t , we obtain the same result.

The permutation matrix P is given by

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Then,

$$\begin{pmatrix} -(E_d A_t E_d^\top)^{-1} E_d A_t E_s^\top \\ I_{N(N-1)/2} \end{pmatrix}^\top P^\top = \begin{pmatrix} -\rho & -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix}$$

and

$$\begin{aligned} \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix} A_t &= \frac{1}{2} \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix} \begin{pmatrix} 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho & \rho & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \\ \rho & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho \\ \rho & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho \\ 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho & \rho & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 - \rho^2 & 1 - \rho^2 & 0 \end{pmatrix} \end{aligned}$$

Finally,

$$\text{vec}(\Theta^{-1} \varepsilon_t \varepsilon_t^\top \Theta^{-1}) = \frac{1}{(1 - \rho^2)^2} \begin{pmatrix} \varepsilon_{1t}^2 - 2\rho \varepsilon_{1t} \varepsilon_{2t} + \rho^2 \varepsilon_{2t}^2 \\ (1 + \rho^2) \varepsilon_{1t} \varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \\ (1 + \rho^2) \varepsilon_{1t} \varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \\ \rho^2 \varepsilon_{1t}^2 - 2\rho \varepsilon_{1t} \varepsilon_{2t} + \varepsilon_{2t}^2 \end{pmatrix}, \quad \text{vec}(\Theta^{-1}) = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 \\ -\rho \\ -\rho \\ 1 \end{pmatrix}$$

such that

$$\begin{pmatrix} 0 & 1 - \rho^2 & 1 - \rho^2 & 0 \end{pmatrix} \text{vec}(\Theta^{-1} \varepsilon_t \varepsilon_t^\top \Theta^{-1} - \Theta^{-1}) = \frac{2}{1 - \rho^2} \left\{ (1 + \rho^2) \varepsilon_{1t} \varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \right\} + 2\rho$$

and

$$u_t = \frac{1}{1 - \rho^2} \left\{ (1 + \rho^2) \varepsilon_{1t} \varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \right\} + \rho$$

which is the score given by Harvey (2013, Section 7.3.1) for the bivariate Gaussian case. He uses the parametrization $\gamma = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$ and

$$g = \frac{\exp \gamma + \exp(-\gamma)}{2}, \quad \bar{g} = \frac{\exp \gamma - \exp(-\gamma)}{2}$$

so that

$$u_t = (2g^2 - 1)\varepsilon_{1t}\varepsilon_{2t} - g\bar{g}(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) + \frac{\bar{g}}{g}.$$

For the information matrix of Theorem 2, the expressions reduce to

$$\begin{aligned} I_{\delta\delta} &= \frac{1}{2}\mathbb{E}\left[\frac{\partial\theta_t^\top}{\partial\delta}M_t\left(\Theta_t^{-1}\otimes\Theta_t^{-1}\right)M_t^\top\frac{\partial\theta_t}{\partial\delta^\top}\right] \\ I_{\tau\tau} &= \frac{1}{2}\mathbb{E}\left[\frac{\partial\text{vec}(H_t)^\top}{\partial\tau}\left(H_t^{-1}\otimes H_t^{-1}\right)\frac{\partial\text{vec}(H_t)}{\partial\tau^\top}\right] \\ I_{\tau\delta} &= \frac{1}{2}\mathbb{E}\left[\frac{\partial\text{vec}(H_t)^\top}{\partial\tau}\left(H_t^{-1}\otimes H_t^{-1}\right)(D_t\otimes D_t)M_t^\top\frac{\partial\theta_t}{\partial\delta^\top}\right] \end{aligned}$$

With $M_t = (1-\rho^2)\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}$, straightforward calculations show that $\frac{1}{2}M_t\left(\Theta_t^{-1}\otimes\Theta_t^{-1}\right)M_t^\top = (1+\rho^2) = 2 - 1/g^2$, and

$$I_{\delta\delta} = \mathbb{E}\left[(2 - 1/g^2)\frac{\partial\theta_t^\top}{\partial\delta}\frac{\partial\theta_t}{\partial\delta^\top}\right]$$

Furthermore,

$$H_t = \begin{pmatrix} \exp(2h_1) & \rho\exp(h_1+h_2) \\ \rho\exp(h_1+h_2) & \exp(2h_2) \end{pmatrix}, \quad H_t^{-1} = \frac{1}{(1-\rho^2)}\begin{pmatrix} \exp(-2h_1) & -\rho\exp(-h_1-h_2) \\ -\rho\exp(-h_1-h_2) & \exp(-2h_2) \end{pmatrix}$$

and

$$\frac{\partial\text{vec}(H_t)}{\partial\tau^\top} = \frac{\partial\text{vec}(H_t)}{\partial\varpi^\top}\frac{\partial\varpi}{\partial\tau^\top}, \quad \frac{\partial\text{vec}(H_t)}{\partial\varpi^\top} = \begin{pmatrix} 2\exp(2h_1) & 0 & 0 \\ \rho\exp(h_1+h_2) & \rho\exp(h_1+h_2) & \exp(h_1+h_2) \\ 0 & 2\exp(2h_2) & 0 \end{pmatrix}$$

where $\varpi := (h_1, h_2, \rho)^\top$. Straightforward calculations show that

$$\frac{1}{2}\frac{\partial\text{vec}(H_t)^\top}{\partial\varpi}\left(H_t^{-1}\otimes H_t^{-1}\right)\frac{\partial\text{vec}(H_t)}{\partial\varpi^\top} = \begin{pmatrix} 1+g^2 & 1-g^2 & -\bar{g}/g \\ 1-g^2 & 1+g^2 & -\bar{g}/g \\ -\bar{g}/g & -\bar{g}/g & 2-1/g^2 \end{pmatrix} =: J$$

which is the matrix given in Proposition 40 of Harvey (2013). Thus,

$$I_{\tau\tau} = \mathbb{E}\left[\frac{\partial\varpi^\top}{\partial\tau}J\frac{\partial\varpi}{\partial\tau^\top}\right]$$

Similarly, we obtain

$$I_{\tau\delta} = \mathbb{E}\left[\frac{\partial\varpi^\top}{\partial\tau}J_3\frac{\partial\theta}{\partial\delta^\top}\right]$$

where J_3 is the third column of J .

2.D Lyapounov exponent of the SRE

In the bivariate case, the SRE is given by Equation (2.13) where $\psi_t(s) = \omega + \phi s + \kappa g(s)Z_t$. Since $g(s)$ is continuously differentiable, we are looking for

$$\Lambda(\psi) = \sup_{s \in \mathbb{R}} |\partial \psi(s) / \partial s|.$$

Straightforward calculations show that $g(s) = \frac{1}{2}(-\rho, 1, 1, -\rho)^\top$, with $\rho = \frac{\exp(2s)-1}{\exp(2s)+1}$, and $Z_t = \text{vec}(w_t \xi_t \xi_t^\top - I_N)$, with $w_t = \frac{\eta + N}{\eta - 2 + \xi_t^\top \xi_t}$. Hence,

$$\frac{\partial \psi(s)}{\partial s} = \phi - \kappa \frac{\exp(2s)}{(\exp(2s) + 1)^2} \left\{ w_t (\xi_{1t}^2 + \xi_{2t}^2) - 2 \right\}$$

and because $\inf_{s \in \mathbb{R}} \frac{\exp(2s)}{(\exp(2s)+1)^2} = 0$ and $\sup_{s \in \mathbb{R}} \frac{\exp(2s)}{(\exp(2s)+1)^2} = 1$, we have

$$\Lambda(\psi) = \max \left(\phi, \phi - \kappa \left[w_t (\xi_{1t}^2 + \xi_{2t}^2) - 2 \right] \right).$$

Using this expression, we can simulate a large number of i.i.d. ξ_t and estimate the Lyapounov exponent $\mathbb{E}[\log \Lambda(\psi)]$ by the sample mean of $\log \Lambda(\psi)$. For a student-t with 5 and 15 degrees of freedom, Figure 2.4 shows the parameter space of (ϕ, κ) for which the Lyapounov exponent is negative. As ϕ is approaching 1, κ is constrained to a shrinking region around zero.

In the general case $N \geq 2$ and $\theta_t \in \mathcal{T}_{N,\epsilon} \subset \mathbb{R}^{N(N-1)/2}$, we denote the spectral (operator) matrix norm by $\|\cdot\|$, and the L_2 norm by $\|\cdot\|_2$. At several occasions we will use that

$$\|\Theta_t\| \leq \|\Theta_t\|_2 \leq N, \quad a.s. \quad (2.24)$$

because Θ_t is a correlation matrix and, hence, $\sup_{i,j} |\Theta_{t,ij}| = 1$.

For the differential of the impact function $g(\cdot)$ we have

$$dg(\theta_t) = \frac{1}{2} \left\{ (\Theta_t^{-1/2} \otimes \Theta_t^{-1/2}) dM_t + M_t d(\Theta_t^{-1/2} \otimes \Theta_t^{-1/2}) \right\} \quad (2.25)$$

Let $Y_t = \Theta_t \otimes \Theta_t$, and Then,

$$\begin{aligned} \|dY_t^{-1/2}\| &\leq \|d\text{vec} Y_t^{-1/2}\|_2 \\ &= \|(Y_t^{-1/2} \otimes Y_t^{-1/2}) d\text{vec} Y_t^{1/2}\|_2 \\ &= \|(Y_t^{-1/2} \otimes Y_t^{-1/2})(Y_t^{1/2} \otimes I_N + I_N \otimes Y_t^{1/2})^{-1} d\text{vec} Y_t\|_2 \\ &= \|(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1} d\text{vec} Y_t\|_2 \\ &\leq \|(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1}\| \|d\text{vec} Y_t\|_2 \end{aligned} \quad (2.26)$$

where in Equation (2.26) we used that the spectral norm $\|\cdot\|$ is induced by the L_2 vector norm

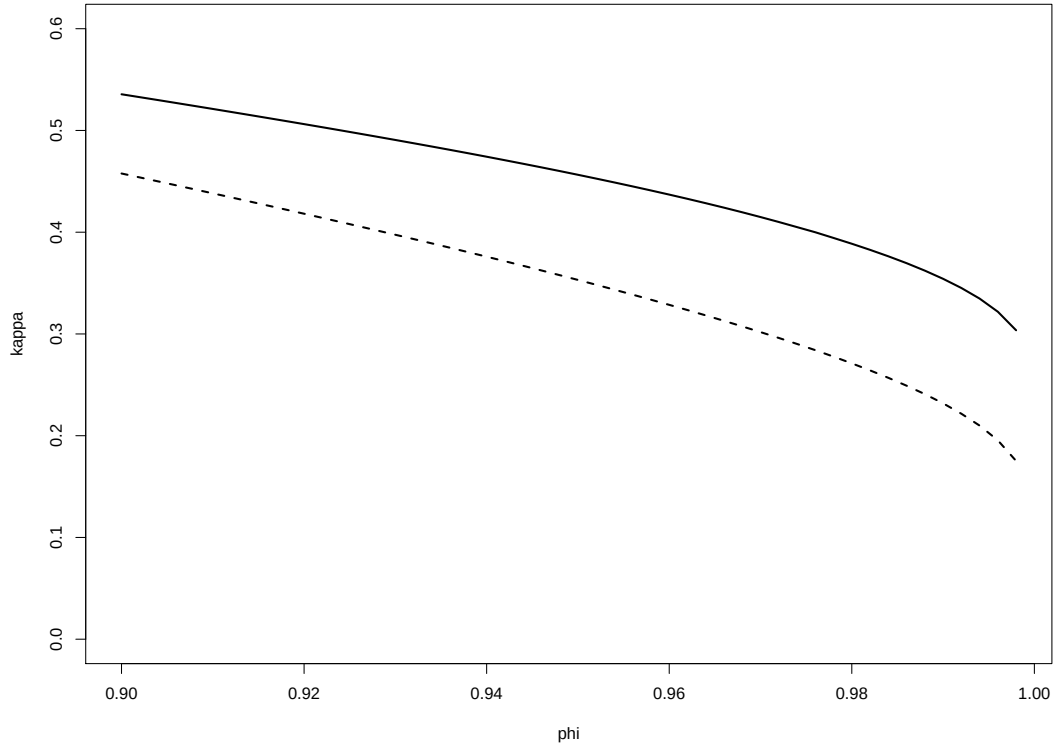


Figure 2.4: The lines show the combinations of (ϕ, κ) for which $\mathbb{E}[\log \Lambda(\psi)] = 0$. Solid line: student-t with 5 df, dashed: 15 df. South of these functions is the region for which the Lyapounov exponent is negative.

$\|\cdot\|_2$, see e.g. [Magnus and Neudecker \(2019\)](#).

Now,

$$\begin{aligned}
 \|(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1}\| &\leq \lambda_{\max}(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1} \\
 &= \lambda_{\min}^{-1}(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t) \\
 &\leq \lambda_{\min}^{-1}(Y_t \otimes Y_t^{1/2}) \\
 &\leq \epsilon^{-3} < \infty, \quad a.s.
 \end{aligned} \tag{2.27}$$

by Assumption (A6), and

$$\begin{aligned}
 d\text{vec}(Y_t) &= (I_N \otimes C_{NN} \otimes I_N)(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})d\text{vec}(\Theta_t) \\
 &= (I_N \otimes C_{NN} \otimes I_N)(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})M_t^\top d\theta_t
 \end{aligned} \tag{2.28}$$

with commutation matrix C_{NN} . We have

$$\begin{aligned}
\|d\text{vec}(Y_t)\|_2 &\leq \|(I_N \otimes C_{NN} \otimes I_N)(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})M_t\| \|d\theta_t\|_2 \\
&\leq \|(I_N \otimes C_{NN} \otimes I_N)\| \|(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})\| \|M_t\| \|d\theta_t\|_2 \\
&= \|(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})\| \|M_t\| \|d\theta_t\|_2
\end{aligned}$$

where the first inequality follows by the fact that the spectral norm is induced by the L_2 vector norm, the second inequality by submultiplicativity of the spectral norm, and the last equality by the fact that $\lambda_{\max}(C_{NN}) = 1$. Now,

$$\begin{aligned}
\|(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})\| &\leq 2\|I_{N^2} \otimes \text{vec}(\Theta_t)\| \\
&= 2\lambda_{\max}^{1/2} \{(I_{N^2} \otimes \text{vec}(\Theta_t))^\top (I_{N^2} \otimes \text{vec}(\Theta_t))\} \\
&= 2\lambda_{\max}^{1/2} \{I_{N^2} \text{vec}(\Theta_t)^\top \text{vec}(\Theta_t)\} \\
&= 2\{\text{vec}(\Theta_t)^\top \text{vec}(\Theta_t)\}^{1/2} \leq 2N, \quad a.s.
\end{aligned}$$

where the last inequality follows by Equation (2.24). Furthermore, note that M_t can be written as $M_t = M_{1t} + M_{2t}$, where $M_{2t} = P_l^\top A_t$, $M_{1t} = -E_s A_t E_d^\top (E_d A_t E_d^\top)^{-1} P_d^\top A_t$, and P_d and P_l are defined by $P_d \text{diag}(Q) = \text{vec}(Q)$ and $P_l \text{vecl}(Q) = \text{vec}(Q)$, respectively, for any symmetric conformable matrix Q .

$$\begin{aligned}
\|M_t\| &\leq \|M_{1t}\| + \|M_{2t}\| \\
&\leq (\|(E_d A_t E_d^\top)^{-1} E_d A_t E_s^\top\| + 1) \|A_t\| \\
&\leq (\lambda_{\min}^{-1}(A_t) \lambda_{\max}(A_t) + 1) \lambda_{\max}(A_t) \\
&\leq (N/\epsilon + 1)N < \infty
\end{aligned} \tag{2.29}$$

where we have used that $E_d A_t E_d^\top$ and $E_d A_t E_s$ are principal submatrices of A_t , and hence $\lambda_{\min}(E_d A_t E_d^\top) \geq \lambda_{\min}(A_t)$ and $\lambda_{\max}(E_d A_t E_s^\top) \leq \lambda_{\max}(A_t)$. Moreover, $\lambda_{\min}(A_t) = \lambda_{\min}(\Theta_t) \geq \epsilon > 0$ a.s. by Assumption (A6), and $\lambda_{\max}(A_t) = \lambda_{\max}(\Theta_t) \leq N$ a.s. because $\|\Theta_t\| \leq \|\Theta_t\|_2$, which is due to Equation (2.24). Thus, $\|d\text{vec}(Y_t^{-1/2})\|_2 \leq 2N^2 \epsilon^{-3} (N/\epsilon + 1) \|d\theta_t\|_2$, and the norm of the second term in Equation (2.25) is

$$\frac{1}{2} \|M_t dY_t^{-1/2}\| \leq (N/\epsilon)^3 (N/\epsilon + 1)^2 \|d\theta_t\|_2.$$

As for the first term, note that $\|\Theta_t^{-1/2} \otimes \Theta_t^{-1/2}\| \leq \lambda_{\min}^{-1}(\Theta) \leq \epsilon^{-1}$ by Assumption (A6), and

therefore

$$\frac{1}{2} \|(\Theta_t^{-1/2} \otimes \Theta_t^{-1/2}) dM_t\| \leq \frac{1}{2} \epsilon^{-1} \|dM_t\|$$

by submultiplicativity of the spectral norm.

Note that $dM_t = dM_{1t} + dM_{2t}$, where

$$\begin{aligned} dM_{1t} &= -E_s dA_t E_d^\top (E_d A_t E_d^\top)^{-1} P_d^\top A_t \\ &+ E_s A_t E_d^\top (E_d A_t E_d^\top)^{-1} (E_d dA_t E_d^\top) (E_d A_t E_d^\top)^{-1} P_d A_t \\ &- E_s A_t E_d^\top (E_d A_t E_d^\top)^{-1} P_d^\top dA_t. \end{aligned}$$

The matrix A_t in Equation (2.8) has an equivalent representation as $A_t = \int_0^1 \Theta_t^r \otimes \Theta_t^{1-r} dr$, see [Linton and McCrorie \(1995\)](#) and [Higham \(2008\)](#). Thus,

$$\begin{aligned} dA_t &= d \int_0^1 \Theta_t^r \otimes \Theta_t^{1-r} dr \\ &= \int_0^1 d\Theta_t^r \otimes \Theta_t^{1-r} dr + \int_0^1 \Theta_t^r \otimes d\Theta_t^{1-r} dr \\ &= \int_0^1 U_r \otimes \Theta_t^{1-r} dr + \int_0^1 \Theta_t^r \otimes V_r dr \end{aligned} \tag{2.30}$$

with

$$U_r := \int_0^1 \Theta_t^{(1-s)r} d(r \log \Theta_t) \Theta_t^{rs} ds \tag{2.31}$$

$$V_r := \int_0^1 \Theta_t^{(1-s)(1-r)} d((1-r) \log \Theta_t) \Theta_t^{(1-r)s} ds \tag{2.32}$$

This yields, using the triangle inequality,

$$\begin{aligned} \|dA_t\| &\leq \left\| \int_0^1 U_r \otimes \Theta_t^{1-r} dr \right\| + \left\| \int_0^1 \Theta_t^r \otimes V_r dr \right\| \\ &\leq \int_0^1 \|U_r \otimes \Theta_t^{1-r}\| dr + \int_0^1 \|\Theta_t^r \otimes V_r\| dr \end{aligned} \tag{2.33}$$

and

$$\begin{aligned} \|U_r\| &\leq \int_0^1 e^{\lambda_{max}(1-s)r} e^{\lambda_{max}rs} ds r \|d \log \Theta_t\| \\ &= \int_0^1 e^{\lambda_{max}r} ds r \|d \log \Theta_t\| \\ &= r e^{\lambda_{max}r} \|d \log \Theta_t\| \\ \|V_r\| &\leq (1-r) e^{\lambda_{max}(1-r)} \|d \log \Theta_t\| \end{aligned} \tag{2.34}$$

Therefore, $\|U_r \otimes \Theta_t^{1-r}\| \leq r e^{\lambda_{max}} \|d \log \Theta_t\|$ and $\|\Theta_t^r \otimes V_r\| \leq (1-r) e^{\lambda_{max}} \|d \log \Theta_t\|$, and thus,

$\|dA_t\| \leq (\int_0^1 r dr + \int_0^1 (1-r) dr) e^{\lambda_{max}} \|d \log \Theta_t\| = e^{\lambda_{max}} \|d \log \Theta_t\|$. Now,

$$\begin{aligned} \|d \log \Theta_t\| &\leq \|d \text{vec} \log \Theta_t\|_2 \\ &\leq \|A_t^{-1} M_t^\top d\theta_t\|_2 \\ &\leq (N/\epsilon + 1) \|d\theta_t\|_2 \end{aligned} \quad (2.35)$$

and therefore $\|dA_t\| \leq N(N/\epsilon + 1) \|d\theta_t\|_2$. Finally, $\|dM_{1t}\| \leq \|dA_t\| (2N/\epsilon + N^2/\epsilon^2)$ and $\|dM_t\| \leq \|dM_{1t}\| + \|dM_{2t}\| \leq \|dA_t\| (2N/\epsilon + N^2/\epsilon^2 + 1) \leq (2N^2/\epsilon + N^3/\epsilon^2 + N)(N/\epsilon + 1) \|d\theta_t\|_2$. To conclude,

$$\sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \left\| \frac{\partial \text{vec} g(\theta_t)}{\partial \theta_t^\top} \right\| = \sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \frac{\|d \text{vec} g(\theta_t)\|_2}{\|d\theta_t\|_2} \quad (2.36)$$

$$\leq (N/\epsilon)(N/\epsilon + 1)^2 \left\{ (N/\epsilon)^2 + (N/\epsilon + 1)/2 \right\} \quad (2.37)$$

$$=: B(\epsilon, N) < \infty \quad (2.38)$$

and the Lipschitz constant is given by

$$\Lambda(\psi) = \sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \left\| \frac{\partial \psi(\theta_t)}{\partial \theta_t^\top} \right\| = \sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \frac{\|d\psi(\theta_t)\|_2}{\|d\theta_t\|_2} = \sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \frac{\|\Phi d\theta_t + K dg(\theta_t) Z_t\|_2}{\|d\theta_t\|_2}$$

Note that

$$\Lambda(\psi) \leq \bar{\Lambda}(\psi) := \|\Phi\| + \|K\| \sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \left\| \frac{\partial \text{vec} g(\theta_t)}{\partial \theta_t^\top} \right\| \|Z_t\|_2 < \infty$$

where the upper bound $\bar{\Lambda}(\psi)$ is equal to $\phi_{max} + \kappa_{max} B(\epsilon, N) \|Z_t\|_2$ if Φ and K are diagonal with maximum diagonal elements given by ϕ_{max} and κ_{max} , respectively. This shows that the SRE is Lipschitz continuous with an upper bound for the Lipschitz constant given by $\bar{\Lambda}(\psi)$. For given parameters, we can simulate a large number of Z_t , calculate $\bar{\Lambda}(\psi)$ for each draw, and then estimate $\mathbb{E} \log |\bar{\Lambda}(\psi)|$ by the simulation average of $\log |\bar{\Lambda}(\psi)|$. Then, $\mathbb{E} \log |\bar{\Lambda}(\psi)| < 0$ implies that the Lyapounov exponent is negative, i.e. $\mathbb{E} \log |\Lambda(\psi)| < 0$.

In large dimensions, or when ϵ is very close to zero, this is unlikely to hold, so it is better to work with $\Lambda(\psi)$ directly and solve for given Z_t the problem

$$\Lambda(\psi) = \sup_{\theta_t \in \mathcal{T}_{N,\epsilon}} \left\| \Phi + (Z_t^\top \otimes K) \frac{\partial \text{vec} g(\theta_t)}{\partial \theta_t^\top} \right\|$$

where the solution for θ_t is found numerically over the space $\mathcal{T}_{N,\epsilon}$, and where Equation (2.25) is used for the differential of $g(\cdot)$. Then, one repeats this for many draws of Z_t and takes the average of $\log |\Lambda(\psi)|$ as an estimate of the Lyapounov exponent. We have done this for our six-dimensional empirical example (Model 2), using the estimated parameters in Table 2.14 and

100 draws of a multivariate student-t distribution. For the lower bound of the eigenvalues of Θ_t we choose $\epsilon = 0.1$, motivated by Figure 2.5 that shows the minimum eigenvalue of simulated Θ_t for the given parameters by increasing the sample size up to $T = 10^6$, which is not smaller than 0.2. We obtain an estimate of the Lyapounov exponent of -0.013, which suggests that Assumption (A7) holds for the estimated model.

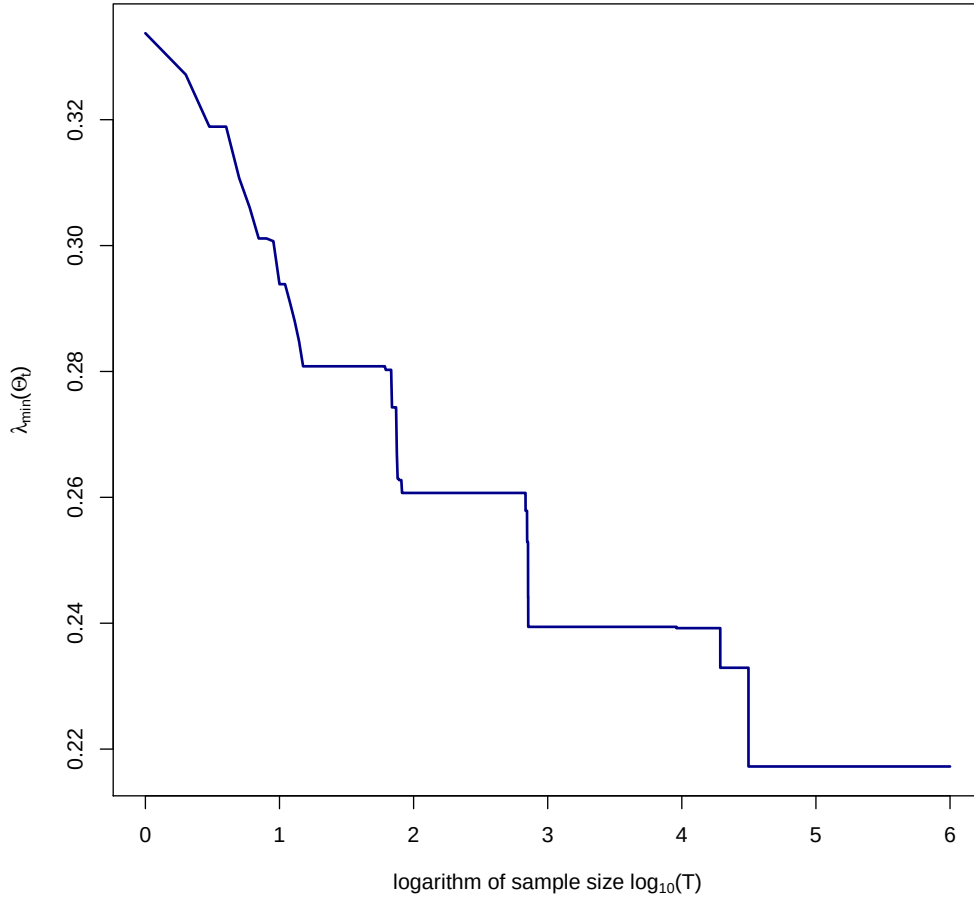


Figure 2.5: Simulated $\lambda_{\min}(\Theta_t)$ for $N = 6$ and different sample sizes up to 1,000,000 observations. The abscissa is the sample size on a log scale. The parameters of Θ_t are set to the estimates of Model 2 in our empirical example. The innovations are drawn from a multivariate student-t distribution, with the degrees of freedom being the average of the ones from the marginal distributions and from the copula parameter.

2.E Expressions for the information matrix

We give analytical expressions for the second derivatives in Equation (2.19) and Equation (2.20).

Note first that

$$\begin{aligned}\frac{\partial^2 l_t^C}{\partial \delta \partial \tau_i^\top} &= \frac{\partial \tilde{\theta}_t^\top}{\partial \delta} \frac{\partial \tilde{u}_t}{\partial \tau_i^\top} + (\tilde{u}_t^\top \otimes I_m) \frac{\partial \text{vec}}{\partial \tau_i^\top} \frac{\partial \tilde{\theta}_t^\top}{\partial \delta} \\ \frac{\partial^2 l_t^C}{\partial \delta \partial \delta^\top} &= \frac{\partial \tilde{\theta}_t^\top}{\partial \delta} \frac{\partial \tilde{u}_t}{\partial \delta^\top} + (\tilde{u}_t^\top \otimes I_m) \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial \tilde{\theta}_t^\top}{\partial \delta},\end{aligned}$$

where $m = \dim(\delta)$ and $\partial \tilde{\theta}_t^\top / \partial \delta$ is given in Equation (2.17), replacing θ_t by $\tilde{\theta}_t$, u_t by $\tilde{u}_t$, z_t by $\tilde{z}_t$ and x_t by $\tilde{x}_t$. Furthermore,

$$\begin{aligned}\frac{\partial \text{vec}}{\partial \tau_i^\top} \frac{\partial \tilde{\theta}_t^\top}{\partial \delta} &= \frac{\partial \text{vec}(\tilde{z}_{t-1}^\top)}{\partial \tau_i^\top} + (\tilde{x}_{t-1} \otimes I_m) \frac{\partial \text{vec}}{\partial \tau_i^\top} \frac{\partial \tilde{\theta}_{t-1}^\top}{\partial \delta} + (I_{N^*} \otimes \frac{\partial \tilde{\theta}_{t-1}^\top}{\partial \delta}) \frac{\partial \text{vec}(\tilde{x}_{t-1}^\top)}{\partial \tau_i^\top} \\ \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial \tilde{\theta}_t^\top}{\partial \delta} &= \frac{\partial \text{vec}(\tilde{z}_{t-1}^\top)}{\partial \delta^\top} + (\tilde{x}_{t-1} \otimes I_m) \frac{\partial \text{vec}}{\partial \delta^\top} \frac{\partial \tilde{\theta}_{t-1}^\top}{\partial \delta} + (I_{N^*} \otimes \frac{\partial \tilde{\theta}_{t-1}^\top}{\partial \delta}) \frac{\partial \text{vec}(\tilde{x}_{t-1}^\top)}{\partial \delta^\top}\end{aligned}$$

Next we calculate the derivatives involving the degrees of freedom parameter η . The first derivative is given by

$$\begin{aligned}\frac{\partial l}{\partial \eta} &= \frac{\partial c(\eta, N)}{\partial \eta} - N \frac{\partial c(\eta, 1)}{\partial \eta} - \frac{1}{2} \sum_{t=1}^T \frac{\partial l_t^C}{\partial \eta} \\ \frac{\partial l_t^C}{\partial \eta} &= \tilde{u}_t^\top \frac{\partial \tilde{\theta}_t^\top}{\partial \eta} + \log \left(1 + \frac{\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\eta - 2} \right) + \frac{\eta + N}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \left\{ 2 \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - \frac{\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\eta - 2} \right\} \\ &\quad - \sum_{i=1}^N \log \left(1 + \frac{\tilde{\varepsilon}_{it}^2}{\eta - 2} \right) + \frac{\eta + 1}{\eta - 2 + \tilde{\varepsilon}_{it}^2} \left\{ 2 \tilde{\varepsilon}_{it} \frac{\partial \tilde{\varepsilon}_{it}}{\partial \eta} - \frac{\tilde{\varepsilon}_{it}^2}{\eta - 2} \right\}\end{aligned}$$

The second derivatives can be obtained by straightforward calculations. First, we have

$$\begin{aligned}\frac{\partial^2 l_t^C}{\partial \eta \partial \tau_i^\top} &= \frac{\partial \tilde{\theta}_t^\top}{\partial \eta} \frac{\partial \tilde{u}_t}{\partial \tau_i^\top} + \tilde{u}_t^\top \frac{\partial^2 \tilde{\theta}_t^\top}{\partial \eta \partial \tau_i^\top} \\ &\quad + \left\{ \frac{\eta + N}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \left(2 \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - \frac{\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\eta - 2} \right) - \frac{N + 2}{\eta - 2} \right\} (\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t)^{-1} \frac{\partial \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \tau_i^\top} \\ &\quad - \left\{ \frac{\eta + 1}{\eta - 2 + \tilde{\varepsilon}_{it}^2} \left(2 \tilde{\varepsilon}_{it} \frac{\partial \tilde{\varepsilon}_{it}}{\partial \eta} - \frac{\tilde{\varepsilon}_{it}^2}{\eta - 2} \right) - \frac{3}{\eta - 2} \right\} (\eta - 2 + \tilde{\varepsilon}_{it}^2)^{-1} 2 \tilde{\varepsilon}_{it} \frac{\partial \tilde{\varepsilon}_{it}}{\partial \tau_i^\top} \\ &\quad + \frac{2(\eta + N)}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \left(\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial^2 \tilde{\varepsilon}_t}{\partial \eta \partial \tau_i^\top} + \frac{\partial \tilde{\varepsilon}_t^\top}{\partial \eta} \frac{\partial \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \tau_i^\top} \right) - \frac{2(\eta + 1)}{\eta - 2 + \tilde{\varepsilon}_{it}^2} \left(\tilde{\varepsilon}_{it} \frac{\partial^2 \tilde{\varepsilon}_{it}}{\partial \eta \partial \tau_i^\top} + \frac{\partial \tilde{\varepsilon}_{it}}{\partial \eta} \frac{\partial \tilde{\varepsilon}_{it}}{\partial \tau_i^\top} \right)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \tau_i^\top} &= 2\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \tau_i^\top} - 2(\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \otimes \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \tau_i^\top} \\ \frac{\partial \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \tau_i^\top} &= \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \tau_i^\top} - 2(\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \otimes \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \tau_i^\top}\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial^2 l_t^C}{\partial \eta \partial \delta^\top} &= \frac{\partial \tilde{\theta}_t^\top}{\partial \eta} \frac{\partial \tilde{u}_t}{\partial \delta^\top} + \tilde{u}_t^\top \frac{\partial^2 \tilde{\theta}_t}{\partial \eta \partial \delta^\top} \\ &+ \left\{ \frac{\eta + N}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \left(2\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - \frac{\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\eta - 2} \right) - \frac{N + 2}{\eta - 2} \right\} (\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t)^{-1} \frac{\partial \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \delta^\top} \\ &+ \frac{2(\eta + N)}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \left(\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial^2 \tilde{\varepsilon}_t}{\partial \eta \partial \delta^\top} + \frac{\partial \tilde{\varepsilon}_t^\top}{\partial \eta} \frac{\partial \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \delta^\top} \right)\end{aligned}$$

and finally,

$$\begin{aligned}\frac{\partial l^2}{\partial \eta^2} &= \frac{\partial^2 c(\eta, N)}{\partial \eta^2} - N \frac{\partial^2 c(\eta, 1)}{\partial \eta^2} - \frac{1}{2} \sum_{t=1}^T \frac{\partial^2 l_t^C}{\partial \eta^2} \\ \frac{\partial^2 l_t^C}{\partial \eta^2} &= \frac{\partial \tilde{\theta}_t^\top}{\partial \eta} \frac{\partial \tilde{u}_t}{\partial \eta} + \tilde{u}_t^\top \frac{\partial^2 \tilde{\theta}_t}{\partial \eta \partial \eta} + (\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t)^{-1} \times \\ &\times \left\{ \left(2\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - \frac{\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\eta - 2} \right) \left[1 - \frac{N + 2}{\eta - 2} - \frac{N + \eta}{\eta - 2 + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t} \left(1 + \frac{\partial \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \eta} \right) \right] \right. \\ &+ 2 \frac{N + 2}{\eta - 2} (\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \otimes \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \eta} + 2(\eta + N) \left[\frac{\partial \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1}}{\partial \eta} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} + \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial^2 \tilde{\varepsilon}_t}{\partial \eta^2} \right] \Big\} \\ &- \sum_{i=1}^N (\eta - 2 + \tilde{\varepsilon}_{it}^2)^{-1} \times \\ &\times \left\{ \left(2\tilde{\varepsilon}_t \frac{\partial \tilde{\varepsilon}_{it}}{\partial \eta} - \frac{\tilde{\varepsilon}_{it}^2}{\eta - 2} \right) \left[\frac{\eta - 5}{\eta - 2} - \frac{1 + \eta}{\eta - 2 + \tilde{\varepsilon}_{it}^2} \left(1 + \frac{\partial \tilde{\varepsilon}_{it}^2}{\partial \eta} \right) \right] + 2(\eta + 1) \left[\left(\frac{\partial \tilde{\varepsilon}_{it}}{\partial \eta} \right)^2 + \tilde{\varepsilon}_{it} \frac{\partial^2 \tilde{\varepsilon}_{it}}{\partial \eta^2} \right] \right\}\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \eta} &= 2\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - 2(\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \otimes \tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \eta} \\ \frac{\partial \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \eta} &= \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - 2(\tilde{\varepsilon}_t^\top \tilde{\Theta}_t^{-1} \otimes \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \eta} \\ \frac{\partial \tilde{\theta}_t}{\partial \eta} &= \Phi \frac{\partial \tilde{\theta}_{t-1}}{\partial \eta} + K \frac{\partial \tilde{u}_{t-1}}{\partial \eta} \\ \frac{\partial^2 \tilde{\theta}_t}{\partial \eta^2} &= \Phi \frac{\partial^2 \tilde{\theta}_{t-1}}{\partial \eta^2} + K \frac{\partial^2 \tilde{u}_{t-1}}{\partial \eta^2}\end{aligned}$$

2.F Monte Carlo simulation

We carry out a Monte Carlo study to investigate the finite sample performance of the two estimation methods. We simulate three return series ($N = 3$) of length $T = 1000$ and $T = 5000$ observations. The parameters used in the data generating process are taken from our empirical example, using the three assets SPX, SBX, and EMU. For the degrees of freedom, we take the average of the degrees of freedom parameters from the marginal distributions and the one from the copula. The values are provided in [Table 2.4](#).

Table 2.4: Parameters used in the data generating process.

Volatility parameters			Correlation parameters			degrees of freedom
ω^v	κ^v	ϕ^v	ω			$\bar{\eta}$
-0.3133	0.1164	0.9355	-0.0250	0.0725	0.0202	0.1335
-0.0348	0.0200	0.9938	K	Φ		
-0.0909	0.0453	0.9803	0.0675	0.8715		

We generate 100 replications of the return series and estimate the model using both the one-step and the two-step estimation procedures. Our two measures of accuracy are the Bias and the root mean squared error (RMSE). For each parameter s , the measures are computed as $\text{Bias} = \frac{1}{100} \sum_{i=1}^{100} \hat{s}_i - s$, and $\text{RMSE} = \frac{1}{100} \sum_{i=1}^{100} \sqrt{(\hat{s}_i - s)^2}$.

The results are presented in [Table 2.5](#). The left-hand panel compares the evolution of the bias in the parameter estimates across different sample sizes for both the one-step and two-step estimation procedures. For the volatility dynamics, we observe downward biases for ω^v and ϕ^v , and upward biases for κ^v . For the correlation dynamics, the estimates ω and κ are only slightly biased, while the estimates for the mean reversion parameter ϕ are downward biased, especially for the one-step method. In all cases, the absolute biases are reduced when increasing the sample size. For the degrees of freedom parameter $\bar{\eta}$, we observe a slightly higher upward bias for the two-step method, which might be due to the aggregation of the marginal degrees of freedom and of the copula one into a single parameter. The bias remains small, however, and decreases with larger samples. When the sample size is increased from $T = 1000$ to $T = 5000$, we observe an overall reduction in the magnitude of the estimation bias for both the one-step and two-step estimation procedures.

The right-hand panel compares the evolution of the RMSE in the parameter estimates across

different sample sizes for both the one-step and two-step estimation procedures. For the volatility dynamics, we notice that the one-step and two-step estimation procedures have comparable RMSE of the parameter estimates. For the correlation dynamics, the two-step estimation procedure exhibits lower RMSE than the one-step estimation case, which is mainly due to the lower biases. When the sample size is increased from $T = 1000$ to $T = 5000$, we observe an overall reduction in the RMSE of the parameter estimates for both the one-step and two-step estimation procedures.

The general picture is that the two-step method is performing better for the considered parameterization and sample sizes. It has faster computation times (the average estimation times are indicated in the last row of [Table 2.5](#)), and generally better performances than the one-step method in terms of RMSE.

Table 2.5: Bias and RMSE of parameter estimates.

	Bias				RMSE			
	T=1000		T=5000		T=1000		T=5000	
	one-step	two-step	one-step	two-step	one-step	two-step	one-step	two-step
ω_1^v	-0.0413	-0.0384	-0.0012	0.0007	0.0753	0.0748	0.0270	0.0273
ω_2^v	-0.0770	-0.0749	-0.0090	-0.0072	0.0844	0.0842	0.0131	0.0129
ω_3^v	-0.0416	-0.0479	-0.0053	-0.0047	0.0506	0.0585	0.0149	0.0159
κ_1^v	0.0019	0.0009	0.0008	-0.0000	0.0106	0.0114	0.0047	0.0053
κ_2^v	0.0014	0.0012	0.0002	0.0001	0.0053	0.0056	0.0022	0.0023
κ_3^v	0.0006	0.0010	0.0001	-0.0005	0.0069	0.0075	0.0033	0.0037
ϕ_1^v	-0.0084	-0.0078	-0.0002	0.0001	0.0155	0.0154	0.0056	0.0056
ϕ_2^v	-0.0137	-0.0133	-0.0016	-0.0013	0.0150	0.0150	0.0023	0.0023
ϕ_3^v	-0.0090	-0.0103	-0.0011	-0.0010	0.0109	0.0127	0.0032	0.0034
ω_1	-0.0019	-0.0013	0.0002	-0.0019	0.0186	0.0110	0.0142	0.0069
ω_2	0.0014	0.0001	-0.0048	0.0027	0.0531	0.0243	0.0430	0.0123
ω_3	0.0028	-0.0020	0.0015	0.0004	0.0154	0.0131	0.0118	0.0034
K	0.0056	-0.0035	0.0101	-0.0029	0.0857	0.0123	0.0612	0.0067
Φ	-0.0665	-0.0131	-0.0621	-0.0074	0.1164	0.0375	0.0819	0.0197
$\bar{\eta}$	0.0023	0.0148	0.0045	0.0081	0.0111	0.0267	0.0070	0.0142
Avg. t	1.9223	0.4938	8.6614	2.4342	1.9223	0.4938	8.6614	2.4342

Note: Avg. t is the average running time of each estimation over 100 replications expressed in hour units.

2.G Additional tables and figures

Table 2.6: Description of the assets considered in the empirical application.

	SECURITY	DESCRIPTION
SPX	SPX Index	S&P 500 Index
SBW	SBWGU Index	FTSE World Government Bond Index (WGBI) USD
EMU	NDDUEMU Index	MSCI EMU Net Total Return Index
OIL	CL1 Comdty	Generic 1st Crude Oil; WTI
GOLD	GOLDLNPM Index	LBMA Gold Price PM USD
CRIX	CRIX	Cryptocurrency Index

Table 2.7: Descriptive statistics for historical log-returns.

	SPX	SBW	EMU	OIL	GOLD	CRIX
N.Obs	1220	1220	1220	1220	1220	1220
Minimum	-0.0999	-0.0225	-0.1514	-0.2799	-0.0526	-0.4466
Maximum	0.0897	0.0181	0.0857	0.3196	0.0513	0.1985
Mean	0.0005	0.0001	0.0003	0.0014	0.0002	0.0011
Stdev	0.0111	0.0038	0.0122	0.0311	0.0089	0.0426
Skewness	-0.1796	-0.3471	-1.8953	1.5560	0.0410	-1.3486
Kurtosis	15.7469	3.8632	25.1767	23.0595	4.1170	13.8111
Sharpe Ratio	0.0470	0.0132	0.0269	0.0448	0.0204	0.0267
Beta	1.1666	0.1127	0.9668	3.1250	0.4091	2.5494

Note: The kurtosis reported in the table corresponds to the excess kurtosis.

Table 2.8: The sample correlation matrix of returns.

	SPX	SBW	EMU	OIL	GOLD	CRIX
SPX	1.0000	-0.1380	0.5753	0.2836	0.0766	-0.0898
SBW	-0.1380	1.0000	0.0831	-0.0288	0.5476	0.0600
EMU	0.5753	0.0831	1.0000	0.2473	0.1097	0.0385
OIL	0.2836	-0.0288	0.2473	1.0000	0.0578	0.0103
GOLD	0.0766	0.5476	0.1097	0.0578	1.0000	0.0515
CRIX	-0.0898	0.0600	0.0385	0.0103	0.0515	1.0000

Table 2.9: Parameter estimates of the volatility using Beta-t-EGARCH(1,1).

	ω_i	κ_i	ϕ_i	$\bar{\eta}_i$
SPX	-0.31327	0.11642	0.93555	0.19126
	(0.00398)	(0.00069)	(0.00081)	(0.00121)
SBW	-0.03482	0.02004	0.99383	0.11880
	(0.00146)	(0.00029)	(0.00026)	(0.00120)
EMU	-0.09091	0.04531	0.98032	0.12895
	(0.00191)	(0.00046)	(0.00041)	(0.00097)
OIL	-0.09397	0.04746	0.97529	0.11695
	(0.00164)	(0.00037)	(0.00042)	(0.00104)
GOLD	-0.01888	0.01576	0.99613	0.18421
	(0.00076)	(0.00021)	(0.00016)	(0.00131)
CRIX	-0.17057	0.17687	0.94220	0.40950
	(0.00196)	(0.00102)	(0.00065)	(0.00150)

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta}_i = \frac{1}{\eta_i} \in (0, 1)$.

Table 2.10: Volatility model diagnostics.

	SPX	SBW	EMU	OIL	GOLD	CRIX
score	0.4224	0.8021	0.1333	0.1478	0.1643	0.9878
ε	0.7751	0.6110	0.4500	0.3886	0.3072	0.0026
ε^2	0.6223	0.0623	0.7234	0.3237	0.2229	0.0577

Note: The table reports the p-values of Box-Pierce test applied to the score with respect to log volatility, the standardized residuals (ε) and squared standardized residuals (ε^2). The number of lags being tested is 30.

Table 2.11: Descriptive statistics for standardized residuals.

	SPX	SBW	EMU	OIL	GOLD	CRIX
N.Obs	845	845	845	845	845	845
Minimum	-6.0957	-4.2156	-6.7406	-4.0590	-4.2477	-5.8446
Maximum	3.3762	3.9590	3.7776	4.1949	4.5662	3.8765
Mean	0.0294	-0.0079	0.0122	0.0018	-0.0139	0.0183
Stdev	1.0268	0.9936	1.0077	0.9954	0.9882	0.7815
Skewness	-1.1159	-0.0773	-0.4852	-0.2146	0.1064	-0.9837
Kurtosis	4.8808	0.8876	2.8387	1.1157	1.6320	8.4533

Note: The kurtosis reported in the table corresponds to the excess kurtosis.

Table 2.12: Two-step parameter estimates for the correlation part of the model. The diagonal elements of K and Φ are not restricted to be the same (Model 1).

	ω_{i*5+1}	ω_{i*5+2}	ω_{i*5+3}	ω_{i*5+4}	ω_{i*5+5}
ω	-0.03331	0.56013	0.12427	-0.00364	-0.01029
	(0.00990)	(0.02140)	(0.02359)	(0.00187)	(0.00833)
	0.00527	-0.01730	0.01239	0.00103	0.00208
	(0.00804)	(0.00460)	(0.03263)	(0.00079)	(0.00618)
	0.00201	0.03624	0.09152	0.00221	0.00022
	(0.00094)	(0.00805)	(0.01145)	(0.00199)	(0.00183)
	κ_{i*5+1}	κ_{i*5+2}	κ_{i*5+3}	κ_{i*5+4}	κ_{i*5+5}
K	0.05506	0.05420	0.02531	0.00771	0.00000
	(0.03334)	(0.03482)	(0.03438)	(0.03439)	(0.03440)
	0.04870	0.02947	0.01731	0.00551	0.01755
	(0.02108)	(0.03419)	(0.02156)	(0.03439)	(0.03233)
	0.04583	0.02583	0.03543	0.01497	0.03328
	(0.03431)	(0.03441)	(0.03447)	(0.03419)	(0.03441)
	ϕ_{i*5+1}	ϕ_{i*5+2}	ϕ_{i*5+3}	ϕ_{i*5+4}	ϕ_{i*5+5}
Φ	0.82329	0.00000	0.52003	0.92355	0.28273
	(0.04426)	(0.03440)	(0.04968)	(0.03543)	(0.03440)
	0.96677	0.59401	0.98322	0.91835	0.98210
	(0.04748)	(0.03699)	(0.04363)	(0.03539)	(0.04079)
	0.89026	0.23319	0.02115	0.94376	0.95013
	(0.06085)	(0.03451)	(0.03440)	(0.04146)	(0.05646)
$\bar{\eta}$	0.02863				
	(0.00670)				

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates, and $i = 0, 1, 2$. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta}_i = \frac{1}{\eta_i} \in (0, 1)$.

Table 2.13: Model descriptions.

	DESCRIPTION	# Pars
Model 1	full ω , diagonal K , diagonal Φ	46
Model 2	full ω , scalar K , scalar Φ	18
Model 3	scalar ω , diagonal K , diagonal Φ	32
Model 4	$\omega = 0$, diagonal K , diagonal Φ	31
Model 5	scalar ω , scalar K , scalar Φ	4
Model 6	$\omega = 0$, scalar K , scalar Φ	3

Note: # Pars is the number of parameters.

Table 2.14: Parameter estimates for the dynamic correlation models.

	Meta-t-log-corr	Meta-t-cDCC	Meta-t-GAS	Meta-t-GAS-h
κ	0.01901	0.02065	0.01919	0.01862
	(0.00015)	(0.00016)	(0.00015)	(0.00015)
ϕ	0.95527	0.93750	0.96162	0.96098
	(0.00057)	(0.00061)	(0.00047)	(0.00051)
$\bar{\eta}$	0.03258	0.03495	0.03326	0.03312
	(0.00026)	(0.00028)	(0.00027)	(0.00027)

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta}_i = \frac{1}{\eta_i} \in (0, 1)$.

Table 2.15: Parameter estimates for the dynamic volatility and correlation models.

	t-cDCC	t-GAS	t-GAS-l	t-GAS-hl
α_1	0.16976	0.16711	0.13417	0.13107
	(0.05958)	(0.00234)	(0.00078)	(0.02919)
α_2	0.01799	0.01880	0.02004	0.01854
	(0.05958)	(0.03680)	(0.00020)	(0.01904)
α_3	0.06530	0.04153	0.04015	0.04063
	(0.05959)	(0.00708)	(0.00045)	(0.03273)
α_4	0.07391	0.07273	0.06185	0.06301
	(0.05964)	(0.00341)	(0.00051)	(0.04186)
α_5	0.02753	0.02004	0.02191	0.02218
	(0.05958)	(0.00874)	(0.00019)	(0.01375)
α_6	0.21436	0.15464	0.16283	0.15848
	(0.05966)	(0.00646)	(0.00071)	(0.02246)
α_7	0.01811	0.01003	0.01355	0.01439
	(0.00113)	(0.03440)	(0.00015)	(0.00578)
β_1	0.77216	0.93407	0.91498	0.90985
	(0.05957)	(0.00603)	(0.00082)	(0.05781)
β_2	0.98046	1.00000	0.99854	0.99921
	(0.00097)	(0.03440)	(0.00013)	(0.02684)
β_3	0.91487	0.99212	0.98609	0.98444
	(0.00017)	(0.00281)	(0.00028)	(0.03189)
β_4	0.89709	0.97438	0.97530	0.97599
	(0.00082)	(0.00517)	(0.00042)	(0.06085)
β_5	0.96657	0.99882	0.99736	0.99713
	(0.00063)	(0.00866)	(0.00011)	(0.02868)
β_6	0.73315	0.97121	0.92509	0.91195
	(0.00025)	(0.00726)	(0.00073)	(0.05214)
β_7	0.93653	0.90664	0.97642	0.96149
	(0.00140)	(0.03440)	(0.00046)	(0.04242)
$\bar{\eta}$	0.13152	0.13068	0.13378	0.12944
	(0.00022)	(0.00948)	(0.00044)	(0.00700)

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta}_i = \frac{1}{\eta_i} \in (0, 1)$. α_i and β_i with $i = 1, \dots, 6$ are the volatility parameters. α_7 and β_7 are the correlation parameters.

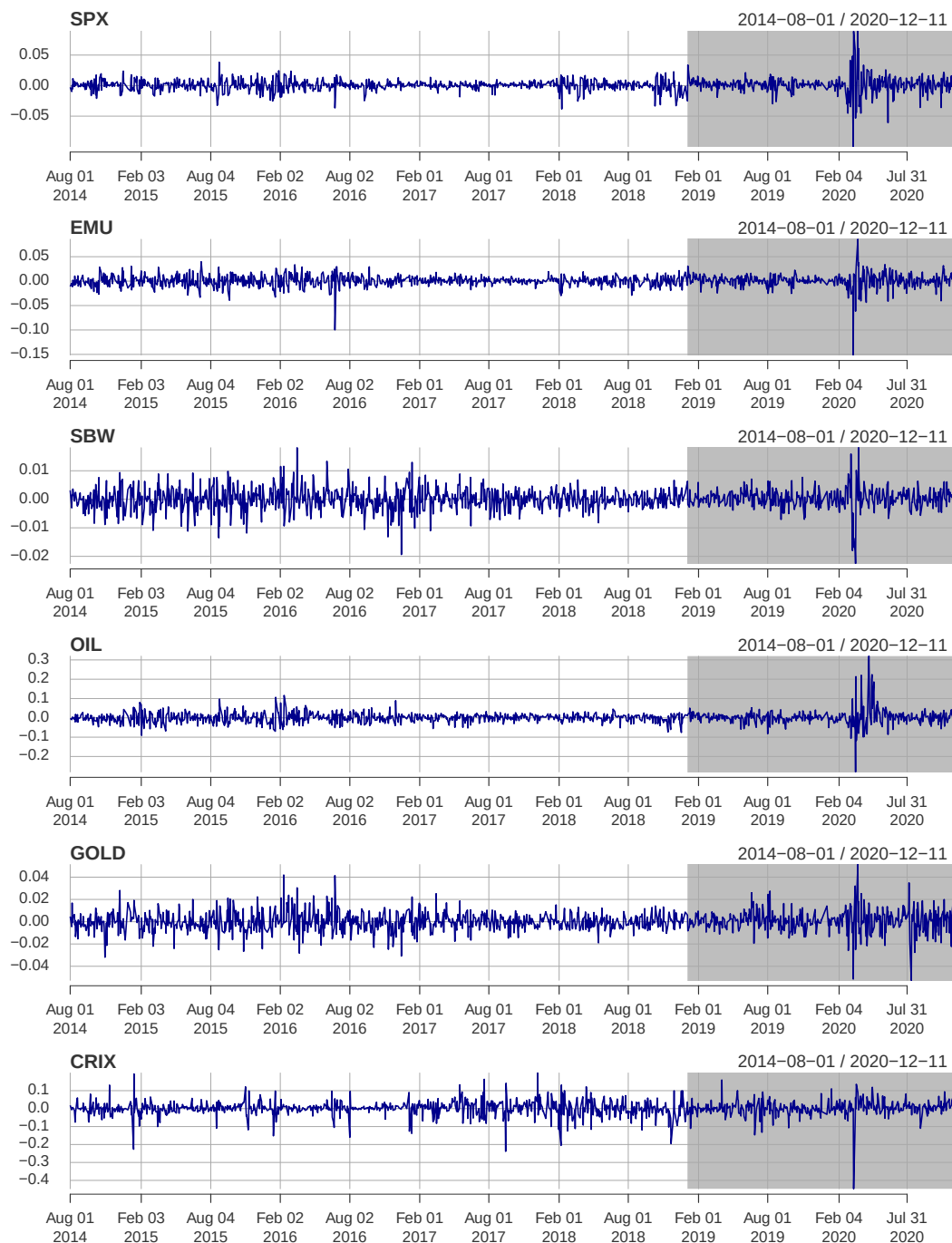


Figure 2.6: Time series of daily log-returns, August 1, 2014 to December 12, 2020, for (from top to bottom) SPX, SBW, EMU, OIL, GOLD, and CRIX. The grey shaded area is the out-of-sample period ranging from the beginning of 2019 to the end of 2020.

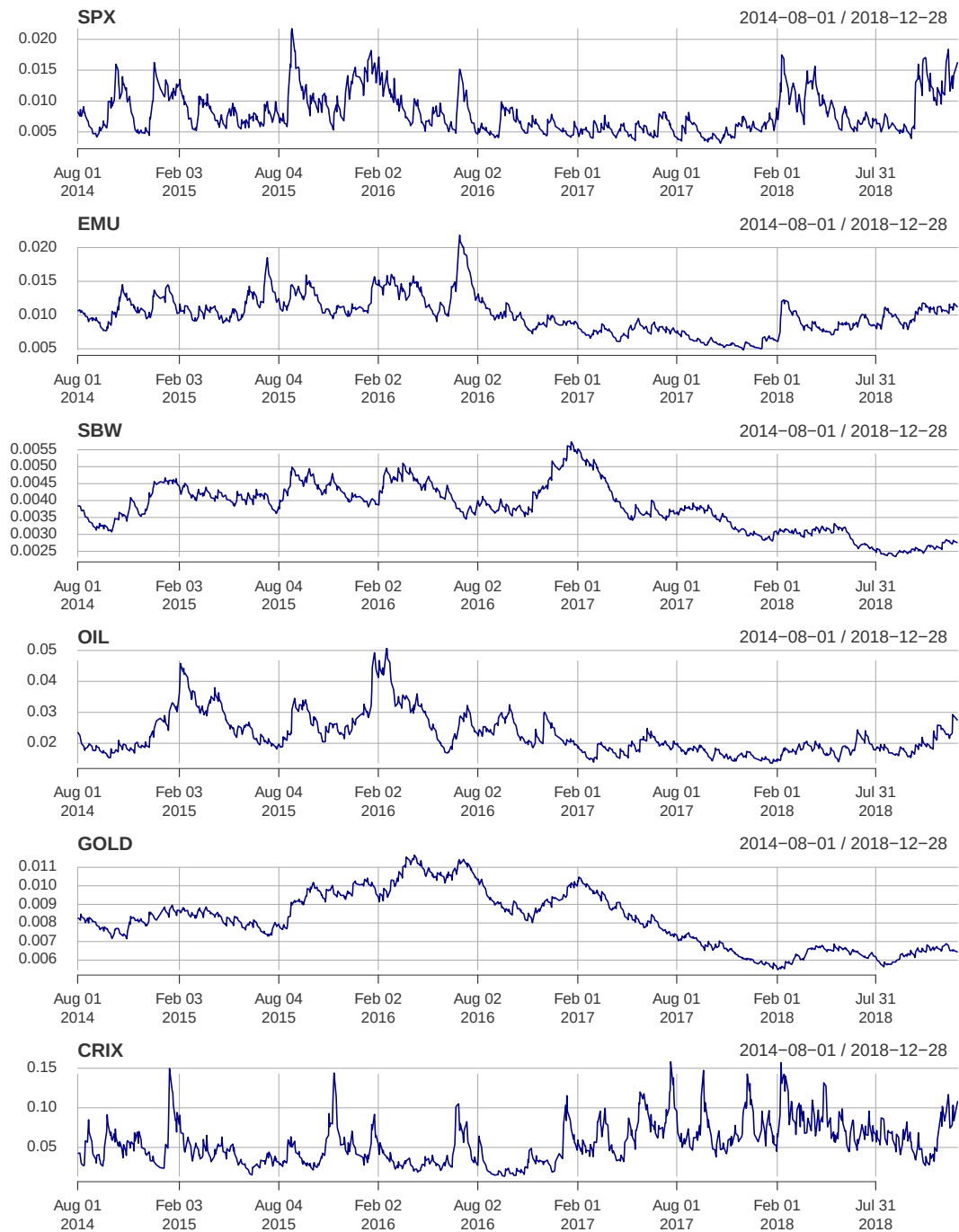


Figure 2.7: Estimated volatility series using the Beta-t-EGARCH model.

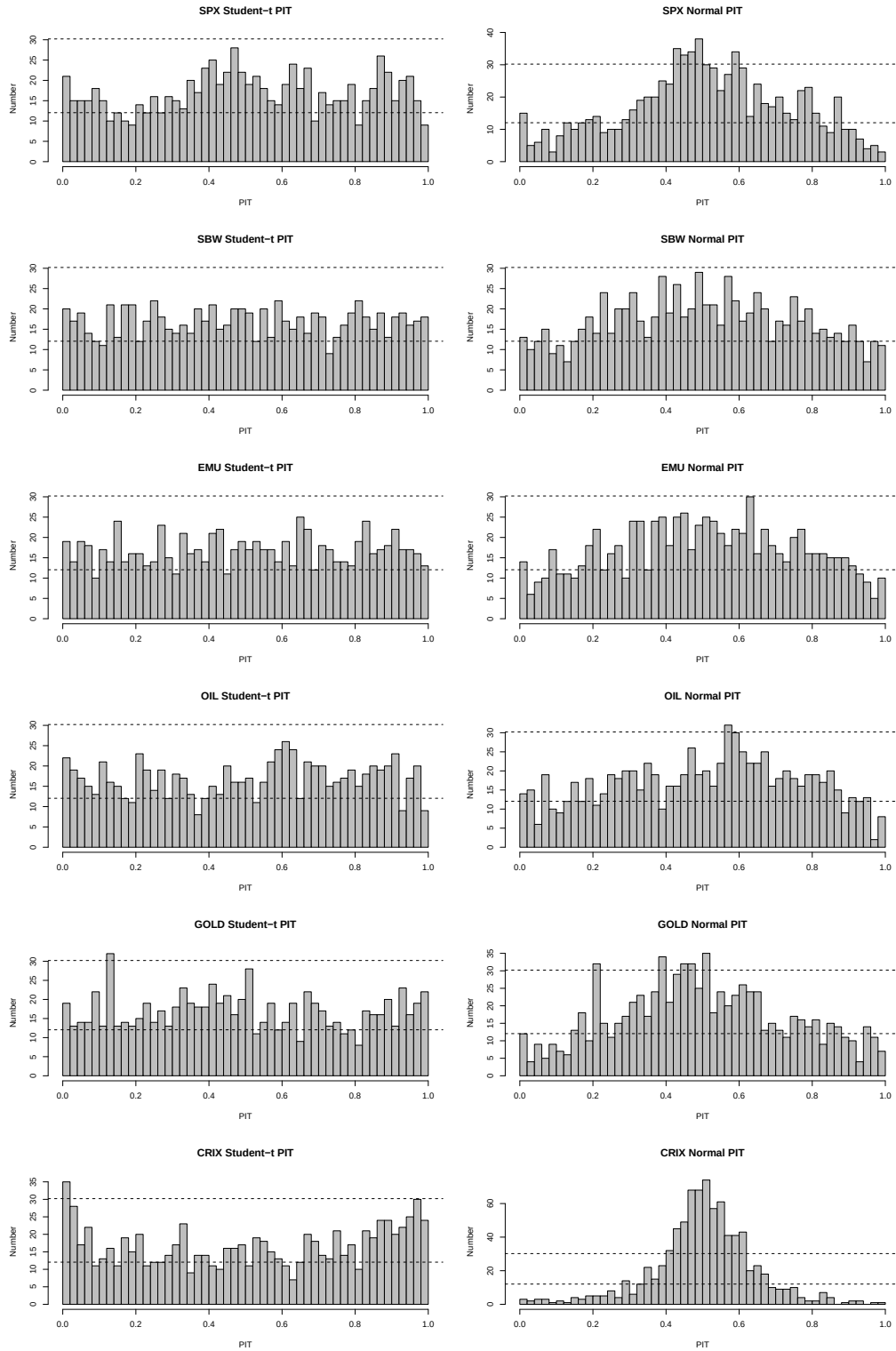


Figure 2.8: Histograms of probability integral transforms (PITs) of standardized residuals of the Beta-t-EGARCH model. We use the student-t distribution with the estimated d.o.f. for figures in the left-hand column. For the figures in the right-hand column, we use the Gaussian cdf for the transformation.

Chapter 3

Dynamic portfolio selection with sector-specific regularization¹

3.1 Introduction

Portfolio selection remains one of the central topics of empirical finance since its inception by [Markowitz \(1952\)](#). It is especially challenging in the context of portfolio selection with a large number of assets, where it is well known that naive implementation of the [Markowitz \(1952\)](#) results using sample means and variance-covariance matrices leads to poor out-of-sample performance due to unstable and ill-conditioned estimates. Remedies have been proposed such as shrinkage methods as in [Ledoit and Wolf \(2004\)](#) and [Ledoit and Wolf \(2017\)](#) and weight regularization as in [DeMiguel et al. \(2009\)](#), [Brodie et al. \(2009\)](#) and [Fastrich et al. \(2015\)](#), which stabilize the ill-conditioned optimization problem. Statistical justifications for regularization techniques in a portfolio selection context are given by e.g. [Puelz et al. \(2015\)](#) and [Fisher et al. \(2020\)](#). [Candelon et al. \(2012\)](#) reformulate traditional shrinkage estimators in terms of a linear regression framework and impose a “double” shrinkage by adding additional weight penalties such as lasso or ridge, which further stabilizes portfolio weights and decreases portfolio turnover. Similarly, [Ao et al. \(2019\)](#) show an equivalent representation of the mean-variance optimization problem as an unconstrained regression, which they augment by lasso-type constraints to obtain

¹This chapter corresponds to Hafner, C.M. and L. Wang (2022): “Dynamic portfolio selection with sector-specific regularization”, forthcoming at *Econometrics and Statistics*, doi: <https://doi.org/10.1016/j.ecosta.2022.01.001>.

sparsity and obtain consistency of mean and risk as both the number of assets and sample size increase. Furthermore, as shown by [Jagannathan and Ma \(2003\)](#), a no-short-sale constraint also serves to stabilize weights and regularize the optimization problem, as it is equivalent to shrinking large elements of the covariance matrix. They show empirically that short-sale constrained portfolios perform as well as those constructed using covariance matrices estimated from factor models and shrinkage estimators. [DeMiguel et al. \(2009\)](#) and [Fan et al. \(2012\)](#) generalize this idea by introducing a gross-exposure constraint that bridges the gap between an unconstrained and no-short-sale constrained portfolio. [Fan et al. \(2012\)](#) give a theoretical justification for the empirical results in [Jagannathan and Ma \(2003\)](#).

Although the main objective of portfolio selection is typically formulated in terms of maximization of the Sharpe ratio, which corresponds to the tangency portfolio of [Markowitz \(1952\)](#), the estimation and prediction of asset mean returns have proven to be particularly difficult and noisy, see e.g. [Best and Grauer \(1991\)](#). Many studies, starting at least with [Chan et al. \(1999\)](#), therefore neglect the mean and concentrate on the minimization of the portfolio variance to obtain estimates of the global minimum variance portfolio (GMV). As shown by [Jagannathan and Ma \(2003\)](#), the GMV outperforms many competing approaches involving estimation of the mean in terms of Sharpe ratio. This has even been shown for the naive equally-weighted portfolio by [DeMiguel et al. \(2009\)](#). We therefore refrain from estimating the mean in this paper and consider the problem of estimating the GMV portfolio.

In a dynamic framework, it is important to take time-varying volatilities and correlations into account. Many alternative modelling strategies exist, including Bayesian dynamic linear models, see e.g. [Puelz et al. \(2020\)](#), but we follow a large part of the financial econometrics literature by implementing a GARCH-type model combined with dynamic conditional correlations (DCC) as in [Engle \(2002\)](#). These models can be used to predict volatilities and correlations up to a certain investment horizon, see [Baillie and Bollerslev \(1992\)](#) and [Engle and Sheppard \(2001\)](#). In large dimensions, issues arise for the estimation of DCC-type models that can be addressed by the composite likelihood method as in [Pakel et al. \(2020\)](#), combined with shrinkage of the sample covariance matrix that is used for correlation targeting as in [Hafner and Reznikova \(2012\)](#). See also [Morana \(2019\)](#) for an alternative approach to DCC estimation in high dimensions based on regularised semiparametric methods. For portfolio selection with many assets, [Engle et al. \(2019\)](#) implement a DCC-GARCH model with composite likelihood estimation and nonlinear

shrinkage and show its superior out-of-sample performance with respect to benchmark portfolios, in particular the ones not taking the dynamics of volatilities and correlations into account.

Despite the vast literature on portfolio selection, not many studies have taken information about the industry sectors explicitly into account, although it is well known that this information helps to improve predictability of mean returns, volatilities and correlations. Companies within a given industry sector often compete in the same product market and co-move regarding product and technology innovations. They react similarly to permanent shifts in supply and demand conditions, as well as the regulatory environment. As the industry goes through expansions and contractions, companies' growth opportunities and investing and financing decisions are correlated. For example, [Moskowitz and Grinblatt \(1999\)](#) document the strong and persistent effects of industry components in stock returns, where industry momentum strategies are significant and more profitable than individual stock momentum strategies. [Hou \(2007\)](#) argue that the lead-lag effect is driven by an intra-industry phenomenon, where returns on big firms lead returns on small firms within the same industry, which is primarily caused by stock prices' slow response to negative information. Hence, there is information clustering at the industry level. [Hong et al. \(2007\)](#) find that stock markets react with a delay to information contained in industry returns about their fundamentals and that information diffuses only gradually across markets. [Brito et al. \(2018\)](#) forecast very large realized covariance matrices of returns using standard firm-level factors (e.g., size, value and profitability) and imposing additional sectoral restrictions in the residual covariance matrix. [Kurose and Omori \(2020\)](#) propose a multi-block equicorrelation structure to estimate multivariate stochastic volatility models in large dimensions and apply their model to an asset allocation exercise with sector as blocks. For portfolio selection, [Chen et al. \(2020\)](#) take a sector-specific structure into account by allowing to incorporate investor preferences with respect to industry sectors. They propose a sparse-group selection with the objective of being sparse across sectors but diversified within favored sectors. [Fan et al. \(2016\)](#) include sector specific information into a factor model where industry ETFs are included as additional factors to the CAPM or the three-factor model of [Fama and French \(1993\)](#).

Sparsity and diversification, although conflicting in nature, both are desirable properties of portfolios, since diversification allows to reduce portfolio weight variability, while sparsity keeps the number of invested assets and hence transaction costs under control. Moreover, both have a stabilizing effect on portfolio weights, if diversification is understood as shrinking towards a

fixed weighting scheme. There are however many ways to introduce these properties into the selection procedure. The choice of [Chen et al. \(2020\)](#) is to impose sparsity between industry sectors and diversification within sectors via the sparse group lasso of [Friedman et al. \(2010\)](#) and [Simon et al. \(2013\)](#), an extension of the group lasso of [Yuan and Lin \(2006\)](#). It is not *a priori* clear whether this is the best strategy, as one might want to allow for sector-wide diversification as well, rather than investing in only a few sectors. Furthermore, the study of [Chen et al. \(2020\)](#) is using a rolling window updating mechanism. From an efficiency point of view, it would be preferable to have a genuine dynamic model that can be estimated using all available historical data, and that produces analytical forecasts of volatilities and correlations, as mentioned above.

In this paper, we adopt the perspective of a portfolio manager who optimally allocates funds across assets subject to a set of criteria motivated by investment preferences or constraints. Relevant examples of such criteria include restrictions on sectoral exposures, guided for instance by preliminary screening of relevant investable sectors ([Chen et al. \(2020\)](#)), or constraints on the distribution of portfolio weights across assets, e.g. by favouring sparse asset allocations ([Puelz et al. \(2020\)](#)), and over time by promoting stable asset allocation minimizing transaction costs ([Hautsch and Voigt \(2019\)](#)).

To this end, we propose a new portfolio selection procedure in a dynamic framework that explicitly accounts for these economically motivated criteria by regularising portfolio weights accordingly. More specifically, building on the DCC-GARCH model of [Engle et al. \(2019\)](#), which allows for the modelling of time-varying volatilities and correlations in large dimensions, we add restrictions on portfolio weights to reflect economically motivated criteria such as between- and within-sector diversification. We also consider penalty terms promoting sparsity in the resulting asset allocation and controlling for transaction costs by stabilizing variations in portfolio weights over time. Our proposed framework is modular as it allows to combine different penalty terms to obtain allocations satisfying multiple criteria simultaneously, e.g. combining between-/within-sector diversification with a cost penalty to obtain diversified portfolios with weights that are more stable over time. In addition, our framework can flexibly accommodate any dynamic input model to generate the portfolio weights, and is easy to implement as the penalty parameter controls the degree to which the resulting allocation will promote the imposition of the economically motivated criteria.

Our contribution is twofold: First, we introduce an algorithm for optimally choosing the

penalization parameters in a data-driven way based on cross-validation in a dynamic context. The algorithm uses a partition of the training period into blocks, for which the combination of DCC-GARCH and regularization delivers portfolio variances that can be minimized with respect to the regularization parameters. These parameters are then used to construct portfolio weights in the out-of-sample evaluation period. The proposed algorithm generalizes classical cross-validation procedures, as e.g. in [DeMiguel et al. \(2009\)](#), to take the time series structure of the data into account.

Our second contribution is a large scale empirical study to investigate how a regularization with respect to the sector structure helps to improve out-of-sample performance of portfolio return variances. We select stocks of the S&P Total index up to dimension 500, with daily frequency for the returns and a monthly frequency for portfolio re-balancing. We obtain several important findings. First, diversification is clearly more important within sectors than between sectors. Second, promoting sparsity of portfolio weights strongly contributes to reducing portfolio variances. This control for sparsity reduces negative weights and effectively moves the optimal portfolio towards a no short-sale constrained portfolio. The best performing model in large cross-sections combines within-sector diversification with a control for sparsity. The optimal portfolios significantly outperform benchmark portfolios such as the equally-weighted or the DCC-GARCH portfolio without sector-wise regularization. Third, when a no-short-sale restriction is added, we observe that the best performing portfolios all feature the within-sector regularization. However, the difference with respect to the DCC-GARCH benchmark is smaller and not statistically significant for larger portfolios. Thus, in all considered scenarios, the penalty term for within-sector variability turns out to be important. This is in line with the results of [Chen et al. \(2020\)](#), extends them to a dynamic framework, and can be explained and motivated by the economic similarities of companies belonging to the same industry.

The remainder of this paper is organized as follows. The next section presents the model and the algorithm to estimate portfolio weights. Section [3.3](#) applies our methodology to stock returns of the S&P Total index, and Section [3.4](#) concludes. Some complementary results related to the empirical study are collected in Appendix [3.A](#).

3.2 The portfolio selection methodology

In this section we develop our portfolio selection methodology. First, some useful notation is introduced. Then, the DCC-GARCH model used for modelling volatilities and correlations is presented, including the way how monthly forecasts are generated from averaging daily forecasts. We then expose competing weight regularizations for the construction of global minimum variance portfolios, and finally present an algorithm for efficiently calibrating the penalization parameters using dynamic cross-validation.

3.2.1 Notation

We adopt the following notation throughout the paper. The subscript $i \in \{1, \dots, N\}$ indexes the assets in the portfolio, where N denotes the total number of assets available; the subscript $k \in \{1, \dots, K\}$ indexes the Global Industry Classification Standard (GICS) sectors, where K is the total number of sectors; and the subscript $t \in \{1, \dots, T\}$ is the time index, where T is the sample size. Furthermore, we use the following notation:

- $y_{i,t}$: return for asset i at date t , stacked into $y_t := (y_{1,t}, \dots, y_{N,t})'$.
- $\mathcal{F}_{t-1}$: information set generated by $\{y_{t-1}, y_{t-2}, \dots\}$.
- $\sigma_{i,t}^2 := \text{Var}(y_{i,t} | \mathcal{F}_{t-1})$: conditional variance of the i -th asset at date t .
- $\varepsilon_{i,t} := y_{i,t} / \sigma_{i,t}$: devolitized series at date t , stacked into $\varepsilon_t := (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})'$.
- D_t : N -dimensional diagonal matrix with the i -th diagonal element being $\sigma_{i,t}$.
- $R_t := \text{Corr}(y_t | \mathcal{F}_{t-1}) = \text{Cov}(\varepsilon_t | \mathcal{F}_{t-1})$: conditional correlation matrix at date t .
- $H_t := \text{Cov}(y_t | \mathcal{F}_{t-1})$: conditional covariance matrix at date t and $\text{Diag}(H_t) = D_t^2$.
- $C := \mathbb{E}(R_t) = \text{Corr}(y_t) = \text{Cov}(\varepsilon_t)$: unconditional correlation matrix.

3.2.2 Dynamic conditional covariance matrix estimation

In our empirical application, we use daily return data to forecast covariance matrices but we adopt the common practice of monthly re-balancing for the portfolio construction. This creates a mismatch between the frequency used for estimation and for forecasting, which we address

with the average-forecasting approach used in [De Nard et al. \(2021\)](#) for DCC-GARCH model forecasting. Based on daily covariance estimates, daily forecasts are generated, and then they are iterated to deliver predictions for the horizon of interest. Specifically, at any portfolio construction date h , we retrieve the forecasts of the covariance matrices for all days in the upcoming month, namely $t = h, h + 1, \dots, h + L - 1$, then average those L daily forecasts and use this averaged forecast to construct the portfolio composition at date h .

For the dynamics of the univariate volatilities, we choose a GARCH(1,1) process:

$$\sigma_{i,t}^2 = \omega_i + \delta_{1,i} y_{i,t-1}^2 + \delta_{2,i} \sigma_{i,t-1}^2, \quad (3.1)$$

where $(\omega_i, \delta_{1,i}, \delta_{2,i})$ are the parameters for asset i . For the evolution of the conditional correlation matrix over time, we assume that it is governed by a DCC(1,1) model:

$$Q_t = (1 - \alpha - \beta) C + \alpha \varepsilon_{t-1} \varepsilon_{t-1}' + \beta Q_{t-1}, \quad (3.2)$$

where (α, β) are the DCC model parameters, and C is the unconditional correlation matrix of ε_t which can be estimated in high dimensions using shrinkage techniques, see our discussion below. The matrix Q_t here can be interpreted as a conditional pseudo-correlation matrix or a conditional covariance matrix of devolatilized residuals. For the reason that its diagonal elements, although close to one, are not exactly equal to one, we obtain the conditional correlation matrix and the conditional covariance matrix as

$$R_t := \text{Diag}(Q_t)^{-1/2} Q_t \text{Diag}(Q_t)^{-1/2}, \quad (3.3)$$

$$H_t := D_t R_t D_t, \quad (3.4)$$

and for estimation by quasi maximum likelihood, we assume a conditional normal distribution, i.e. $y_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t)$.

We use approximate correlation targeting by first estimating C using shrinkage of the sample correlation matrix of ε_t , and then maximize the likelihood with respect to α and β . Due to the inconsistency argument of [Aielli \(2013\)](#) this is not an exact correlation targeting, which would require a modification of the DCC model. However, results are typically very similar, so for simplicity we stick to the original version of the DCC model.

With respect to the shrinkage method for C , we adopt the nonlinear shrinkage estimator of [Ledoit and Wolf \(2017\)](#) and [Ledoit and Wolf \(2020\)](#). This estimator is particularly appropriate to handle the problem of ill-conditioned sample correlation matrices arising in high dimensions,

see also Zhao et al. (2020). Furthermore, the composite likelihood method developed by Pakel et al. (2020) provides a way to overcome the computational hurdle associated with the estimation of the DCC model in high dimensions. Engle et al. (2019) provide empirical evidence that a DCC model combining nonlinear shrinkage technique with an estimation carried out via composite likelihood method - called DCC-NL - has superior out-of-sample performance, compared to a DCC model with linear shrinkage, in a Markowitz portfolio setting with a large number of assets ($N \geq 100$). Note that the DCC-NL model, by regularizing the sample covariance matrix, also implicitly stabilizes portfolio weights. We will investigate in the empirical section whether additional explicit sector-wise penalties are further stabilizing the weights and improving the performance of the DCC-NL Markowitz portfolio.

To determine the average of the L forecasts of the conditional covariance matrix $H_{h+l} = D_{h+l}R_{h+l}D_{h+l}$, for $l = 0, 1, \dots, L-1$, we propose a three-step approach where D_{h+l} and R_{h+l} are predicted separately, as described in the following.

For the multi-step ahead forecast of conditional univariate volatilities, we follow the approach of Baillie and Bollerslev (1992). The forecasts that minimize the mean square prediction error, for given parameters, are determined by the conditional expectations of conditional variances. Let us denote the l -step ahead forecast by $\hat{\sigma}_{i,h}^2(l) := \mathbb{E} \left[\sigma_{i,h+l}^2 | \mathcal{F}_{h-1} \right]$, which for the GARCH(1,1) case can be written as

$$\hat{\sigma}_{i,h}^2(l) = \sum_{j=0}^{l-1} \omega_i (\delta_{1,i} + \delta_{2,i})^j + (\delta_{1,i} + \delta_{2,i})^l \sigma_{i,h}^2. \quad (3.5)$$

The forecasts of the diagonal matrix D_{h+l} can be constructed as $\hat{D}_h(l) = \text{Diag} (\hat{\sigma}_{1,h}, \dots, \hat{\sigma}_{N,h})$.

Due to the nonlinearity of the DCC model, there is no exact analytical expression for the conditional expectations $\mathbb{E} [R_{h+l} | \mathcal{F}_{h-1}]$. However, Monte Carlo simulation evidence of Engle and Sheppard (2001) suggests that the approximation $\mathbb{E} [R_{h+l} | \mathcal{F}_{h-1}] \approx \hat{R}_h(l)$ has a negligible bias, where

$$\hat{R}_h(l) = \sum_{j=0}^{l-1} (1 - \alpha - \beta) \hat{C} (\alpha + \beta)^j + (\alpha + \beta)^l R_h, \quad (3.6)$$

where $\hat{C}$ is a nonlinear shrinkage estimator of the unconditional correlation matrix of ε_t proposed by Ledoit and Wolf (2020).

By using the forecasts $\hat{D}_h(l)$ and $\hat{R}_h(l)$, the forecasts of the conditional covariance matrix using DCC-GARCH can be finally computed as $\hat{H}_h(l) := \hat{D}_h(l) \hat{R}_h(l) \hat{D}_h(l)$, for $l = 0, 1, \dots, L-1$. Thus, to obtain the estimated conditional covariance matrix on portfolio construction day h , we

average over the L forecasts:

$$\hat{H}_h := \frac{1}{L} \sum_{l=0}^{L-1} \hat{H}_h(l). \quad (3.7)$$

This matrix $\hat{H}_h$ will then be used for portfolio selection.

3.2.3 Global minimum-variance portfolio

We follow a large part of the empirical literature and consider the problem of estimating the global minimum variance (GMV) portfolio. Several reasons motivate our choice. First, it is more difficult to estimate means accurately than the covariance matrix of asset returns and the errors in estimating means have a larger impact on portfolio weights than the errors in the estimates of covariance matrices. Furthermore, as demonstrated by extensive empirical evidence, e.g. [Jagannathan and Ma \(2003\)](#), GMV portfolios have shown very good out-of-sample performance even in terms of criteria that take the mean into account, such as the Sharpe ratio.

The standard GMV problem for a given covariance matrix H_t is formulated as

$$\begin{aligned} \min_w & w' H_t w \\ \text{s.t. } & w' \mathbf{1} = 1, \end{aligned} \quad (3.8)$$

where $\mathbf{1}$ denotes a vector of ones with dimension $N \times 1$. It has the analytical solution

$$w_t = \frac{H_t^{-1} \mathbf{1}}{\mathbf{1}' H_t^{-1} \mathbf{1}}. \quad (3.9)$$

The natural strategy in practice is to replace the unknown H_t in Equation (3.9) by an estimator $\hat{H}_t$, yielding a feasible portfolio allocation strategy

$$\hat{w}_t = \frac{\hat{H}_t^{-1} \mathbf{1}}{\mathbf{1}' \hat{H}_t^{-1} \mathbf{1}}. \quad (3.10)$$

In order to take a sector structure into account, we modify the optimization problem above to the following

$$\begin{aligned} \min_{w \in \mathbf{R}^N} & w' H_t w + P_{\lambda t}(w) \\ \text{s.t. } & w' \mathbf{1} = 1, \end{aligned} \quad (3.11)$$

where $P_{\lambda t}(w)$ is a penalty function, to be specified in the following. The penalty function is parameterized by a parameter $\lambda \geq 0$, possibly a vector, that determines the strength of the imposed penalty. It will also depend on time, so that t is added as a subscript of the penalty function. The optimization problem reduces to the standard GMV problem if $\lambda = 0$, with

analytical solution given in Equation (3.9). For $\lambda > 0$, Equation (3.11) does not have an analytical solution in general, but can typically be written as a convex programming problem for which efficient numerical algorithms are available. In our numerical implementations, we use the R package CVXR for disciplined convex programming, see Fu et al. (2020).

We denote the K industry sectors by $G_1, G_2, \dots, G_K$, and by p_k the number of stocks in the k -th sector G_k , with average weight $m_{kt} = \sum_{i \in G_k} w_{it}/p_k$. In this paper, we conduct a comparison of the following candidates for the penalty function $P(w)$.

1. WITHIN (W): penalize weight differences of assets within the same sector

$$P_{\lambda t}(w) = \lambda \sum_{k=1}^K \sum_{i \in G_k} |w_{it} - m_{kt}|. \quad (3.12)$$

2. BETWEEN (B): penalize weight differences of assets between sectors

$$P_{\lambda t}(w) = \lambda \sum_{k=1}^K |m_{kt} - \frac{1}{N}|. \quad (3.13)$$

3. GROUP LASSO (G):

$$P_{\lambda t}(w) = \lambda \sum_{k=1}^K |m_{kt}|. \quad (3.14)$$

4. SPARSE (S):

$$P_{\lambda t}(w) = \lambda \sum_{i=1}^N |w_{it}|. \quad (3.15)$$

5. COST (C):

$$P_{\lambda h}(w) = \lambda \sum_{i=1}^N |w_{ih} - w_{i,h-1}^+|, \quad h \geq 2, \quad (3.16)$$

where $w_{i,h}^+ := w_{i,h} y_{i,h} / (1 + w_h' y_h)$ with $y_{i,h}$ being the return of asset i between two consecutive re-balancing dates $h-1$ and h .

The first two penalty functions encourage weight similarities of assets belonging to the same sector (“within”), and across different sectors (“between”), respectively. If one believes that the dynamic properties of stock returns are more homogeneous within a sector than between different sectors, then it would intuitively make sense to obtain better GMV portfolio results with the penalty Equation (3.12) than with the penalty Equation (3.13). This is however an empirical question that we are going to address in the next section. Note that an alternative

specification of the within and between penalties could use an L_2 norm rather than L_1 , or a combination of both as in the elastic net of [Zou and Hastie \(2005\)](#).

The third penalty function displayed in Equation (3.14) promotes sparsity of portfolio weights between sectors via a lasso-type penalty as introduced by [Yuan and Lin \(2006\)](#). The fourth penalty function in Equation (3.15) promotes global sparsity of portfolio weights via a lasso-type penalty as introduced by [Tibshirani \(1996\)](#) in a regression context. It is well known that as the strength of this penalty increases, the solution converges to the short-sale constrained portfolio, see [DeMiguel et al. \(2009\)](#). This term thus controls the overall degree of allowed short sales.

Finally, the penalty function in Equation (3.16) promotes a reduction of variability of portfolio weights from one period to the next, and hence intends to reduce transaction costs, similar to [Chen et al. \(2020\)](#) and [Hautsch and Voigt \(2019\)](#). The vector w_{h-1}^+ is the allocation of portfolio weights right before re-balancing at time $h - 1$. As the strength of this penalty increases, the optimal portfolio allocation, i.e. the number of shares per asset, would be decided at the beginning of the evaluation period. The portfolio weights would vary during the evaluation period only because of changing asset prices. We note that the transaction costs are not explicitly taken into account in the portfolio construction and performance evaluation, we instead control for the weight variation over time and use it as an indicator on the relative performance in terms of transaction costs for different portfolios.

We can combine the above penalties in various forms. For example, if we want to penalize both the weights differences between and within sectors, we construct

$$P_{\lambda t}(w) = \lambda_1 \sum_{k=1}^K |m_{kt} - \frac{1}{N}| + \lambda_2 \sum_{k=1}^K \sum_{i \in G_k} |w_{it} - m_{kt}|. \quad (3.17)$$

where now the penalty parameter is a vector of two components, $\lambda = (\lambda_1, \lambda_2)'$. We will call this penalty “BW” for between- and within-sector regularization. The BW portfolio shrinks the asset allocation towards the equally-weighted portfolio, which has often been chosen as a benchmark for investment strategies and was shown in [DeMiguel et al. \(2009\)](#) to perform well out of sample. Analogously, we call SW and CW penalties that combine the within term with sparsity and transaction cost control, respectively. Furthermore, combining the sparse and group lasso penalties, we obtain the sparse group lasso regularization of [Friedman et al. \(2010\)](#). See also [Babii et al. \(2020\)](#) and [Babii et al. \(2021\)](#) who provide theoretical results for the sparse

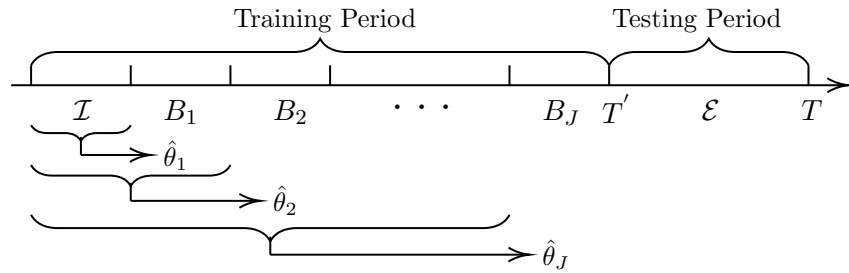
group lasso in a time series framework.

We first investigate individually how the above regularization terms perform compared to the DCC-GARCH estimation without penalty. We will then study whether all three sector-wise penalization terms are equally important and whether combining sector-wise penalties with global controls for sparsity and cost efficiency further improves the performance.

3.2.4 The dynamic portfolio selection algorithm

The previous section presented the objective function for portfolio optimization including various penalty terms depending on the sector structure. The problem is how to choose the parameters λ that determine the strength of the corresponding penalties in a dynamic framework. In the following we propose an algorithm that allows to combine estimation of the dynamic conditional covariance matrix with a cross-validation method to select the penalty parameter. We call θ the vector of all unknown parameters contained in the DCC-GARCH model.

1. Divide the data into a training set $\mathcal{T} = [1, T']$ and a evaluation set, $\mathcal{E} = (T', T]$.
2. Divide $\mathcal{T}$ into an initialization period $\mathcal{I}$ and J blocks B_j : $\mathcal{T} = \mathcal{I} \cup \bigcup_{j=1}^J B_j$. The length of each block is m .
3. Estimate a DCC-GARCH model on $\mathcal{I}$ to get $\hat{\theta}_1$ and the conditional covariance matrix forecasts of $H_t(\hat{\theta}_1)$, on $\mathcal{I} \cup B_1$ to get $\hat{\theta}_2$ and $H_t(\hat{\theta}_2)$, etc. until $\mathcal{I} \cup \bigcup_{j=1}^{J-1} B_j$ to get $\hat{\theta}_J$ and $H_t(\hat{\theta}_J)$.



4. Fix a $\lambda > 0$ and calculate the optimal portfolio allocation for each re-balancing date t in B_1 as

$$w_t(\lambda) = \arg \min_w (w' H_t(\hat{\theta}_1) w + P_{\lambda t}(w)), \quad t \in B_1, \quad (3.18)$$

with $P_{\lambda}(w)$ defined for different cases from Equation (3.12) to Equation (3.17), and the

portfolio variance over B_1 is

$$Q_1(\lambda) = \frac{1}{m} \sum_{t \in B_1} (w_t(\lambda)' y_t)^2. \quad (3.19)$$

5. Repeat step 5 for $j = 2, \dots, J$ to get $Q_2(\lambda), \dots, Q_J(\lambda)$, and calculate $Q(\lambda) = \frac{1}{J} \sum_{j=1}^J Q_j(\lambda)$.

6. Repeat steps 5 and 6 for different values of λ on a grid, and calculate

$$\lambda^* = \arg \min_{\lambda} Q(\lambda). \quad (3.20)$$

7. Estimate a DCC-GARCH model on the training set to get $\hat{\theta}$ and the conditional covariance matrix forecasts of $H_t(\hat{\theta})$.

8. For $t \in \mathcal{E}$, using the optimal penalty parameter λ^* of step 6 and the forecasts $H_t(\hat{\theta})$ of step 7, calculate the optimal portfolio allocation as

$$w_t = \arg \min_w (w' H_t(\hat{\theta}) w + P_{\lambda^* t}(w)), \quad t \in \mathcal{E}. \quad (3.21)$$

9. Calculate the out-of-sample portfolio variance

$$V = \frac{1}{T - T'} \sum_{t \in \mathcal{E}} (w_t' y_t)^2. \quad (3.22)$$

10. Compare V with benchmarks such as the classical GMV without sector regularization ($\lambda = 0$), and one that imposes weight equality across assets ($1/N$).

The proposed algorithm nests a block-version of time series cross-validation (steps 4 to 6) into the portfolio selection problem. We use blocks rather than single observations as test sets, because the forecasting targets are variances, not observations themselves. The choice of the block-size m should balance the estimation uncertainty of out-of-sample volatilities $Q_j(\lambda)$ in Equation (3.19) with that of their expectation, $Q(\lambda)$. Note that in this form, the proposed portfolio selection algorithm is novel. In the following empirical analysis we investigate its performance when applied to a large asset data set.

3.3 Empirical application

In our empirical analysis, we investigate the question of how adding a sector-specific regularization structure will impact portfolio performance with respect to benchmark portfolios. We

will examine the out-of-sample performance of Markowitz portfolios based on combining a dynamic conditional correlation model with intra- and inter-sector penalization using historical stock data.

3.3.1 Data description

We use historical data on daily returns for the S&P Total Index component stocks with data available since at least January 03, 2000. The selected stocks belong to 10 different Global Industry Classification Standard (GICS) sectors which have at least 50 companies in the S&P Total Index. Therefore, Communication Services is the only sector in the S&P Total Index which is not covered in the investment universes considered in our analysis. The sample period spans from January 03, 2000 to March 31, 2021, with a total of 5304 observations per stock. For simplicity, we adopt the common convention that 21 consecutive trading days constitute one month. In this manner, the cross-validation period for the penalization terms ranges from January 17, 2007 to July 13, 2016, resulting in a total of 112 months. The out-of-sample period is composed of the remaining 56 months, starting on July 14, 2016 and ending on March 16, 2021.

To be consistent with common practice, all the portfolios are re-balanced monthly to achieve a lower turnover and avoid an unreasonable amount of transaction costs. For simplicity, we assume the portfolio weights are fixed from one day to the next within a month. This will reduce the transaction cost but not eliminate it as the number of shares over time does not remain constant. At any investment date, the conditional covariance matrix is estimated using the 3-step approach for the DCC-GARCH model described in Section 3.2.2. The availability of daily returns ensures a sufficient estimation precision of the DCC-GARCH model at each step of the proposed algorithm. We consider the following portfolio sizes: $N \in \{50, 100, 250, 500\}$, composed respectively of the first 5, 10, 25, and 50 largest market capitalization stocks – evaluated on March 31, 2021 – from the S&P Total Index component companies belonging to the 10 GICS sectors. In our application, each investment universe is nested by the larger ones, e.g. our investment universe with $N = 50$ is a subset of the one with $N = 100$.

3.3.2 Dynamic covariance matrix estimation

We follow the procedure described in Section 3.2.2, where a univariate GARCH(1,1) model is adopted for the conditional variances of asset returns, and a DCC(1,1) model is fitted to obtain time-varying conditional correlation matrices using historical daily returns. To address the challenges arising in large dimensions, we use nonlinear shrinkage for the unconditional correlation matrix targeting and the estimation is carried out via the composite likelihood method. In order to gain insights on the robustness of the parameter estimates over time and on the effect of sample size, we consider five samples of increasing size. Each period starts on 2000-01-04, and ends respectively on 2007-01-12, 2009-07-09, 2011-11-04, 2014-03-12, and 2016-07-12. In this manner, the sample size for estimation starts from roughly 1800 observations and then increases by around 600 observations each period. These periods will also serve as blocks for our cross-validation algorithm introduced in Section 3.2.4, so that after the initialization period (Period 1) with 1764 observations, there are $J = 4$ blocks, each consisting of $m = 588$ observations.

Figure 3.1 summarizes the GARCH estimates via box-plots for the different sub-periods and dimensions. It can be observed that the estimates for all parameters (ω, α, β) are quite stable over time, especially after Period 2 where the effect of adding more data becomes marginal. Although there are some outliers, the estimates of ω and α are close to zero for most of the assets, while the estimates for the persistence coefficient β are close to one, which are typical findings for financial asset return data. Parameter estimates for the DCC model are reported in Table 3.1, where again estimates of α are close to zero and estimates of β are close to 1, indicating high persistence in the dynamic correlations. Furthermore, similar to the volatility models, parameter estimates stabilize as the sample size increases. In addition, all parameter estimates are significantly different from zero using estimated asymptotic standard errors.

After estimation, we have used various diagnostics such as Portmanteau tests for autocorrelation in standardized residuals, as well as their squares and cross-products, which do not yield any evidence against the chosen model specifications. To save space, we do not report these diagnostics here, but keep them available upon request. We therefore continue with the construction of GMV portfolios under alternative types of regularization.

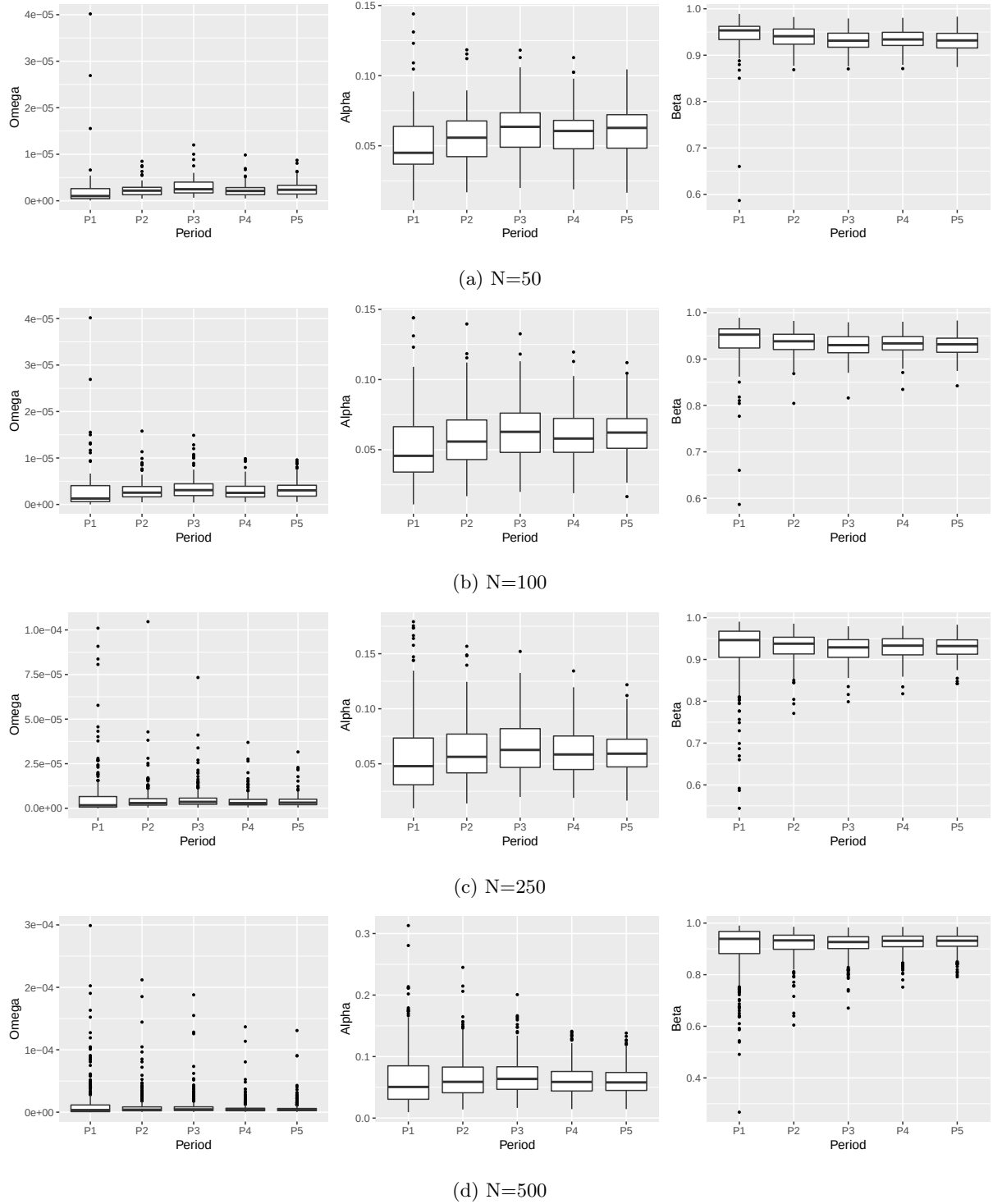


Figure 3.1: Boxplots of GARCH(1,1) parameter estimates for the five periods of augmenting sample sizes (P1 to P5), and for four investment universes $N \in \{50, 100, 250, 500\}$. The sub-periods considered in the estimation are: P1 from 2000-01-04 to 2007-01-12; P2 from 2000-01-04 to 2009-07-09; P3 from 2000-01-04 to 2011-11-04; P4 from 2000-01-04 to 2014-03-12; P5 from 2000-01-04 to 2016-07-12.

Table 3.1: DCC parameter estimates for the five periods of augmenting sample sizes (Period 1 to Period 5), and for four investment universes $N \in \{50, 100, 250, 500\}$.

	$N = 50$		$N = 100$		$N = 250$		$N = 500$	
	α	β	α	β	α	β	α	β
Period 1	0.0126	0.9797	0.0116	0.9815	0.0109	0.9825	0.0095	0.9856
	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0001)	(0.0003)	(0.0004)	(0.0015)
Period 2	0.0145	0.9782	0.0137	0.9803	0.013	0.9821	0.0116	0.9844
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Period 3	0.0196	0.9711	0.0192	0.9725	0.0178	0.9754	0.0156	0.9794
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Period 4	0.018	0.9737	0.0171	0.9762	0.0167	0.9773	0.0149	0.9803
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Period 5	0.0183	0.9735	0.0168	0.9766	0.0167	0.9773	0.0146	0.9807
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

Note: The numbers in parentheses are asymptotic standard errors of the corresponding parameter estimates. The sub-periods considered in the estimation are: Period 1 from 2000-01-04 to 2007-01-12; Period 2 from 2000-01-04 to 2009-07-09; Period 3 from 2000-01-04 to 2011-11-04; Period 4 from 2000-01-04 to 2014-03-12; Period 5 from 2000-01-04 to 2016-07-12.

3.3.3 GMV portfolio performance without short-sale constraints

We study the performance of investment strategies including sector-specific regularization with respect to benchmark portfolios in the context of the GMV portfolio where the most important performance measure is the out-of-sample volatility. Since the GMV portfolio is designed to minimize the variance rather than to maximize the Sharpe Ratio, all portfolios are primarily evaluated by the magnitude of the volatility reduction. We also report the Sharpe ratio to demonstrate that GMV portfolios typically perform well even in terms of criteria that include average out-of-sample returns (see e.g. [Jagannathan and Ma \(2003\)](#)).

We include the following eleven portfolios in our empirical analysis.

1. EQ: the equally-weighted portfolio with weights given by $1/N$, which is a standard benchmark advocated by [DeMiguel et al. \(2009\)](#).

2. NP (no penalty): the DCC-GARCH with nonlinear shrinkage and maximum composite likelihood.
3. W: as NP but with additional penalty for within-sector variation.
4. B: as NP but with additional penalty for between-sector variation.
5. S: as NP but with additional L_1 -norm penalty to promote sparsity.
6. Gl: as NP but with additional L_1 -norm penalty to promote sparsity in the selected sectors.
7. C: as NP but with additional penalty to control transaction costs.
8. BW: as W but with additional penalty for between-sector variation.
9. SW: as W but with additional L_1 -norm penalty to promote sparsity.
10. SGl: as Gl but with additional L_1 -norm penalty to promote sparsity within the selected sectors.
11. CW: as W but with additional penalty to control transaction costs.

In order to obtain the optimal value of the penalization parameter(s) λ , we solve the optimization program specified in Equation (3.20) via grid search. The resulting optimal values of $\log(\lambda)$ are reported in Table 3.6 of Appendix 3.A. We then construct the asset allocation for the eleven portfolios considered on each monthly re-balancing date, and the portfolio weights are assumed to remain constant within the month. The out-of-sample portfolio performance measurements are computed with daily return data and then annualized by following the convention of multiplying the average return by 252 and the standard deviation by $\sqrt{252}$. The (annualized) Sharpe ratio is obtained as the ratio of average return and standard deviation. The statistics, with volatility in percentage points, are presented in Table 3.2. The best performing portfolio with respect to the volatility criterion is marked in bold.

Focusing first on the three portfolios with sector-wise regularizations, namely the W, B, and Gl portfolios, we conclude that the W portfolio has the highest contribution to reducing portfolio volatility. While the combination of within- and between-sector (BW) regularization is further improving the performance of the W portfolio in terms of volatility for the moderate

Table 3.2: Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio.

	W	B	S	GI	C	BW	SW	SGI	CW	NP	1/N
N=50											
SD	17.410**	18.013	16.936***	18.049	16.909***	16.662***	16.933***	16.449***	16.884***	18.191	19.868
SR	0.819	0.908	0.955	0.984	0.954	1.031	0.898	0.973	0.899	0.860	0.729
N=100											
SD	17.231***	18.854	16.800***	18.871	16.833***	17.073***	16.886***	16.806***	16.860***	18.755	20.269
SR	0.703	0.917	1.044	0.884	1.035	0.722	0.795	1.048	0.784	0.872	0.648
N=250											
SD	18.650***	19.579	17.568**	19.634	17.573**	19.060	17.592**	17.572**	17.581**	19.269	20.539
SR	0.644	0.644	0.919	0.680	0.919	0.599	0.898	0.920	0.894	0.675	0.573
N=500											
SD	18.825**	19.657	17.636***	19.604	17.632***	19.464	17.577***	17.635***	17.588***	19.464	21.511
SR	0.777	0.682	0.942	0.682	0.914	0.723	0.900	0.942	0.865	0.699	0.518

Note: Significant outperformance of the portfolios over the NP portfolio in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. All portfolios outperform the equally-weighted benchmark at a significance level of 5% or lower.

investment universes ($N \in \{50, 100\}$), adding additional between-sector penalization is deteriorating the volatility performance of the W portfolio in larger dimensions (i.e. $N \in \{250, 500\}$). Additionally, the observation that shrinking different sector weights towards equivalence lacks in efficacy is consistent with the intuition that assets belonging to different sectors may have quite different characteristics and tend to be less correlated. Promoting within-sector diversification is more important than between-sector diversification or concentrating the allocation on a few sectors. Furthermore, promoting sparsity concentrates the weight allocation on fewer assets, and adding a cost penalization term stabilizes the weight evolution of each asset over time. We can observe from Table 3.2 that both approaches contribute to better portfolio performance in terms of volatility.

In addition, for each of the portfolios with regularizations, we conduct the test of equality in standard deviations with respect to the two benchmark portfolios, with a two-sided p-value obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011). It can be concluded that the nine regularized portfolios outperform the equally-weighted benchmark and the outperformance is statistically significant at all conventional levels. Furthermore, controlling

for “within”, “sparse” and “cost” significantly improves the portfolio volatility (at the 5% level) when considering the NP portfolio as benchmark. We also test whether the regularized portfolios are significantly outperforming each other. The p-values are reported in [Table 3.8](#) and [Table 3.9](#) of [Appendix 3.A](#). We observe that the “within” portfolio is not significantly outperformed by other penalized portfolios at the 5% level for $N \in \{100, 250\}$ and at the 1% level for $N = 500$. However, the “between” and “group sparse” portfolios are almost always significantly outperformed for $N \in \{100, 250, 500\}$ at the 1% level. This confirms that promoting within-sector diversification is more beneficial than between-sector diversification or concentrating the allocation on a few sectors. In addition, portfolios controlling for sparsity and transaction cost are rarely significantly outperformed, especially in moderate and large investment universes.

In terms of the Sharpe ratio, we conclude that adding a sparsity regularization and a cost penalty term contributes to an improved portfolio performance. Additionally, the “between” regularized portfolio (B) and the portfolio focusing on a few selected sectors (G1) display an improvement in performance relative to both benchmarks in the small investment universes ($N \in \{50, 100\}$). The outperformance in terms of Sharpe ratio tends to diminish for portfolios with a larger number of assets ($N \in \{250, 500\}$). We note that if we changed our performance target from GMV portfolio to Sharpe ratio maximization, including additional features such as a momentum signal for expected returns could further improve the performance, see e.g. [De Nard et al. \(2021\)](#).

We also provide some descriptive statistics for the portfolio weights w_t over time. In each holding period, namely one month, we compute the following four characteristics:

- Min: Minimum weight of all assets in the portfolio.
- Max: Maximum weight of all assets in the portfolio.
- SD: Standard deviation of weights of all assets in the portfolio.
- MAD-EW: Mean absolute deviation from equal weights, $1/N$.
- MDiv-Sec-W: Mean diversification within sectors, computed as

$$\text{MDiv-Sec-W} = \frac{1}{T - T'} \sum_{t=1}^{T-T'} \sum_{k=1}^K \sum_{i \in G_k} |w_{it} - m_{kt}|.$$

- MDiv-Sec-B: Mean diversification between sectors, computed as

$$\text{MDiv-Sec-B} = \frac{1}{T - T'} \sum_{t=1}^{T-T'} \sum_{k=1}^K |m_{kt} - \frac{1}{N}|.$$

- No. Active: Number of stocks with absolute weight larger than $1\% * 1/N$.

For each characteristic, we then report the average statistics over the out-of-sample period, as summarized in [Table 3.3](#). We observe that the portfolios with the least dispersed weights are those including within-sector regularization, either alone or in combination with one of the other regularization terms. For each dimension, portfolios with “within”, “sparse” and “cost” penalizations contribute the most to reducing the distance to the equally-weighted portfolio as measured by the MAD-EW criterion. Additionally, we observe that portfolios with significantly improved volatilities (i.e. W, S, C) have overall smaller values for the MDiv-Sec-W statistic indicating higher within-sector diversification, but larger values for the MDiv-Sec-B statistic suggesting lower between-sector diversification. Conversely, portfolios with poor out-of-sample performance (i.e. B, Gl) exhibit higher between-sector diversification and lower within-sector diversification. This observation is consistent with the conclusion that promoting within-sector diversification is more important than between-sector diversification or concentrating the allocation on only a few sectors. Furthermore, the number of active positions indicates that the portfolios with sector-wise regularizations (i.e. W, B, Gl) are diversified across all available assets while portfolios controlling for sparsity and cost are concentrating their weights on 40% of the investment universe for the moderate portfolio sizes ($N \in \{50, 100\}$) and on 20% of the investment universe for larger portfolio sizes ($N \in \{250, 500\}$).

We also present the weight evolution during the out-of-sample period as weight variations provide us with further indications on the relative performance of the different portfolios in terms of transaction costs. [Figure 3.2](#) illustrates the weight variation over time for component stocks in the case where the investment universe is composed of 50 assets. The asset with the highest sector-wise market capitalization is chosen as a representative of each sector, and in such a way we obtain ten stocks representing the ten sectors in the portfolio. From [Figure 3.2](#), it can be concluded that the DCC-GARCH portfolio (NP) is subject to large changes from one rebalancing date to another, which further deteriorates the portfolio return when transaction costs are taken into account. The black curve indicates that adding the within-sector penalization structure is stabilizing the weights across time, resulting in a further improved relative portfolio

performance – compared to the benchmark NP portfolio – when accounting for transaction costs. Additionally, it can be observed from the blue curve that including another regularization term based on transaction costs is further stabilizing the weight evolution over time.

3.3.4 GMV portfolio performance with short-sale constraints

Given that some markets and fund managers are explicitly subject to short-sale restrictions in their activities, we also consider the added benefits of our approach when a non-negativity constraint is applied to all portfolio weights. We report the values of the optimal penalization parameters λ for all cases in Table 3.7 of Appendix 3.A.

Comparing the GMV portfolio results in Table 3.2 with the no-short-sale portfolio results in Table 3.4, we observe that the optimal no-short-sale portfolios have smaller out-of-sample volatility than the GMV portfolios for all considered investment universes. This finding suggests that the imposed no-short-sale constraint helps decreasing the actual portfolio risk, but further improvements can still be obtained. In addition, our previous conclusion that including within-sector regularization is strongly beneficial still holds, especially in small and moderate investment universes. Whereas adding between-sector shrinkage effectively improves performance for small to moderate dimensions, it tends to deteriorate overall portfolio performance in large dimensions ($N = 500$), although the differences are small.

Furthermore, although within-sector penalization reduces portfolio volatility, this improvement in performance is only statistically significant in the case of $N = 50$. The outperformance in terms of volatility is not statistically significant for larger portfolios ($N \in \{100, 250, 500\}$).

Finally, we take a closer look at the portfolio construction by reporting the weights statistics in Table 3.5. Clearly, there is now a stronger effect of the penalization term, which is shrinking the weights towards equality, especially in larger portfolios. Moreover, the W portfolio has higher within-sector diversification while the between-sector diversification increases for larger portfolios ($N \in \{250, 500\}$) and decreases for moderate portfolio sizes ($N \in \{50, 100\}$). Additionally, the number of active positions decreases significantly compared to the corresponding portfolios without short-sale constraint, and this effect gets stronger for larger investment universes.

Table 3.3: Statistics of portfolio weights.

	W	B	S	GI	C	BW	SW	SGI	CW	NP	1/N
N=50											
Min	-0.070	-0.084	0.000	-0.078	-0.001	-0.073	0.000	0.002	-0.001	-0.074	0.020
Max	0.212	0.212	0.332	0.232	0.329	0.191	0.318	0.277	0.310	0.235	0.020
SD	0.049	0.065	0.057	0.066	0.057	0.046	0.053	0.042	0.052	0.066	0.000
MAD-EW	0.034	0.050	0.032	0.050	0.032	0.031	0.028	0.020	0.027	0.050	0.000
MDiv-Sec-W	1.060	2.475	1.352	2.407	1.356	1.237	1.085	0.937	1.075	2.317	0.000
MDiv-Sec-B	0.265	0.124	0.193	0.158	0.193	0.203	0.200	0.122	0.201	0.191	0.000
No. Active	50	50	20	50	27	50	27	50	32	50	50
N=100											
Min	-0.046	-0.062	0.000	-0.062	-0.001	-0.045	0.000	0.000	-0.001	-0.061	0.010
Max	0.200	0.163	0.306	0.170	0.305	0.188	0.333	0.307	0.329	0.174	0.010
SD	0.033	0.041	0.038	0.041	0.039	0.033	0.039	0.038	0.039	0.041	0.000
MAD-EW	0.020	0.030	0.017	0.030	0.017	0.020	0.016	0.017	0.016	0.030	0.000
MDiv-Sec-W	1.752	2.943	1.566	2.915	1.610	1.808	1.347	1.584	1.356	2.887	0.000
MDiv-Sec-B	0.124	0.063	0.087	0.072	0.087	0.113	0.103	0.087	0.104	0.082	0.000
No. Active	100	100	40	100	25	99	55	29	46	100	100
N=250											
Min	-0.030	-0.028	0.000	-0.028	-0.001	-0.029	0.000	0.000	-0.001	-0.028	0.004
Max	0.126	0.127	0.305	0.132	0.302	0.103	0.291	0.305	0.290	0.140	0.004
SD	0.018	0.019	0.024	0.019	0.024	0.017	0.024	0.024	0.024	0.019	0.000
MAD-EW	0.011	0.013	0.007	0.013	0.007	0.012	0.007	0.007	0.007	0.013	0.000
MDiv-Sec-W	2.831	3.229	1.779	3.230	1.808	2.914	1.779	1.779	1.795	3.208	0.000
MDiv-Sec-B	0.023	0.015	0.038	0.015	0.038	0.013	0.038	0.038	0.038	0.019	0.000
No. Active	249	249	57	249	60	249	57	70	59	249	250
N=500											
Min	-0.014	-0.016	0.000	-0.016	-0.000	-0.014	0.000	0.000	-0.000	-0.016	0.002
Max	0.067	0.072	0.238	0.072	0.235	0.058	0.232	0.238	0.229	0.078	0.002
SD	0.009	0.009	0.015	0.009	0.015	0.009	0.015	0.015	0.015	0.010	0.000
MAD-EW	0.006	0.006	0.004	0.006	0.004	0.006	0.004	0.004	0.004	0.006	0.000
MDiv-Sec-W	2.866	3.137	1.845	3.157	1.870	2.936	1.847	1.851	1.862	3.129	0.000
MDiv-Sec-B	0.009	0.007	0.017	0.007	0.017	0.006	0.017	0.017	0.017	0.009	0.000
No. Active	499	499	111	498	111	498	73	99	104	498	500

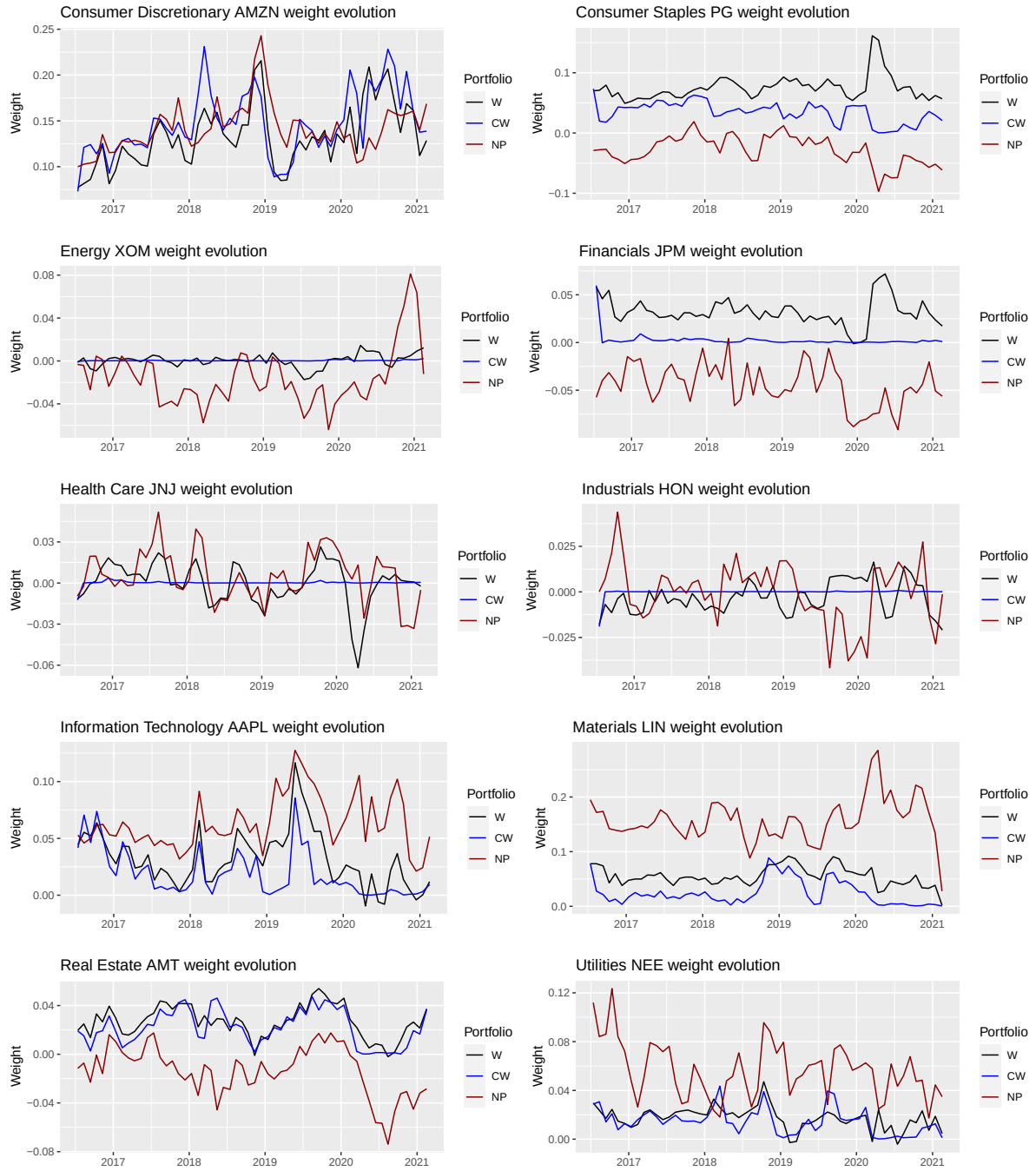


Figure 3.2: Out-of-sample period weight evolution.

Table 3.4: Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio (with short-sale constraint).

	W	B	C	BW	CW	NP	1/N
N=50							
SD	16.370**	16.401**	16.910**	16.472**	16.356**	16.931	19.868
SR	0.866	1.109	0.959	1.014	0.873	0.953	0.729
N=100							
SD	16.589	16.681*	16.793	16.669	16.602	16.798	20.269
SR	0.781	1.093	1.050	0.809	0.781	1.046	0.648
N=250							
SD	16.381	17.565	17.582	16.378	16.403	17.554	20.539
SR	0.598	0.915	0.905	0.596	0.601	0.914	0.573
N=500							
SD	17.515	17.648	17.610*	17.605	17.501	17.625	21.511
SR	0.828	0.949	0.940	0.838	0.811	0.946	0.518

Note: Significant outperformance of the portfolios over the NP portfolio in terms of SD is denoted by asterisks: *, **, and *** indicate significance at the 10%, 5%, and 1% level respectively. The outperformances of all portfolios against the equally-weighted benchmark are significant at a level of 5% or lower.

Table 3.5: Statistics of portfolio weights (with short-sale constraint).

	W	B	C	BW	CW	NP	1/N
N=50							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.020
Max	0.206	0.299	0.330	0.206	0.207	0.331	0.020
SD	0.040	0.054	0.057	0.038	0.040	0.057	0.000
MAD-EW	0.025	0.030	0.031	0.024	0.025	0.031	0.000
MDiv-Sec-W	0.496	1.343	1.340	0.586	0.494	1.345	0.000
MDiv-Sec-B	0.214	0.184	0.192	0.204	0.213	0.193	0.000
No. Active	33	28	25	39	36	23	50
N=100							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.010
Max	0.290	0.285	0.306	0.278	0.285	0.306	0.010
SD	0.035	0.037	0.038	0.034	0.034	0.038	0.000
MAD-EW	0.015	0.017	0.017	0.014	0.014	0.017	0.000
MDiv-Sec-W	1.136	1.572	1.569	1.162	1.110	1.565	0.000
MDiv-Sec-B	0.106	0.086	0.087	0.104	0.106	0.087	0.000
No. Active	68	51	43	75	72	51	100
N=250							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.004
Max	0.050	0.303	0.304	0.049	0.053	0.304	0.004
SD	0.009	0.024	0.024	0.009	0.010	0.024	0.000
MAD-EW	0.005	0.007	0.007	0.005	0.005	0.007	0.000
MDiv-Sec-W	0.799	1.762	1.759	0.767	0.872	1.762	0.000
MDiv-Sec-B	0.042	0.038	0.038	0.042	0.042	0.038	0.000
No. Active	141	111	113	147	144	117	250
N=500							
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.002
Max	0.223	0.190	0.237	0.183	0.222	0.238	0.002
SD	0.015	0.013	0.015	0.013	0.015	0.015	0.000
MAD-EW	0.004	0.004	0.004	0.004	0.004	0.004	0.000
MDiv-Sec-W	1.829	1.858	1.829	1.852	1.819	1.834	0.000
MDiv-Sec-B	0.018	0.011	0.017	0.012	0.018	0.017	0.000
No. Active	125	168	188	117	153	187	500

3.4 Conclusions and outlook

In a dynamic framework, considering various regularizations for portfolio weights, we have shown that both sparsity and sector-wise regularizations are important for reducing out-of-sample portfolio volatilities. Among the sector-wise regularizations, we observe that controlling for within-sector weight variations has the largest contribution to reducing out-of-sample portfolio volatility compared to promoting between-sector diversification or concentrating the portfolio weights on a few sectors. In a scenario without short-sale constraints, the best-performing portfolio always includes a control for sparsity that implicitly reduces negative weights and moves the portfolio closer to one with short sales constraints. In large dimensions, the optimal portfolio combines the sparsity regularization with a penalty for within-sector variation of weights. The optimal portfolios significantly outperform two benchmark portfolios. Adding short-sale constraints, the best-performing portfolio always includes a “within” penalty, but the statistical significance disappears when portfolio size is larger.

Our results can be extended in various directions. First, other portfolios than the GMV can be considered such as the tangency portfolio, or factor-risk-parity portfolios as in [Lassance et al. \(2022\)](#). Second, our methodology could be applied to international diversification, where the structure is not only based on industry sectors but also countries. In a similar vein as our paper, it would be interesting to investigate whether, for example, within-country diversification is more relevant than between-country diversification. Moreover, additional classes of assets can be considered such as bonds, foreign exchange, commodities and alternative assets, for which other types of regularizations might be useful, depending on the objectives. We leave these topics for future research.

3.A Additional tables

Table 3.6: Logarithm of penalty tuning parameters, $\log(\lambda)$, via cross validation (without short-sale constraint).

	W	B	S	GI	C	BW	SW	SGI	CW				
N=50	-15.8	-16.6	-9.3	4.9	-8.1	-16.6	-15.8	-9.4	-15.8	-9.3	4.9	-8.2	-15.8
N=100	-15.5	-18.1	-11.4	-11.5	-13.8	-18.2	-15.6	-11.6	-15.6	-11.4	-11.4	-14.0	-15.6
N=250	-17.6	-17.6	-10.9	-5.8	-10.8	-17.6	-17.5	-11.0	-17.5	-11.1	-5.7	-10.9	-17.6
N=500	-17.8	-18.4	-11.8	-4.6	-12.7	-18.5	-17.8	-11.8	-17.9	-11.8	-4.8	-2.0	-17.8

Table 3.7: Logarithm of penalty tuning parameters, $\log(\lambda)$, via cross validation (with short-sale constraint).

	W	B	C	BW		CW	
N=50	-15.0	-16.6	-8.3	-16.7	-15.1	-8.4	-15.0
N=100	-14.9	-18.0	-12.4	-17.9	-15.0	-12.2	-14.9
N=250	-13.6	-20.9	-11.2	-21.0	-13.6	-11.0	-13.7
N=500	-17.0	-17.3	-11.6	-17.4	-17.0	-11.4	-16.9

Table 3.8: P-values for statistical test of difference in volatility (without short-sale constraint).

	W	B	S	Gl	C	BW	SW	SGl	CW	NP	1/N
N=50											
W		0.212		0.261						0.038	0.001
B				0.842						0.314	0.010
S	0.130	0.062		0.088						0.005	0.000
Gl										0.556	0.029
C	0.100	0.049	0.462	0.076			0.788			0.003	0.000
BW	0.014	0.000	0.546	0.002	0.570		0.446		0.508	0.000	0.000
SW	0.082	0.052	0.970	0.078						0.002	0.000
SGl	0.013	0.009	0.056	0.021	0.044	0.638	0.072		0.095	0.001	0.000
CW	0.054	0.042	0.576	0.078	0.765		0.011			0.002	0.000
NP											0.016
N=100											
W		0.000		0.000						0.000	0.000
B				0.806							0.002
S	0.306	0.000		0.000	0.205	0.492	0.466	0.226	0.642	0.000	0.000
Gl											0.005
C	0.334	0.000		0.000		0.539	0.658		0.834	0.000	0.000
BW	0.087	0.000		0.000						0.000	0.000
SW	0.387	0.000		0.000		0.619				0.000	0.000
SGl	0.312	0.000		0.000	0.205	0.501	0.496		0.672	0.000	0.000
CW	0.352	0.000		0.000		0.564	0.602			0.000	0.000
NP		0.333		0.059							0.004

Note: The p-values are from the tests of outperformance where the row portfolios are benchmarked against the column portfolios, i.e. we test whether portfolio i significantly outperforms portfolio j in terms of volatility for row i and column j of the table. An empty cell indicates portfolio i is not outperforming portfolio j .

Table 3.9: P-values for statistical test of difference in volatility (without short-sale constraint).

	W	B	S	Gl	C	BW	SW	SGl	CW	NP	1/N
N=250											
W		0.000		0.000		0.000				0.001	0.000
B				0.390							0.017
S	0.174	0.006		0.004	0.762	0.058	0.889	0.454	0.929	0.021	0.001
Gl											0.035
C	0.181	0.007		0.004		0.061	0.922		0.956	0.023	0.001
BW		0.002		0.004						0.158	0.000
SW	0.139	0.004		0.002		0.040				0.014	0.000
SGl	0.176	0.006		0.004	0.960	0.059	0.908		0.952	0.022	0.001
CW	0.143	0.004		0.002		0.042	0.911			0.014	0.000
NP		0.000		0.004							0.001
N=500											
W		0.002		0.001		0.000				0.011	0.000
B											0.000
S	0.044	0.000		0.001		0.003				0.001	0.000
Gl		0.189									0.000
C	0.043	0.000	0.799	0.000		0.003		0.825		0.001	0.000
BW		0.502		0.560						0.999	0.000
SW	0.025	0.000	0.266	0.000	0.278	0.001		0.270	0.262	0.000	0.000
SGl	0.044	0.000	0.679	0.001		0.003				0.001	0.000
CW	0.026	0.000	0.416	0.000	0.434	0.001		0.422		0.000	0.000
NP		0.000		0.000							0.000

Note: The p-values are from the tests of outperformance where the row portfolios are benchmarked against the column portfolios, i.e. we test whether portfolio i significantly outperforms portfolio j in terms of volatility for row i and column j of the table. An empty cell indicates portfolio i is not outperforming portfolio j .

Table 3.10 reports the total running time for the out-of-sample period of 56 months using the optimal penalization parameter found in the cross-validation step. The running time increases moderately with respect to portfolio size. Note that the optimal weights of the portfolios with W, B, S, Gl penalty terms at time t are independent of the optimal weights of other re-balancing dates and thus the computation can be implemented with parallel programming. However, the optimal allocation at time t of the portfolios controlling for the cost penalty is dependent on the optimal weights of the previous re-balancing date $t - 1$, and thus can only be computed sequentially.

Table 3.10: Running time for the out-of-sample period (in minutes).

	W	B	S	Gl	C	BW	SW	SGl	CW
N=50	0.33	0.05	0.03	0.03	0.13	0.37	0.33	0.04	2.05
N=100	0.32	0.06	0.03	0.03	0.13	0.36	0.33	0.05	2.09
N=250	0.32	0.07	0.05	0.05	0.22	0.38	0.37	0.07	2.17
N=500	0.53	0.19	0.16	0.13	0.65	0.56	0.53	0.22	2.67

Note: Computations were carried out on a 2020 iMac with the following specifications: Intel Core i9 10-Core processor (3.6GHz) and 72 GB RAM (2133 MHz DDR4).

Chapter 4

Dynamic Autoregressive Liquidity (DARLiQ)¹

4.1 Introduction

Liquidity is a fundamental property of a well-functioning market, and lack of liquidity is generally at the heart of many financial crises and disasters. Common ways of measuring liquidity using high frequency data include bid-ask spreads, effective spreads, realized spreads, depth, weighted depth, and transaction volume. There is a big literature that uses such measures to compare market quality across markets, across time, and before and after interventions of various sorts. For example, it has been a big part of the debate around high frequency trading, i.e., whether such trading activity has improved or degraded market liquidity, see e.g. [Brogaard \(2010\)](#), [Hendershott et al. \(2011\)](#), [Beddington et al. \(2012\)](#), [O’Hara and Ye \(2011\)](#). There are many complex issues in working with high frequency trade and quote data in a legally integrated market such as the US, where separate venues exist without synchronized timestamps so that for example establishing the time priority of messages across different venues is difficult. There are several methods widely used to measure liquidity using lower frequency data, i.e., daily data, see [Goyenko et al. \(2009\)](#) for a review of such measures. We focus on the Amihud illiquidity measure as proposed in [Amihud \(2002\)](#). This measure has proven to be very popular in the

¹This chapter corresponds to Hafner, C.M. and Linton, O.B. and L. Wang (2022): “Dynamic Autoregressive Liquidity (DARLiQ)”, *Cambridge Working Papers in Economics CWPE2214*, doi: <https://doi.org/10.17863/CAM.82497>.

empirical literature. It is easy to implement and by all accounts relatively robust. It has been shown to influence the cross-sectional asset returns through the so-called illiquidity premium, see the review of [Amihud and Mendelson \(2015\)](#).

We propose a dynamic semiparametric model for illiquidity as measured by the daily component of the Amihud measure. Specifically, we propose a multiplicative model that contains a nonparametric long-run trend and a parametric short-run autoregressive process as in [Engle et al. \(2012\)](#). The trend part is important for many datasets where liquidity has improved in a secular fashion such as the S&P 500 over the last hundred years and Bitcoin over the much more recent period of its operation. The nonparametric trend is comparable with the conventional monthly averaged measure that is widely used in the literature, except that our measure is available daily and the implicit length of averaging is controlled by a bandwidth parameter to be chosen by the practitioner. Further, the dynamic component of the model measures the short-run variation in liquidity that may be of equal interest.

We approach estimation through GMM based on the first conditional moment restriction, as well as through a semiparametric likelihood procedure that assumes i.i.d. shocks. In the latter approach we consider two cases, one where the shock distribution is parametric such as the Weibull distribution and a further case in which the shock distribution is not specified and is treated nonparametrically. We develop the distribution theory and efficiency bound in both cases.

We also develop methodology for detecting permanent and transitory changes in liquidity that might arise from structural changes in financial markets such as the upgrade of a stock exchange's matching engine or from stock specific events such as stock splits. In our approach, permanent effects are captured by discontinuous changes in the nonparametric trend function, whereas temporary effects are measured by dummy variables in the dynamic part of the process. We develop the inference tools required to test for the null hypothesis of no changes versus the alternatives of permanent or temporary shifts in the illiquidity process.

In the spirit of [Amihud \(2002\)](#), who studies the effect of expected and unexpected illiquidity on stock excess returns, we also consider the regression modelling of the market risk premium driven by the separate components of liquidity from our model. In particular, we study the link between the stock excess returns and the long-run trend, short-run dynamics as well as the unexpected shocks of the illiquidity process.

We implement our framework on a panel composed by the five largest US technology company stocks and the Bitcoin asset. We demonstrate the model performance in terms of fitting the relevant features of the illiquidity data, and provide various model diagnostics and specification tests. We show that the efficient semiparametric maximum likelihood estimator, assuming a parametric Weibull distribution for the error term, captures well the salient features of the illiquidity process. In addition, we also demonstrate that using a nonparametric density estimator for the error term can further improve the model estimation in terms of likelihood.

We study the impact of stock splits on the illiquidity dynamics of the five largest US technology company stocks. One explanation for why companies split their stock is the theory that this creates “wider” markets, that is, reducing the price level makes it easier for a wider pool of retail investors to buy into the stock and allows existing investors more easily to sell part of their holding to other investors thereby increasing the investor base and the volume of transactions. This in turn should lead to greater liquidity as measured for example by the Amihud measure. However, there are other theoretical arguments presented in [Copeland \(1979\)](#) that may point to a decrease in liquidity following a stock split, and as he says “liquidity changes following stock splits is an empirical question”. [Copeland \(1979\)](#) found: nonstationarities in trading behavior, volume increases less than proportionately, brokerage revenues increased, and increases in proportional bid-ask spreads following stock splits. He argues that “these results lead to the conclusion that there is a permanent decrease in liquidity following the split”. Our results broadly support these findings in our more recent sample data on a special subset of stocks, the tech stocks. Specifically, we document that stock splits cause significant shifts in the long-term illiquidity trend while no significant effects on short-run liquidity dynamics are detected.

We also investigate how the different components of the illiquidity process obtained from our model relate to the stock market risk premium using data on the S&P 500 stock market index. We find that the detrended market risk premium is positively affected by the anticipated short run illiquidity process and negatively associated with the unanticipated component of market illiquidity in agreement with the results of [Amihud \(2002\)](#) (which were based on an AR model fit to monthly illiquidity).

The remainder of the paper is organized as follows. In [Section 4.2](#), we discuss the Amihud illiquidity measure and its time series properties. [Section 4.3](#) introduces our DArLiQ model and

we discuss in Section 4.4 estimation via GMM based on the first conditional moment restriction, as well as through a semiparametric likelihood procedure that assumes i.i.d. shocks. The large sample properties of our procedures are provided in Section 4.5. We develop in Section 4.6 the methodology to detect permanent and temporary changes in the liquidity process and in Section 4.7 the framework to study the effect of illiquidity components on risk premium. Section 4.8 presents a detailed empirical application of the model, and Section 4.9 concludes. Theoretical materials including proofs of the theorems are collected in section 4.A to section 4.D. Additional tables and figures for the empirical application are presented in section 4.E.

4.2 Amihud illiquidity

The Amihud (2002) illiquidity measure of a stock at time t , A_t , is defined as

$$A_t = \frac{1}{n_t} \sum_{j=1}^{n_t} \ell_{t_j}, \quad \ell_{t_j} = \frac{|R_{t_j}|}{V_{t_j}}, \quad (4.1)$$

where R_t is the stock return and V_t is the (dollar) trading volume at time t . Intuitively, the Amihud measure captures the fact that a stock is less liquid if a given trading volume generates a larger move in its price. Typically, the measure is computed over periods ranging from a day to a year by averaging the daily illiquidity ratio ℓ_{t_j} over the corresponding period n_t . The Amihud illiquidity measure is a good proxy for high-frequency measures of price impact (Goyenko et al. (2009); Hasbrouck (2009)) with the advantage of only requiring daily data on stock prices and trading volumes.

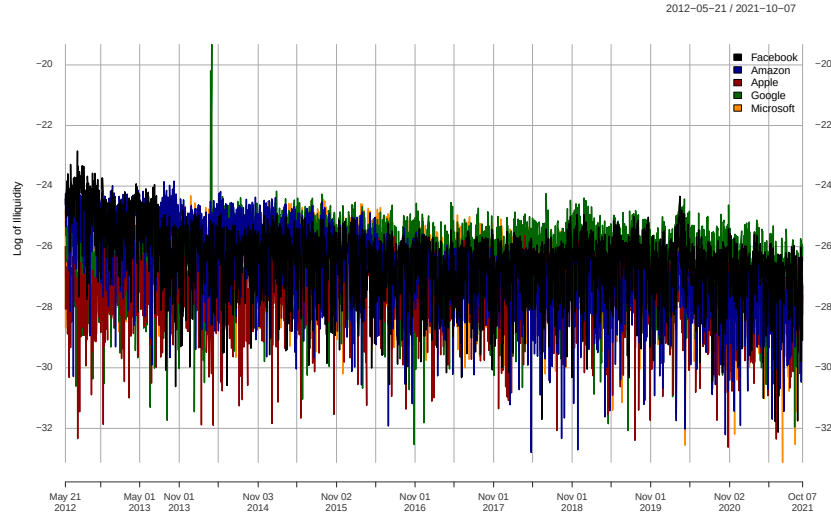
We show in Figure 4.1 the daily stock log illiquidity series for the five largest US information technology companies (the “Fab 5”) – Amazon, Apple, Facebook, Google, and Microsoft – over the period from May 2012 to October 2021. Note that there is spike in the illiquidity series for Google around end-March 2014 which is caused by a stock split on March 27, 2014.² As this event caused irregularity in the trading activities for a few days, we replace the volume data on those dates using the average volume level of the day before and the day after that period. The daily log illiquidity series using the adjusted data are shown in Figure 4.1b. The illiquidity time series appear broadly stationary during this period, although a slight downward trend can be observed. To gain more insights into the conditional dynamics the data, we fit an AR(5) model

²The two-for-one stock split was associated with the introduction of a new non-voting share class (Class C shares). See [press release](#).

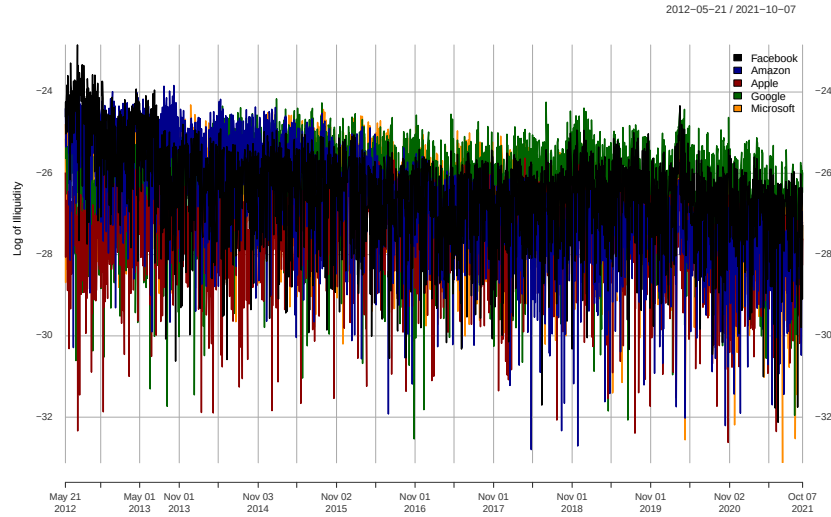
with a quadratic polynomial trend function to the scaled illiquidity series, $y_t = \ell_t \times 10^{10}$, i.e.

$$y_t = \alpha + \beta(t/T) + \gamma(t/T)^2 + \sum_{j=1}^5 \phi_j y_{t-j} + \varepsilon_t,$$

where the coefficients β and γ respectively capture the linear and quadratic components of the polynomial trend. The estimated coefficients with their corresponding t-statistics are provided in Table 4.1. We observe that most of the autoregressive coefficients are statistically significant, indicating some degree of persistence in the stock illiquidity dynamics.



(a) Original data.



(b) Data of Google stock processed.

Figure 4.1: Feb 5 daily log illiquidity $-\log \ell_t$.

Table 4.1: Estimated parameters of an AR(5) with trend.

	Facebook	Amazon	Apple	Google	Microsoft
AR(1)	0.055 (2.683)	0.004 (0.175)	0.025 (1.222)	0.002 (0.095)	0.028 (1.339)
AR(2)	0.196 (9.720)	0.003 (0.136)	0.019 (0.929)	0.102 (4.961)	0.029 (1.398)
AR(3)	0.146 (7.182)	0.088 (4.295)	0.106 (5.218)	0.168 (8.306)	0.025 (1.235)
AR(4)	0.170 (8.426)	0.097 (4.730)	0.027 (1.316)	0.113 (5.506)	0.075 (3.647)
AR(5)	0.122 (5.971)	0.061 (2.953)	0.097 (4.723)	0.063 (3.053)	0.053 (2.592)
Con	0.059 (8.946)	0.131 (16.206)	0.010 (9.969)	0.029 (8.235)	0.072 (15.583)
t/T	-0.169 (-7.263)	-0.297 (-13.433)	0.027 (6.405)	0.049 (3.539)	-0.062 (-5.076)
(t/T) ²	0.131 (6.346)	0.178 (10.402)	-0.031 (-7.482)	-0.060 (-4.411)	-0.001 (-0.118)
Adj. R ²	0.475	0.499	0.086	0.104	0.245

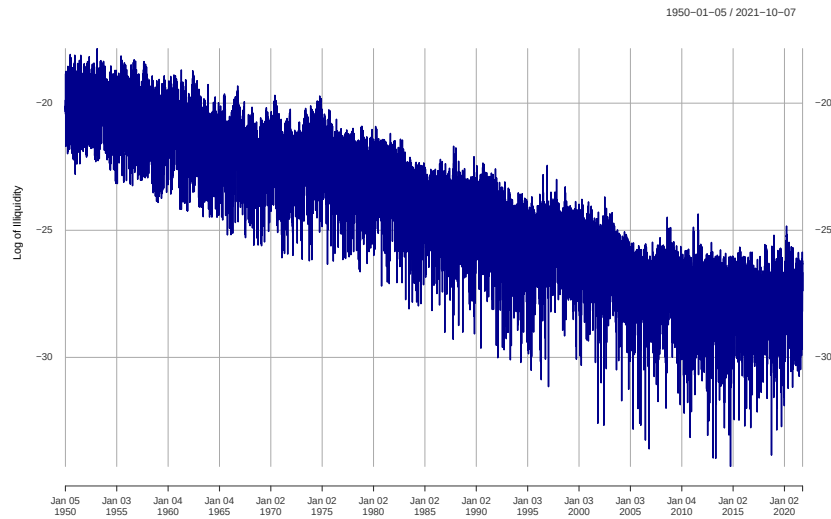
Note: Models are fitted on $y_t = \ell_t \times 10^{10}$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

In addition, the coefficient estimates for the trend function are also significant. One exception is the quadratic term for Microsoft, meaning that this stock exhibits a linear trend over the sample period. Consistent with the visual inspection of [Figure 4.1](#), all the estimated polynomial trend functions are overall downward trending.

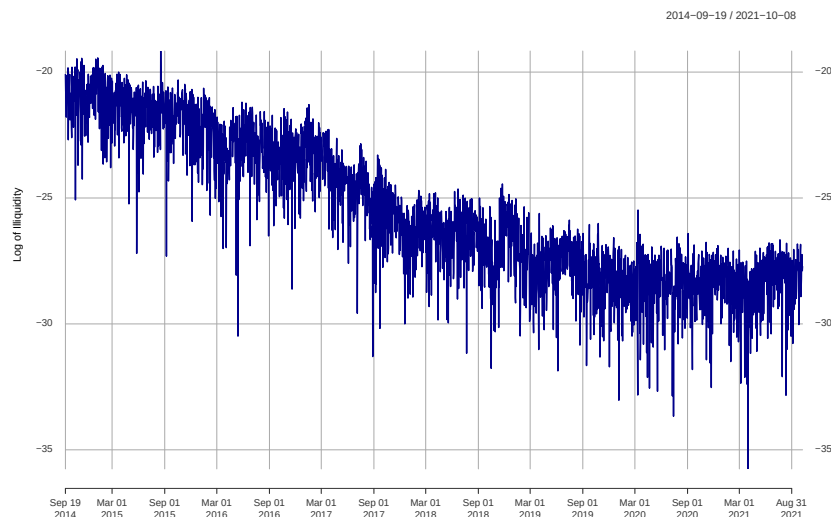
Trends in illiquidity series are not circumscribed to the Fab 5 stocks used in our illustration. To emphasize how prevalent this feature is across financial markets, [Figure 4.2a](#) shows the evolution of $\log \ell_t$ for the S&P 500 stock market index over a longer period (1950–2021) and [Figure 4.2b](#) for Bitcoin-USD (2014–2021) – both exhibiting strong downward trends over their respective sample periods. Taken together, these evidence point to the existence of factors driv-

ing low-frequency variations in illiquidity dynamics in addition to higher-frequency variations.

This motivates our modelling approach for the Amihud illiquidity measure which weakens the requirement on stationarity and develop a new class of dynamic autoregressive liquidity (DARLiQ) models. A key feature of our framework is that it captures both the slow-varying long-term trend and short-run autoregressive components relevant for the modelling of illiquidity series.



(a) S&P 500 index.



(b) Bitcoin asset.

Figure 4.2: S&P 500 index and Bitcoin asset daily log illiquidity – $\log \ell_t$.

4.3 The model

We suppose that ℓ_t , $t = 1, \dots, T$ is an observed non-negative stochastic process. We assume that the observed daily series ℓ_t follows a multiplicative process as in [Engle and Gallo \(2006\)](#).

$$\ell_t = \lambda_t \zeta_t \quad (4.2)$$

$$\lambda_t = \omega + \sum_{j=1}^p \beta_j \lambda_{t-j} + \sum_{k=1}^q \gamma_k \ell_{t-k}, \quad (4.3)$$

where ζ_t is a sequence of positive random variables with conditional mean one and finite unconditional variance. Provided $\omega > 0$, the process $\lambda_t \geq \omega > 0$ with probability one (provided the initialization is also positive). Furthermore, provided $\sum_{j=1}^p \beta_j + \sum_{k=1}^q \gamma_k < 1$, the process ℓ_t is stationary in mean and follows an ARMA(p,q) process. We may further assume that ζ_t has constant conditional variance denoted σ_ζ^2 , in which case $E(\ell_t | \mathcal{F}_{t-1}) = \lambda_t$ and $\text{var}(\ell_t | \mathcal{F}_{t-1}) = \lambda_t^2 \sigma_\zeta^2$, i.e., both z_t and $z_t^2 - 1$ are martingale difference sequences, where $z_t = (\frac{\ell_t}{\lambda_t} - 1)/\lambda_t \sigma_\zeta$. We may further assume that ζ_t is i.i.d. with Lebesgue density function f on the positive real line.

In practice, it may be important to account for nonstationarity or trend. We allow for a nonparametric trend, so let

$$\ell_t = g(t/T) \lambda_t \zeta_t \quad (4.4)$$

$$\lambda_t = \omega + \sum_{j=1}^p \beta_j \lambda_{t-j} + \sum_{k=1}^q \gamma_k \ell_{t-k}^*, \quad (4.5)$$

where $g(\cdot)$ is a smooth but unknown function of rescaled time, and $\ell_t^* = \ell_t / g(t/T)$ is the rescaled liquidity. There is an identification issue because we can multiply and divide the two components g, λ by constants. We suppose that $E(\lambda_t) = 1$, which is achieved by setting $\omega = 1 - \sum_{j=1}^p \beta_j - \sum_{k=1}^q \gamma_k$. The series $\ell_t^* = \lambda_t \zeta_t$ possesses the same stationarity properties as ℓ_t from the model without a trend. Likewise, the error process ζ_t may possess a constant conditional variance or even be i.i.d. with some density function f . This is important for estimation but it may also be important for calculation of “Liquidity at Risk”, which would require some further assumption about the conditional quantiles of ζ_t . Note that in this model, the process ℓ_t actually depends on T and forms a triangular array, $\ell_{t,T}$, but for notational economy we generally suppress this from the notation. We note that the process $\ell_{t,T}$ is locally stationary according to [Vogt \(2012\)](#). In the sequel, we suppose that $p = 1, q = 1$ for simplicity.

We may wish to consider the effects of interventions at some times $t_1, \dots, t_J$. We model

temporary effects by dummy variables in the dynamic equation, that is, we let

$$\lambda_t = \omega + \beta\lambda_{t-1} + \sum_{j=1}^J \alpha_j D_{jt} + \gamma\ell_{t-1}^*,$$

where D_{jt} is one if an intervention occurs in period t_j . The null hypothesis of interest here is $\alpha_1 = \dots = \alpha_J = 0$ in which case the model collapses to Equation (4.5). We may allow the possibility of permanent effects by allowing the function g to be discontinuous at a known point $u_0 = t_0/T$, that is,

$$\lim_{u \uparrow u_0} g(u) = g_-(u_0), \quad \lim_{u \downarrow u_0} g(u) = g_+(u_0)$$

are both well defined but $g_-(u_0)$ may not be equal to $g_+(u_0)$. The size of the jump is the magnitude of the permanent effect (that is, the effect that remains permanently in the absence of further changes). The null hypothesis of interest here is that $g_-(u_0) = g_+(u_0)$ in which case the analysis of the trend is may exploit this property.

4.4 Estimation

Estimation is guided by assumptions made about the error ζ_t . The minimalist approach is to assume only that with probability one

$$E(\zeta_t - 1 | \mathcal{F}_{t-1}) = 0,$$

and $E(\zeta_t^2) = \sigma_\zeta^2 < \infty$. In that case, one can estimate the function $g(\cdot)$ and the identified parameters β, γ by conditional mean smoothing of ℓ_t and by the GMM approach. Provided that additional high level weak dependence conditions are satisfied, one can ensure a CLT for the resulting estimators. One may wish to additionally specify a second conditional moment restriction whereby $E(\zeta_t^2 - (1 + \sigma_\zeta^2) | \mathcal{F}_{t-1}) = 0$ with probability one. This additional moment restriction permits more efficient estimation provided this restriction is true, but if it is not true, using this additional moment restriction will bias the parameter estimates.

We may further assume that ζ_t is i.i.d. with a Lebesgue density f . In this case, we may either assume that f is of unknown functional form or we may assume that f is parametrically specified, i.e., f_φ for some unknown shape parameters φ such as Exponential, Weibull, Gamma etc. The enlarged vector $(\beta, \gamma, \varphi)^\top$ can be estimated by MLE. In the semiparametric case, one needs also to estimate the error density $f(\cdot)$ along with the trend $g(\cdot)$ and the identified parameters. For forecasting future values of ℓ_t , one does not need the shock distribution, but predicting intervals

and LAR (Liquidity at Risk) requires the estimation of some features of the error distribution.

Finally, we note that under the smoothness conditions on $g(\cdot)$,

$$\frac{\ell_{t,T}}{\ell_{t-1,T}} = \left(1 + O(T^{-1})\right) \frac{\lambda_t \zeta_t}{\lambda_{t-1} \zeta_{t-1}}$$

is approximately stationary and one could design a model specification test based on this “differencing” along the lines of [Yatchew \(2000\)](#).

4.4.1 Estimation based on conditional moment restriction

We have the unconditional moment restriction

$$E(\ell_t) = g(t/T).$$

We first use this condition to obtain an initial consistent estimators of g by kernel smoothing method, specifically

$$\hat{g}(u) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \ell_t, \quad (4.6)$$

where K is a kernel function symmetric about zero supported on $[-1, 1]$ satisfying $\int K(u) du = 1$ and $\int K(u) u^2 du < \infty$, while h is a bandwidth sequence. Because of the equally spaced observations in time, the denominator of the Nadaraya-Watson estimator is unnecessary here (this may also be called the Priestley-Chao estimator). Another estimator of historical interest is the Gasser-Müller (1979) estimator for time trend. One could use local linear or local polynomial estimators here to manage the boundary issue or alternatively modify the kernel in the boundary region $[-1, -1 + h] \cup [1 - h, 1]$. One can interpret the widely computed measure A_t defined in Equation (4.1) as a crude estimator of g at the appropriate time point. We define the detrended liquidity $\hat{\ell}_t^* = \ell_t / \hat{g}(t/T)$, $t = 1, \dots, T$.

Second, we use the GMM approach to estimate the parameters from the conditional moment restriction

$$E(\ell_t^* | \mathcal{F}_{t-1}) = \lambda_t,$$

where $\ell_t^* = \ell_t / g(t/T)$, $t = 1, \dots, T$. We work with $e_t(\theta) = \ell_t^* - \lambda_t(\theta)$, where $\theta = (\beta, \gamma)^\top$, which is a martingale difference sequence at the true parameter values $\beta = \beta_0, \gamma = \gamma_0$. In practice, we define $\hat{e}_t(\theta) = \hat{\ell}_t^* - \hat{\lambda}_t(\theta)$ and

$$\hat{\lambda}_t(\theta) = 1 - \beta - \gamma + \beta \lambda_{t-1} + \gamma \hat{\ell}_{t-1}^*.$$

Then we define $\rho_t(\theta, \hat{g}) = z_{t-1}(\hat{\ell}_t^* - \hat{\lambda}_t(\theta))$ and

$$M_T(\theta, \hat{g}) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, \hat{g})$$

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} \|M_T(\theta, \hat{g})\|_W,$$

where W is a weighting matrix, while $z_t \in \mathcal{F}_t$ are instruments. This provides initial consistent estimators of θ . One can optimize the GMM procedure by choosing the instruments and weight matrix optimally, but we shall not pursue this here.

Given consistent estimates of θ , one can improve the estimate of g . Note that

$$E\left(\frac{\ell_t}{\lambda_t}\right) = g(t/T),$$

which provides an alternative local moment condition for estimation, that is, we let

$$\tilde{g}(u) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t}, \quad (4.7)$$

where $\hat{\lambda}_t = \hat{\lambda}_t(\hat{\theta}_{GMM}, \hat{g})$ are estimated in the previous procedure. We may repeat to update the estimates of θ .

The other approach to this is to use profiling, that is, for given θ , we estimate $g_\theta(u)$ by smoothing $\ell_t/\lambda_t(\theta, g_\theta)$ against time and then we optimize the profiled GMM objective function. This approach is more time consuming and we prefer the direct approach here.

4.4.2 Estimation based on i.i.d. assumption

In this case, we assume that the error is i.i.d. with mean one and density f . We consider several cases. First, where f is known completely. Second, where f is known up to a vector of parameters φ . Third, where f is of unknown form. [Drost and Klaassen \(1997\)](#) suppose that ζ_t is a scale random variable with $\zeta_t = \sigma_\zeta z_t + 1$, where z_t is a standardized random shock with mean zero and variance one and density f_0 . The density f_0 may be treated parametrically, i.e., depending on unknown shape parameters φ or it may be treated nonparametrically.

In the case where f is treated parametrically, one may prefer to work directly with ζ_t and parameterize f . In that case, we have a semiparametric model with parameters $\theta = (\beta, \gamma, \varphi^\top)^\top$ and unknown functions $\tau(\cdot) = g(\cdot)$. In the case where f is of unknown form, we have a semiparametric model with parameters $\theta = (\beta, \gamma)^\top$ and unknown functions $\tau(\cdot) = (f(\cdot), g(\cdot))$. In that case, we can equivalently write the model in terms of parameters $\theta = (\beta, \gamma, \sigma_\zeta^2)^\top$ and unknown functions $\tau(\cdot) = (f_0(\cdot), g(\cdot))$.

4.4.2.1 Parametric density case

Suppose that f depends on some unknown parameters φ , denoted as f_φ . If $g(\cdot)$ were known, the log likelihood function is apart from a term to do with $g(\cdot)$ that does not depend on parameters equal to

$$L(\theta, \varphi | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) + \sum_{t=1}^T \ln f_\varphi(\zeta_t(\theta)),$$

$$\zeta_t(\theta) = \frac{\ell_t}{\lambda_t(\theta)g(t/T)}.$$

In practice, given a consistent estimate of $g(\cdot)$, we maximize an estimated version of this likelihood. In fact, in the semiparametric model, the efficient score functions (derived in the appendix) for θ, φ in the presence of unknown $g(\cdot)$ are

$$L_\theta^* = \sum_{t=1}^T \ell_{\theta t}^*, \quad \ell_{\theta t}^* = s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \quad (4.8)$$

$$L_\varphi^* = \sum_{t=1}^T \ell_{\varphi t}^*, \quad \ell_{\varphi t}^* = \frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} - \frac{E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right)}{I_2(f) E \left(\frac{1}{\lambda_t^2} \right)} s_2(\zeta_t) \frac{1}{\lambda_t}. \quad (4.9)$$

To obtain fully efficient estimates of θ, φ , we use one-step updating from initial root-T consistent estimates. Denote $\eta = (\theta, \varphi)$ and $\tilde{\eta} = (\tilde{\theta}^\top, \tilde{\varphi}^\top)^\top$ and let $\ell_{\eta t}^* = (\ell_{\theta t}^{*\top}, \ell_{\varphi t}^{*\top})^\top$, then let

$$\tilde{\eta} = \hat{\eta} + \left(\frac{1}{T} \sum_{t=1}^T \ell_{\eta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) \ell_{\eta t}^{*\top}(\hat{\theta}, \hat{\varphi}, \hat{g}) \right)^{-1} \frac{1}{T} \sum_{t=1}^T \ell_{\eta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) \quad (4.10)$$

$$\ell_{\theta t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) = \hat{s}_2(\hat{\zeta}_t) \frac{1}{\hat{\lambda}_t} \left(\frac{\partial \hat{\lambda}_t}{\partial \theta} - \frac{\frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \right)$$

$$\ell_{\varphi t}^*(\hat{\theta}, \hat{\varphi}, \hat{g}) = \frac{\partial f_{\hat{\varphi}}(\hat{\zeta}_t) / \partial \varphi}{f_{\hat{\varphi}}(\hat{\zeta}_t)} - \frac{\frac{1}{T} \sum_{t=1}^T \frac{\partial f_{\hat{\varphi}}(\hat{\zeta}_t) / \partial \varphi}{f_{\hat{\varphi}}(\hat{\zeta}_t)} \hat{s}_2(\hat{\zeta}_t) \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \hat{s}_2^2(\hat{\zeta}_t) \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \hat{s}_2(\hat{\zeta}_t) \frac{1}{\hat{\lambda}_t}$$

$$\hat{s}_2(\zeta) = - \left(1 + \zeta \frac{f'_{\hat{\varphi}}(\zeta)}{f_{\hat{\varphi}}(\zeta)} \right), \quad \hat{\lambda}_t = 1 - \hat{\beta} - \hat{\gamma} + \hat{\beta} \hat{\lambda}_{t-1} + \hat{\gamma} \frac{\ell_{t-1}}{\hat{g}((t-1)/T)}.$$

Note that we may obtain initial consistent estimates of φ by the method of moments. For example, in the gamma case, parameterized to have mean one, the parameter φ can be estimated as one over the variance.

Under the i.i.d. structure, one can also improve the estimation of g by using local likelihood. Suppose that f, θ were known, then the local likelihood estimator of $g(u)$ based on data ℓ_t is

given by the maximizer of

$$\sum_{t=1}^T K_h(t/T - u) \left(\ln f(\zeta_t(g)) - \ln g \right), \quad (4.11)$$

$$\zeta_t(g) = \frac{\ell_t}{\lambda_t g}, \quad t = 1, \dots, T, \quad (4.12)$$

with respect to the parameter $g \in \mathbb{R}_+$. Following [Fan and Chen \(1999\)](#), we may update the estimator of g by

$$\tilde{g}_{LL}(u) = \hat{g}(u) - \hat{L}_{gg}^{-1}(\hat{g}(u); u) \hat{L}_g(\hat{g}(u); u), \quad (4.13)$$

where $\hat{L}_g(g; u) = \partial \hat{L}(g; u) / \partial g$ and $\hat{L}_{gg}(g; u) = \partial^2 \hat{L}(g; u) / \partial g^2$ with

$$\hat{L}(g; u) = \sum_{t=1}^T K_h(t/T - u) \left(\ln f_{\hat{\varphi}}(\tilde{\zeta}_t(g)) - \ln g \right) \quad (4.14)$$

$$\tilde{\zeta}_t(g) = \frac{\ell_t}{g \lambda_t(\hat{\theta}, \hat{g})}, \quad t = 1, \dots, T. \quad (4.15)$$

4.4.2.2 Nonparametric density case

Suppose we have initial consistent estimators of $\theta, g(\cdot)$. Then, one can estimate the density function f using the residuals

$$\hat{\zeta}_t = \frac{\ell_t}{\hat{g}(t/T) \hat{\lambda}_t(\hat{\theta})}, \quad t = 1, \dots, T.$$

In particular, the kernel estimator of the density is given by

$$\hat{f}(\zeta) = \frac{1}{T} \sum_{t=1}^T K_{h_f}(\hat{\zeta}_t - \zeta),$$

where h_f is another bandwidth sequence.

Next, one may proceed in a fourth step to improve estimates of all parameters and the function g using the estimated density $\hat{f}$ as the basis of a likelihood estimation. The efficient score function (derived in the appendix) for θ in the semiparametric model with unknown f, g is

$$L_{\theta}^{**} = \sum_{t=1}^T \ell_{\theta t}^{**},$$

$$\begin{aligned} \ell_{\theta t}^{**} = s_2(\zeta_t) & \left(\left(\frac{\partial \log \lambda_t}{\partial \theta} - E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \left(\frac{1}{\lambda_t} - E \left(\frac{1}{\lambda_t} \right) \right) \frac{E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right) \\ & + \frac{\zeta_t - 1}{\sigma_\zeta^2} \left(E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \frac{E \left(\frac{1}{\lambda_t} \right) E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right). \end{aligned}$$

Efficient estimation can be conducted by two-step estimation based on initial consistent estimates of θ, f, g :

$$\tilde{\theta} = \hat{\theta} + \left(\frac{1}{T} \sum_{t=1}^T \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g}) \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g})^\top \right)^{-1} \frac{1}{T} \sum_{t=1}^T \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g}) \quad (4.16)$$

$$\begin{aligned} \ell_{\theta t}^{**}(\hat{\theta}, \hat{f}, \hat{g}) = \hat{s}_2(\hat{\zeta}_t) & \left(\left(\frac{\partial \log \hat{\lambda}_t}{\partial \theta} - \frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} - \left(\frac{1}{\hat{\lambda}_t} - \frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t} \right) \frac{\frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \right) \right) \\ & + \frac{\hat{\zeta}_t - 1}{\hat{\sigma}_\zeta^2} \left(\frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} - \frac{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial \log \hat{\lambda}_t}{\partial \theta} \frac{1}{\hat{\lambda}_t}}{\frac{1}{T} \sum_{t=1}^T \frac{1}{\hat{\lambda}_t^2}} \right). \end{aligned}$$

where $\hat{\sigma}_\zeta^2 = \sum_{t=1}^T (\hat{\zeta}_t - \bar{\zeta})^2 / T$ with $\bar{\zeta} = \sum_{t=1}^T \hat{\zeta}_t / T$.

In practice, we may update the estimator of g by the one-step improvement

$$\tilde{g}_{LL}(u) = \hat{g}(u) - \hat{L}_{gg}^{-1}(\hat{g}(u); u) \hat{L}_g(\hat{g}(u); u),$$

where $\hat{L}_g(g; u) = \partial \hat{L}(g; u) / \partial g$ and $\hat{L}_{gg}(g; u) = \partial^2 \hat{L}(g; u) / \partial g^2$ with

$$\hat{L}(g; u) = \sum_{t=1}^T K_h(t/T - u) \left(\ln \hat{f}(\tilde{\zeta}_t(g)) - \ln g \right), \quad (4.17)$$

$$\tilde{\zeta}_t(g) = \frac{\ell_t}{g \hat{\lambda}_t(\hat{\theta})}, \quad t = 1, \dots, T. \quad (4.18)$$

These procedures can be iterated, that is, given consistent initial estimators of g, θ , we estimate f . Then we use the estimated f to update our estimate of θ taking the initial estimator of g , then we update our estimator of g using the estimated f and updated θ . We can continue this operation until it meets the defined criteria of convergence.

4.4.3 Sieve semiparametric estimation procedure

An alternative approach is based on the sieve method, [Chen \(2007\)](#). The advantage of this method is that it only requires a single optimization, albeit one with many parameters to choose.

We just consider the cases where the error ζ_t is i.i.d. with mean one and density f : (a) f is parametric with f_φ that reflects the unit mean constraint; (b) f is nonparametric. We suppose in both cases that

$$\ln g(u) = \sum_{j=0}^{\infty} \psi_j H_j(u)$$

for some basis functions H_j defined on $[0, 1]$. In case (b) we will also consider

$$\ln f(v) = \sum_{j=0}^{\infty} \alpha_j \Gamma_j(v)$$

for some basis functions Γ_j defined on $\mathbb{R}_+$.

We first consider case (a) where f is parametric (and φ is a scalar). In that case, the sieve log likelihood for $\eta = (\theta^\top, \varphi, \psi^\top)^\top$ is

$$L(\eta | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \sum_{j=0}^M \psi_j H_j(t/T) + \sum_{t=1}^T f_\varphi(\zeta_t(\theta, \psi)),$$

$$\zeta_t(\theta, \psi) = \frac{\ell_t}{\lambda_t(\theta) \exp \left(\sum_{j=0}^M \psi_j H_j(u) \right)},$$

where $M = M(T)$ is the truncation parameter. We just maximize this criterion function with respect to $\eta \in \mathbb{R}^{M+3}$. Under some restrictions, the estimates $\hat{\theta}, \hat{\varphi}$ of θ, φ are consistent and asymptotically normal with mean zero and finite variance, while we can show that the estimated function

$$\hat{g}(u) = \exp \left(\sum_{j=0}^M \hat{\psi}_j H_j(u) \right)$$

is also consistent and asymptotically normal at a slower rate. Specifically, the standard errors are calculated as follows. Define the matrices

$$\hat{\mathcal{I}}_{\eta\eta} = \frac{\partial^2 L(\hat{\eta} | \ell_1, \dots, \ell_T)}{\partial \eta \partial \eta^\top}, \quad \hat{\mathcal{I}}_{\eta\eta}^{-1} = \begin{pmatrix} \hat{\mathcal{I}}^{\theta\theta} & \hat{\mathcal{I}}^{\theta\varphi} & \hat{\mathcal{I}}^{\theta\psi} \\ \hat{\mathcal{I}}^{\varphi\theta} & \hat{\mathcal{I}}^{\varphi\varphi} & \hat{\mathcal{I}}^{\varphi\psi} \\ \hat{\mathcal{I}}^{\psi\theta} & \hat{\mathcal{I}}^{\psi\varphi} & \hat{\mathcal{I}}^{\psi\psi} \end{pmatrix}.$$

Then it can be shown that

$$\left(\hat{\mathcal{I}}^{\theta\theta} \right)^{-1/2} (\hat{\theta} - \theta) \Rightarrow N(0, I_2), \quad \left(\hat{\mathcal{I}}^{\varphi\varphi} \right)^{-1/2} (\hat{\varphi} - \varphi) \Rightarrow N(0, 1), \quad (4.19)$$

$$\left(H^\top \hat{\mathcal{I}}^{\psi\psi} H \right)^{-1/2} \exp(-\hat{g}(u)) (\hat{g}(u) - g(u)) \Rightarrow N(0, 1), \quad (4.20)$$

where $H = (H_1, \dots, H_M)^\top$ and we have suppressed the dependence on u .

We now turn to case (b) where f is treated nonparametrically. In this case we also need to impose the restrictions that f is a density and that ζ has mean one. In this case, the restricted

sieve log likelihood for $\eta = (\theta^\top, \psi^\top, \alpha^\top, \delta)^\top$ is

$$L(\theta, \psi, \alpha, \delta | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \sum_{j=0}^M \psi_j H_j(t/T) + \sum_{t=1}^T \sum_{j=0}^N \alpha_j \Gamma_j(\zeta_t(\theta, \psi)) \\ - \ln \left(\int \exp \left(\sum_{j=0}^N \alpha_j \Gamma_j(z) \right) dz \right) - \delta \left(\int z \exp \left(\sum_{j=0}^N \alpha_j \Gamma_j(z) \right) dz - 1 \right),$$

where $\zeta_t(\theta, \psi)$ is as defined above and δ is a Lagrange multiplier parameter that imposes the unit mean restriction on f . We find the solution to the first order condition with respect to $\eta \in \mathbb{R}^{M+N+2}$. Define for this η the matrices

$$\hat{\mathcal{I}}_{\eta\eta} = \frac{\partial^2 L(\hat{\eta} | \ell_1, \dots, \ell_T)}{\partial \eta \partial \eta^\top}, \quad \hat{\mathcal{I}}_{\eta\eta}^{-1} = \begin{pmatrix} \hat{\mathcal{I}}^{\theta\theta} & \hat{\mathcal{I}}^{\theta\psi} & \hat{\mathcal{I}}^{\theta\alpha} & \hat{\mathcal{I}}^{\theta\delta} \\ & \hat{\mathcal{I}}^{\psi\psi} & \hat{\mathcal{I}}^{\psi\alpha} & \hat{\mathcal{I}}^{\psi\delta} \\ & & \hat{\mathcal{I}}^{\alpha\alpha} & \hat{\mathcal{I}}^{\alpha\delta} \\ & & & \hat{\mathcal{I}}^{\delta\delta} \end{pmatrix}.$$

Then it can be shown that

$$\left(\hat{\mathcal{I}}^{\theta\theta} \right)^{-1/2} \left(\hat{\theta} - \theta \right) \Rightarrow N(0, I_2), \quad (4.21)$$

$$\left(H^\top \hat{\mathcal{I}}^{\psi\psi} H \right)^{-1/2} \exp(-\hat{g}(u)) \left(\hat{g}(u) - g(u) \right) \Rightarrow N(0, 1), \quad (4.22)$$

$$\left(\Gamma^\top \hat{\mathcal{I}}^{\alpha\alpha} \Gamma \right)^{-1/2} \exp(-\hat{f}(\zeta)) \left(\hat{f}(\zeta) - f(\zeta) \right) \Rightarrow N(0, 1), \quad (4.23)$$

where $H = (H_1, \dots, H_M)^\top$, $\Gamma = (\Gamma_1, \dots, \Gamma_N)^\top$, and we have suppressed the dependence on u, ζ .

The difficulty with this approach is the potential high dimensional parameter space that has to be navigated when the truncation parameters M, N are large.

4.5 Large sample properties

The distribution theory for this requires some extension of [Hafner and Linton \(2010\)](#). The presence of two nonparametric functions g, f , albeit scalar ones, makes this a little challenging. We make some assumptions.

Definition. A triangular array process $\{X_{t,T}, t = 0, 1, 2, \dots, T = 1, 2, \dots\}$ is said to be alpha mixing if

$$\alpha(k) = \sup_{T \geq 1} \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty} |P(AB) - P(A)P(B)| \rightarrow 0, \quad (4.24)$$

as $k \rightarrow \infty$, where $\mathcal{F}_{-\infty}^{n,T}$ and $\mathcal{F}_{n+k,T}^\infty$ are two σ -fields generated by $\{X_{t,T}, t \leq n\}$ and $\{X_{t,T}, t \geq n+k\}$ respectively. We call $\alpha(\cdot)$ the mixing coefficient.

We suppose that ℓ_t^* is stationary and alpha mixing. This can be shown to hold under the parameter restrictions provided ζ_t . It may also hold when ζ_t itself is only described as a stationary mixing process although this can be difficult to establish. Instead, one can work with the more general near epoch dependence condition, see [Lu and Linton \(2007\)](#).

We define the long run variance for a stationary mixing process x_t as

$$\text{lrvar}(x_t) = \sum_{j=-\infty}^{\infty} \text{cov}(x_t, x_{t-j}).$$

4.5.1 Conditional moment restrictions

We first consider the properties of the GMM estimator based on the first conditional moment restriction. This estimator makes the weakest assumptions about the process ζ_t and so it is more robust than the subsequent procedures we analyze. We do not address the efficient use of this information but it follows from standard arguments.

4.5.1.1 Nonparametric trend

We first consider the estimator $\hat{g}(u)$, $u \in (0, 1)$, that is based on smoothing of the raw liquidity. Let $v_t = \lambda_t \zeta_t - 1$.

ASSUMPTION A1. *Suppose that g is twice continuously differentiable at $u \in (0, 1)$.*

ASSUMPTION A2. *Suppose that $\{v_t\}$ is an alpha mixing sequence with $E(v_t) = 0$ and $E(|v_t|^{2+\delta}) \leq C < \infty$ for some $\delta > 0$ for $t = 1, 2, \dots$ such that*

$$\sum_{k=1}^{\infty} \alpha(k)^{\frac{\delta}{2+\delta}} < \infty.$$

ASSUMPTION A3. *Suppose that K is symmetric about zero with compact support $[-1, 1]$ and K is differentiable on its support.*

Theorem 4. *Suppose that assumptions A1-A3 hold and that $h \rightarrow 0$ and $Th \rightarrow \infty$. Then*

$$\sqrt{Th} \left(\hat{g}(u) - g(u) - h^2 b(u) \right) \implies N(0, V(u)),$$

$$V(u) = g^2(u) \|K\|^2 \times \text{lrvar}(v_t).$$

This estimator is consistent and asymptotically normal with a potentially optimal rate of $T^{-2/5}$ based on the smoothness assumption. The bandwidth that achieves this rate is of order $h = T^{-1/5}$ and balances squared bias with variance. We know that $\ell_t^* = \ell_t/g(t/T) = \lambda_t \zeta_t$ is an ARMA(1,1) process with $A(L)\ell_t^* = B(L)e_t$ for some MDS shock e_t and lag polynomials A, B .

In this case, the long run variance of ℓ_t^* is $\sigma_e^2(B(1)/A(1))^2$ and it could be estimated by the plug in of estimated θ from the second step.

Instead, it may be preferable to work with the refined estimator $\tilde{g}(u)$ that is based on the estimator of θ . We have for this estimator the following CLT.

Theorem 5. *Suppose that assumptions A1-A3 hold and that $\hat{\theta}$ is $\sqrt{T}$ -consistent. Suppose that $h \rightarrow 0$ and $Th \rightarrow \infty$. Then*

$$\sqrt{Th} \left(\tilde{g}(u) - g(u) - h^2 b(u) \right) \implies N(0, V(u)),$$

$$V(u) = g^2(u) \|K\|^2 \times \text{var}(\zeta_t).$$

In this case, the limiting variance is proportional to the variance of ζ_t , which is generally smaller and easier to estimate. When ζ_t is i.i.d.,

$$E(\lambda_t^2(\zeta_t - 1)^2) = E(\lambda_t^2) \times \text{var}(\zeta_t) \geq \text{var}(\zeta_t),$$

because by assumption $E(\lambda_t) = 1$. For this estimator, consistent standard errors can be based on

$$\hat{V}(u) = \tilde{g}^2(u) \|K\|^2 \times \hat{\sigma}_\zeta^2, \quad (4.25)$$

where $\hat{\sigma}_\zeta^2$ is an estimator of the unconditional variance of ζ_t such as defined above.

4.5.1.2 Parametric components

Let $\theta = (\beta, \gamma)^\top$ and let W be some given positive definite weighting matrix. Let

$$w_t = \lambda_t(\zeta_t - 1)z_{t-1} + \frac{1 - \beta - \gamma}{1 - \beta}(\lambda_t \zeta_t - 1)E(z_{t-1}).$$

We consider for any function g

$$M_T(\theta, g) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, g), \quad \rho_t(\theta, g) = z_{t-1} \left(\frac{\ell_t}{g(t/T)} - \lambda_t(\theta, g) \right)$$

$$\lambda_t(\theta, g) = 1 - \beta - \gamma + \beta \lambda_{t-1} + \gamma \frac{\ell_{t-1}}{g((t-1)/T)} = \frac{1 - \beta - \gamma}{1 - \beta} + \gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{\ell_{t-j}}{g((t-j)/T)}.$$

ASSUMPTION A4. *Define:*

$$M(\theta, g) = \lim_{T \rightarrow \infty} E(M_T(\theta, g))$$

$$\Gamma_1(\theta, g_0) = \frac{\partial M(\theta, g_0)}{\partial \theta}$$

$$\Gamma_2(\theta, g_0) \circ (g - g_0) = \frac{\partial}{\partial \tau} M(\theta, g_0 + \tau(g - g_0)),$$

which are assumed to exist in all directions θ, g .

ASSUMPTION A5. For all $\delta > 0$, there is a $\epsilon > 0$ such that

$$\inf_{\|\theta - \theta_0\| > \delta} \|M(\theta, g_0)\| \geq \epsilon.$$

Uniformly for all $\theta \in \Theta$, the function $M(\theta, g)$ is continuous in g (with the respect to the L_2 metric) at $g = g_0$. Furthermore,

$$\sup_{\theta \in \Theta, \|g - g_0\| \leq \delta_T} \|M_T(\theta, g) - M(\theta, g_0)\| = o_P(1).$$

ASSUMPTION A6. For all sequences of positive numbers $\delta_T \rightarrow 0$

$$\sup_{\|\theta - \theta_0\| \leq \delta_T, \|g - g_0\| \leq \delta_T} \delta_T^{-2} \|M(\theta, g) - M(\theta, g) - \Gamma_2(\theta, g_0) \circ (g - g_0)\| \leq C$$

$$\sup_{\|\theta - \theta_0\| \leq \delta_T, \|g - g_0\| \leq \delta_T} \delta_T^{-1} \|\Gamma_2(\theta, g_0) \circ (g - g_0) - \Gamma_2(\theta_0, g_0) \circ (g - g_0)\| = o(1)$$

$$\sup_{\|\theta - \theta_0\| \leq \delta_T, \|g - g_0\| \leq \delta_T} \sqrt{T} \|M_T(\theta, g) - M(\theta, g) - M_T(\theta_0, g_0)\| = o_P(1)$$

Theorem 6. Suppose that Assumptions A1-A6 hold. Then as $T \rightarrow \infty$

$$\sqrt{T} (\hat{\theta} - \theta) \Rightarrow N(0, V)$$

$$\Omega = \lim_{T \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T w_t \right), \quad \Gamma = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left(\frac{\partial e_t(\theta_0)}{\partial \theta} z_{t-1} \right)$$

$$V = (\Gamma^\top W \Gamma)^{-1} (\Gamma^\top W \Omega W \Gamma) (\Gamma^\top W \Gamma)^{-1}.$$

In general, the asymptotic variance will depend on the long run variance of the process w_t and so inference procedures are complicated by that.

4.5.2 I.i.d. shocks

We suppose here that ζ_t is i.i.d. with mean one and density f . Define the so-called Fisher score functions and informations

$$s_1(\zeta) = \frac{f'(\zeta)}{f(\zeta)}, \quad s_2(\zeta) = 1 + \zeta \frac{f'(\zeta)}{f(\zeta)} \quad (4.26)$$

$$I_j(f) = \int s_j^2(\zeta) f(\zeta) d\zeta, \quad j = 1, 2. \quad (4.27)$$

We suppose that we have initial consistent estimators of $g(\cdot), \theta$ available from the GMM procedure described above, say.

4.5.2.1 Parametric density

In the case where f is parametrically specified with parameters φ , the model is semiparametric with parameters $\eta = (\theta^\top, \varphi)^\top$ and unknown function g . We first consider the local likelihood estimator of the trend function. We note that regarding the estimation of $g(\cdot)$ it may be assumed that the error density is known, whether this is estimated parametrically or nonparametrically, [Linton and Xiao \(2001\)](#).

Theorem 7. *Suppose that assumptions A1-A3 hold and that $\hat{\theta}$ is $\sqrt{T}$ -consistent. Suppose that $h \rightarrow 0$ and $Th \rightarrow \infty$. The local likelihood estimator has for some bias $b(u)$,*

$$\sqrt{Th} \left(\tilde{g}_{LL}(u) - g(u) - h^2 b(u) \right) \Rightarrow N(0, V(u))$$

$$V(u) = \|K\|^2 I_2^{-1}(f) g(u)^2.$$

To construct pointwise confidence bands we use:

$$\tilde{V}(u) = \|K\|^2 I_2^{-1}(f_{\tilde{\varphi}}) \tilde{g}(u)^2,$$

$$I_2(f_{\tilde{\varphi}}) = \frac{1}{T} \sum_{t=1}^T \left(1 + \zeta \frac{f'_{\tilde{\varphi}}(\tilde{\zeta}_t)}{f_{\tilde{\varphi}}(\tilde{\zeta}_t)} \right)^2,$$

where $\tilde{\zeta}_t$ are the estimated residuals.

We next turn to the properties of the estimated parametric components. Define the efficient information matrix:

$$\mathcal{I}_{\eta\eta}^* = \begin{pmatrix} \mathcal{I}_{\theta\theta}^* & \mathcal{I}_{\theta\varphi}^* \\ \mathcal{I}_{\varphi\theta}^* & \mathcal{I}_{\varphi\varphi}^* \end{pmatrix}, \quad \mathcal{I}_{\theta\theta}^* = E(\ell_{\theta t}^* \ell_{\theta t}^{*\top}), \quad \mathcal{I}_{\varphi\varphi}^* = E(\ell_{\varphi t}^* \ell_{\varphi t}^{*\top}), \quad \mathcal{I}_{\theta\varphi}^* = E(\ell_{\theta t}^* \ell_{\varphi t}^{*\top}).$$

Theorem 8. *Suppose that Assumptions A1-A6 hold. Then as $T \rightarrow \infty$*

$$\sqrt{T}(\tilde{\eta} - \eta) \Rightarrow N(0, \mathcal{I}_{\eta\eta}^{*-1}).$$

Furthermore, the asymptotic variance may be estimated by $\tilde{\mathcal{I}}^{*-1}$, where

$$\tilde{\mathcal{I}}_{\eta\eta}^* = \frac{1}{T} \sum_{t=1}^T \ell_{\eta t}^*(\tilde{\theta}, \tilde{\varphi}, \tilde{g}) \ell_{\eta t}^{*\top}(\tilde{\theta}, \tilde{\varphi}, \tilde{g}).$$

4.5.2.2 Nonparametric density case

In the case where f is nonparametrically specified, the model is semiparametric with parameters θ and unknown functions g, f . First, the large sample properties of the estimated g are as if f is known. We turn to the properties of the estimated parametric components.

Define the efficient information matrix

$$\mathcal{I}_{\theta\theta}^{**} = E(\ell_{\theta t}^{**} \ell_{\theta t}^{**\top}).$$

Theorem 9. *Suppose that Assumptions A1-A6 hold. Then as $T \rightarrow \infty$*

$$\sqrt{T} \left(\tilde{\theta} - \theta \right) \Rightarrow N(0, \mathcal{I}_{\theta\theta}^{*-1}).$$

Furthermore, the asymptotic variance may be estimated by $\tilde{\mathcal{I}}^{*-1}$, where

$$\tilde{\mathcal{I}}_{\theta\theta}^{**} = \frac{1}{T} \sum_{t=1}^T \ell_{\theta t}^{**}(\tilde{\theta}, \tilde{f}, \tilde{g}) \ell_{\theta t}^{**}(\tilde{\theta}, \tilde{f}, \tilde{g})^\top.$$

The presence of unknown τ does not affect the efficiency of θ once the scale σ_ζ^2 is accounted for, but the scale does affect the achievable efficiency.

4.6 Testing for temporary and permanent shifts

We can estimate the function g allowing for a discontinuity at the point $u_0 = t_0/T$ by considering

$$\hat{g}^+(u) = \frac{1}{T} \sum_{t=1}^T K_h^+(t/T - u) \ell_t, \quad \hat{g}^-(u) = \frac{1}{T} \sum_{t=1}^T K_h^-(t/T - u) \ell_t,$$

where K^+ is a kernel supported on $[0, 1]$ with $\int K^+(u) du = 1$ and $\int K^+(u) u du = 0$ and K^- is a kernel supported on $[-1, 0]$ with $\int K^-(u) du = 1$ and $\int K^-(u) u du = 0$. We may test for the presence of a discontinuity by computing

$$\tau(u_0) = \sqrt{Th} \frac{\hat{g}^+(u_0) - \hat{g}^-(u_0)}{\sqrt{\hat{\sigma}^{2+}(u_0) \|K^+\|^2 + \hat{\sigma}^{2-}(u_0) \|K^-\|^2}},$$

where $\hat{\sigma}^{2\pm}(u_0) \|K^\pm\|^2 / Th$ are estimates of the asymptotic variance of $\hat{g}^\pm(u_0)$. In general,

$$\hat{\sigma}^{2\pm}(u_0) = \hat{g}^\pm(u_0)^2 \times \widehat{\text{Irvar}}(\lambda_t \zeta_t),$$

because $\ell_t - g(t/T) = g(t/T) (\lambda_t \zeta_t - 1)$ has smoothly varying variance that is exactly $g^2(t/T)$ and because the series $\lambda_t \zeta_t$ is in general weakly dependent. On the other hand if we work with the improved estimator that works with a smooth of $\ell_t / \hat{\lambda}_t$, we can assume that the error process is a MDS and so the variance of the estimator is proportional to $g^\pm(t/T)^2 \sigma_\zeta^2$. That is, we may define

$$\tilde{g}^+(u) = \frac{1}{T} \sum_{t=1}^T K_h^+(t/T - u) \frac{\ell_t}{\hat{\lambda}_t}, \quad \tilde{g}^-(u) = \frac{1}{T} \sum_{t=1}^T K_h^-(t/T - u) \frac{\ell_t}{\hat{\lambda}_t}.$$

In this case, we may choose

$$\tilde{\sigma}^{2\pm}(u_0) = \tilde{g}^\pm(u_0)^2 \times \hat{\sigma}_\zeta^2.$$

Likewise for the local likelihood estimator, but this also takes care of the error shape and heteroskedasticity. Given the studentized statistic, $\tau(u_0)$ we compare this with the standard normal distribution as in [Delgado and Hidalgo \(2000\)](#). Under the null hypothesis this should lie between $\pm z_{\alpha/2}$ with probability $1 - \alpha$.

We also consider how to include a control group to eliminate common trends at the change time. This amounts to a diff in diff test. Specifically, suppose that we have a “treatment” stock labelled with an S subscript and a “control” stock labelled with an C subscript. We suppose that model (4.4) holds for both stocks and that ζ_{St} and ζ_{Ct} may be correlated. We define the diff-in-diff statistic as

$$\tau_{did}(u_0) = \sqrt{Th} \frac{(\tilde{g}_S^+(u_0) - \tilde{g}_S^-(u_0)) - (\tilde{g}_C^+(u_0) - \tilde{g}_C^-(u_0))}{\sqrt{(\tilde{\sigma}_S^{2+}(u_0) + \tilde{\sigma}_C^{2+}(u_0) - 2\tilde{\sigma}_{S,C}^+(u_0)) \|K^+\|^2 + (\tilde{\sigma}_S^{2-}(u_0) + \tilde{\sigma}_C^{2-}(u_0) - 2\tilde{\sigma}_{S,C}^-(u_0)) \|K^-\|^2}},$$

where:

$$\begin{aligned} \tilde{\sigma}_{S,C}^\pm(u) &= \tilde{g}_S^\pm(u_0) \tilde{g}_C^\pm(u_0) \times \hat{\sigma}_{\zeta_S, \zeta_C}, \\ \hat{\sigma}_{\zeta_S, \zeta_C} &= \frac{1}{T} \sum_{t=1}^T \left(\hat{\zeta}_{St} - \bar{\zeta}_S \right) \left(\hat{\zeta}_{Ct} - \bar{\zeta}_C \right). \end{aligned}$$

What happens if we reject this test at some point $u_0 = t_0/T$? Note that the above procedures (estimation of θ) can allow $\hat{g}(u)$ to be inconsistent at a finite number of points, because the estimators of θ, f are averages of $\hat{g}(t/T)$ over all points $t = 1, \dots, T$. However, the rate of convergence may be affected, since the jump points will contribute a bias of order $O(h)$ rather than the usual order $O(h^2)$. Therefore, one should choose a smaller bandwidth such that $\sqrt{Th} \rightarrow 0$ instead of $\sqrt{Th^2} \rightarrow 0$. Having identified the jump points one can use this information in the rescaling. We suppose in general that the function g is continuous from the right with limits from the left. In that case for $t \geq t_0$ we may normalize by $\hat{g}^+(u)$ and for $t < t_0$ we normalize by $\hat{g}^-(u)$. That is, we let

$$\hat{\ell}_t^* = \begin{cases} \frac{\ell_t}{\hat{g}^+(t/T)} & t = t_0, \dots, T \\ \frac{\ell_t}{\hat{g}^-(t/T)} & t = 1, \dots, t_0 - 1 \end{cases},$$

and proceed to estimate the dynamic parameters as before.

We next discuss the estimation of temporary effects in the dynamic equation

$$\lambda_t = 1 - \beta - \gamma + \beta \lambda_{t-1} + \sum_{j=0}^{J-1} \alpha_j D_{jt} + \gamma \ell_{t-1}^*,$$

where J is fixed. Here, D_{jt} are dummy variables indicating times $t_0, \dots, t_{J-1}$ and we focus on the case where $t_j = t_0 + j$. In this case it is not possible to consistently estimate the parameters α_j , however, it is possible to provide a consistent test of the null hypothesis that $\alpha_1 = \dots = \alpha_J = 0$ against the general alternative even in the full semiparametric model. The efficient score function with respect to α in the semiparametric model with known f is

$$\sum_{t=1}^T \ell_{\theta t}^* = \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \alpha} - \frac{\frac{1}{T} \sum_{t=1}^T E \left[\frac{\partial \lambda_t}{\partial \alpha} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right)$$

where

$$\frac{\partial \lambda_t(\theta, \alpha)}{\partial \alpha_j} = \beta \frac{\partial \lambda_{t-1}(\theta, \alpha)}{\partial \alpha_j} + D_{jt} = \begin{cases} \beta^{t-t_j} & \text{if } t \geq t_j \\ 0 & \text{if } t < t_j. \end{cases}$$

$$\frac{1}{T} \sum_{t=1}^T E \left[\frac{\partial \lambda_t}{\partial \alpha} \frac{1}{\lambda_t^2} \right] = \frac{1}{T} \sum_{t=t_j}^T \beta^{t-t_j} E \left[\frac{1}{\lambda_t^2} \right].$$

It follows that the efficient score function at $\alpha = 0$ is

$$\frac{\partial L}{\partial \alpha_j}(\theta, \sigma_\zeta^2, 0) = \sum_{t=t_j}^T s_2(\zeta_t) \frac{1}{\lambda_t(\theta, 0)} \left(\left[\beta^{t-t_j} - \frac{1}{T} \sum_{t=t_j}^T \beta^{t-t_j} \right] \right) \simeq \sum_{t=t_j}^T s_2(\zeta_t) \frac{\beta^{t-t_j}}{\lambda_t(\theta, 0)}.$$

In practice we must replace the unknown quantities by estimates. Define the test statistic (we call CAR to recognize the event study literature where this quantity originates):

$$\widehat{CAR}(\tau) = \sum_{j=0}^{\tau} \sum_{t=t_j}^T s_2(\hat{\zeta}_t) \frac{\hat{\beta}^{t-t_j}}{\lambda_t(\hat{\theta}, 0)}, \quad \tau = 0, \dots, J-1. \quad (4.28)$$

The test statistic does not satisfy a central limit theorem (even when θ is known) because of the summability of $\sum_{t=t_j}^T \beta^{2(t-t_j)}$ (that is, essentially only a finite number of periods matter). Nevertheless, if the distribution of ζ_t were known along with the parameter values θ , we can calculate the distribution numerically using data $w_r(\tau) = \sum_{j=1}^{\tau} \sum_{t=r}^T s_2(\zeta_t) \beta^{t-r} / \lambda_t(\theta, 0)$ for r some time before t_0 . Let F_w denote the distribution of the series $\{w_r\}$. We assume that t_0 is large, i.e., $t_0 \rightarrow \infty$ so that there is a long sample of data available before the intervention.

We compare $\widehat{CAR}(\tau)$ with the critical values $\hat{F}_w^{-1}(\alpha/2)$, $\hat{F}_w^{-1}(1 - \alpha/2)$, where $\hat{F}_w(\cdot)$ is estimated using the data

$$\hat{w}_r(\tau) = \sum_{j=1}^{\tau} \sum_{t=r}^T s_2(\hat{\zeta}_t) \frac{\hat{\beta}^{t-r}}{\lambda_t(\hat{\theta}, 0)}, \quad r = 1, \dots, t_0 - J. \quad (4.29)$$

4.7 Risk premium

Amihud (2002) considers an autoregressive model for annual and monthly liquidity and then relates this to the stock risk premium. Specifically, he writes

$$E(R_{mt} - R_{ft} | liq_t^e) = a + b \times liq_t^e$$

$$liq_t = c_0 + c_1 \times liq_{t-1} + \eta_t,$$

where $liq_t^e = E(liq_t | \mathcal{F}_{t-1}) = c_0 + c_1 liq_{t-1}$ and liq_t is the annual or monthly average that we have called A_t . He also considers a specification with unexpected liquidity as a regressor where $liq_t^u = liq_t - liq_t^e$.

It seems a bit strange to wait for a whole year to update ones estimate of liquidity. We may consider the following specification for daily stock returns

$$E(R_{mt} - R_{ft} | \mathcal{F}_{t-1}) = a + b \times g(t/T) + c \times \lambda_t + d \times \zeta_t, \quad (4.30)$$

where λ_t is defined above. This allows the risk premium to depend on long run trend liquidity on short run predictable dynamic variation and also on unanticipated liquidity shocks.

We also consider the alternative regression with detrended equity premium, that is,

$$E(R_{mt} - R_{ft} - m(t/T) | \mathcal{F}_{t-1}) = \alpha + \gamma \times \lambda_t + \delta \times \zeta_t, \quad (4.31)$$

where $m(t/T) = E(R_{mt} - R_{ft})$ is the time varying unconditional equity premium. In practice we can estimate $m(\cdot)$ by kernel smoothing methods.

4.8 Empirical application

The ability to accurately model illiquidity series, and the availability of a framework to conduct inference on potential structural changes in their dynamics, are useful tools to investigate liquidity conditions in financial markets and their evolution over time. In our empirical application, we consider the Fab 5 tech stocks and the Bitcoin asset introduced in Section 4.2 to analyze their illiquidity series using our DARLiQ model. The rest of this section is organized as follows: Section 4.8.1 presents the data description and the key descriptive statistics of the illiquidity series. Section 4.8.2.1 covers the estimation results adopting a GMM approach based on moment restrictions. Sections 4.8.2.2 and 4.8.2.3 present the analysis based on a semiparametric ML estimation procedure with, respectively, parametrically specified and nonparametrically estimated

error densities. Section 4.8.3 focuses on the detection of permanent and temporary breaks in the illiquidity series arising from stock splits. Finally, We investigate in Section 4.8.4 how market liquidity dynamics – particularly short-run autoregressive dynamics and unexpected liquidity shocks – impact equity risk premia.

4.8.1 Data description

We use historical daily return and volume data from a panel composed of the five tech stocks and the bitcoin asset to compute the Amihud illiquidity series.³ The sample period starts from the date of the first available data point for each asset until October 7th, 2021. The descriptive statistics of the illiquidity series are summarized in Table 4.2.⁴ It can be observed that the Bitcoin asset is less liquid compared to the technology company stocks during this period. In addition, the illiquidity series of Bitcoin is more volatile, exhibits higher skewness and has thicker tails. We further note that the five tech companies have comparable levels of liquidity – although Apple stock is slightly more liquid than others. Moreover, the illiquidity of Facebook stock has higher skewness and thicker tails compared to the other four tech companies.

Table 4.2: Summary statistics for daily illiquidity – $\ell_t \times 10^{10}$.

	Facebook	Amazon	Apple	Google	Microsoft	Bitcoin
Mean	0.0372	0.0313	0.0187	0.0615	0.0424	1.7013
StdDev	0.0295	0.0389	0.0148	0.0499	0.0398	4.1201
Skewness	1.3673	2.5656	1.1146	1.1921	1.6408	4.0626
Kurtosis	6.2978	10.5467	4.1849	4.5345	6.1514	24.9266

We plot in Figure 4.3 and Figure 4.4 respectively the illiquidity and log illiquidity series over the corresponding sample period for each of the six assets. To manage boundary issues, we obtain an initial consistent estimator of the trend function $g(t/T)$ using a local linear estimator. The red curves in the two figures represent respectively the estimated trend functions and their logarithms. From Figure 4.3, we observe that the estimated trend function $g(t/T)$ serves as a good approximation for the time-varying mean of the illiquidity series.⁵ Furthermore, a strong

³Data were retrieved from Yahoo Finance.

⁴To make it comparable, we use the daily Amihud illiquidity ratios in the common sample period of September 19, 2014 to October 7th, 2021 to compute the descriptive statistics for all assets.

⁵Note that the trend function $g(t/T)$ is the mean level of the illiquidity ℓ_t , i.e. $E(\ell_t) = g(t/T)$, which is

downward trend is observed in the evolution of most illiquidity series, indicating an overall improvement in liquidity conditions over time. Lastly, it is worth noticing that a temporary worsening in liquidity conditions is occurring during significant market events such as the burst of the dot-com bubble and 2007-2009 Global Financial Crisis.

4.8.2 Estimation results

We introduce the detrended illiquidity series, $\ell_t^* = \ell_t/g(t/T)$, which are assumed to be mean stationary. We then estimate the parameters θ of the λ_t process based on moment restrictions and an i.i.d. assumption for the shock distributions. We consider two model specifications for λ_t . Firstly, in the classic specification, we suppose $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^*$ and use an expectation targeting approach to obtain ω by setting $\omega = 1 - \beta - \gamma$. Secondly, in the specification with asymmetric effect, we suppose $\lambda_t = \omega + \beta\lambda_{t-1} + \gamma\ell_{t-1}^* + \gamma^-\ell_{t-1}^*I_{R_{t-1}<0}$ where R_t is the return at time t . We further assume that conditional on $\mathcal{F}_{t-1}$, R_t has a zero median and is uncorrelated with ℓ_{t-1}^* . Therefore, it implies that $E[\ell_{t-1}^*I_{R_{t-1}<0} | \mathcal{F}_{t-1}] = \lambda_t/2$ and we can also use a targeting approach for ω by setting $\omega = 1 - \beta - \gamma - \gamma^-/2$.

4.8.2.1 Estimation based on conditional moment restrictions

We use the GMM approach based on the conditional moment restrictions to acquire an initial consistent estimators of the λ_t process parameters θ . We consider the minimalist case where the model is estimated using only the first conditional moment restriction, i.e. $E\left[\frac{\ell_t^*}{\lambda_t} - 1 | \mathcal{F}_{t-1}\right] = 0$. We further improve the estimates of the $g(t/T)$ function using the estimated $\hat{\lambda}_t = \hat{\lambda}_t(\hat{\theta}_{GMM})$ obtained in the previous step. This, in turn, allows us to further improve the estimates of the θ parameters.

We report the obtained estimates with associated t-statistics in [Table 4.3](#). It can be observed that the parameter estimates in the classic specification are almost always statistically significant at the 5% level. However, the γ and γ^- estimates in the asymmetric model specification are in general not significant. The overall lack of statistical significance indicates that the asymmetric effect does not contribute to improving the empirical fit of the model based on the first moment

estimated with a local linear estimator. Therefore, $g(t/T)$ is roughly moving around the mid-level of ℓ_t but this is not the case for the log illiquidity series as $\log g(t/T)$ is higher than the mean level of $\log \ell_t$ due to a Jensen's inequality effect. In the interest of space, we will focus on the plot of the log illiquidity series hereafter.

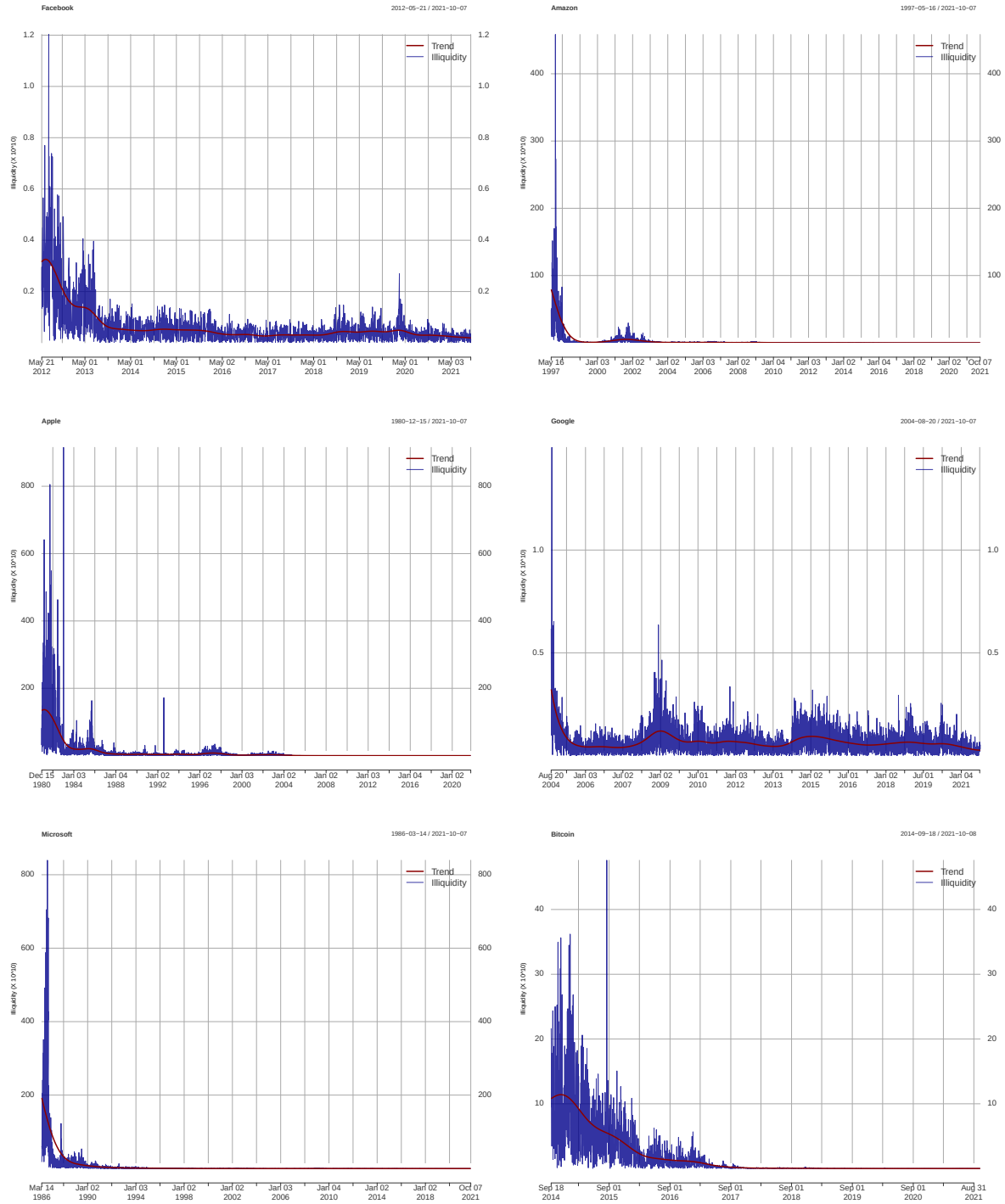


Figure 4.3: Fab 5 and Bitcoin illiquidity series and trend functions ($\times 10^{10}$).

restriction. We will further investigate whether including an asymmetric term is beneficial in the case where the models are estimated via the MLE approach under an i.i.d. shock assumption.

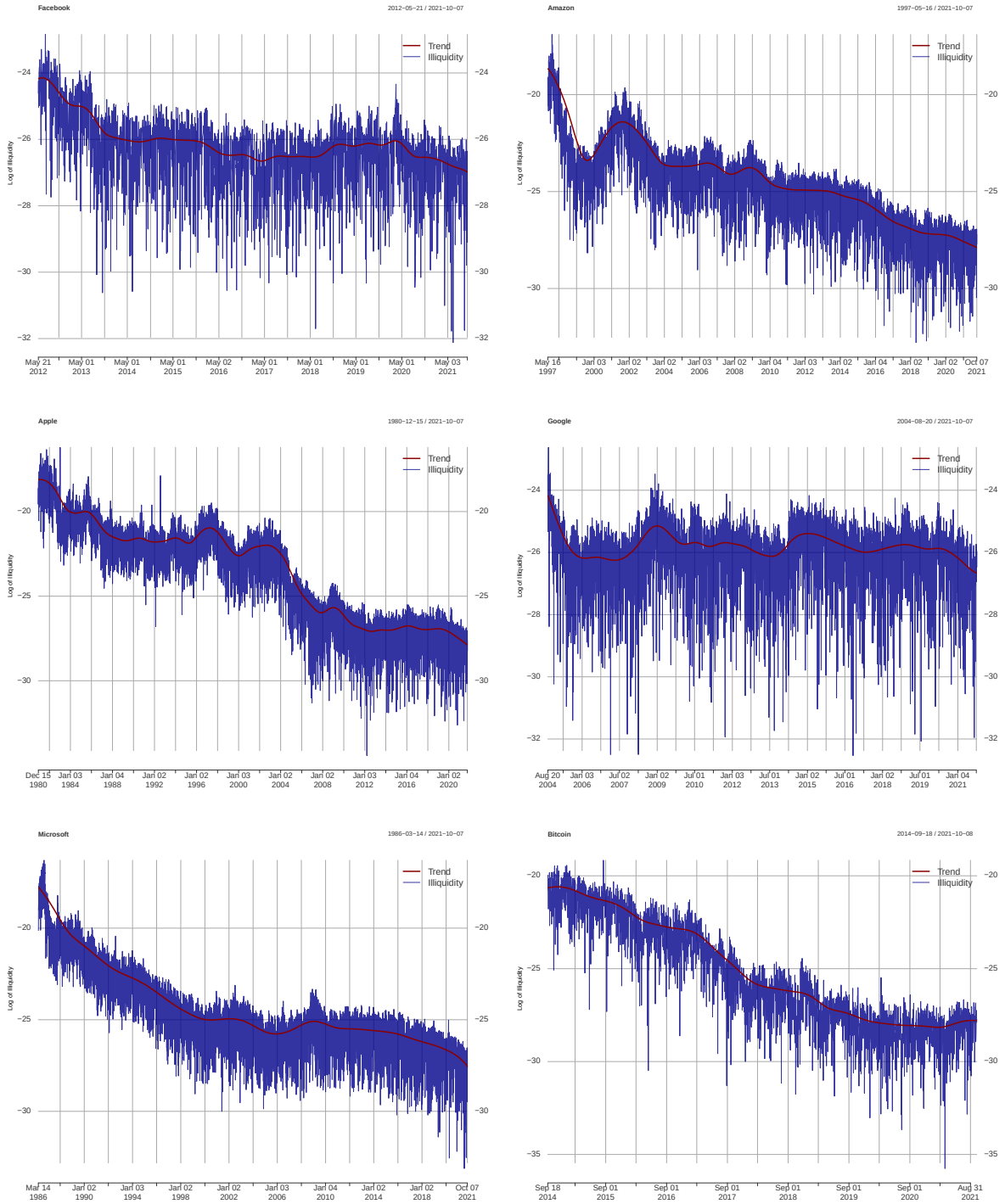


Figure 4.4: Fab 5 and Bitcoin log illiquidity series and trend functions.

Finally, the coefficient β is close to one, indicating high persistence in the short-run dynamics of the illiquidity series.

We improve the estimates of the trend function based on the estimated $\hat{\lambda}_t$ process. The log transforms of the initial and updated estimates of the trend function, i.e. $\log g(t/T)$, are plotted in Figure 4.7 of Appendix 4.E.1 together with the log illiquidity series. We observe that the updated trend function estimates – under both the classic model and the asymmetric specification – are different from the initial estimate but only to a minor extent. This observation indicates that a 2-step approach consisting in first using a local linear estimator for the trend function and then estimating the λ_t process and its associated parameters θ can be a viable option in empirical applications.

Table 4.3: Estimated parameters of the λ_t process based on first moment restriction.

	Classic		Asymmetric		
	β	γ	β	γ	$\gamma^{(-)}$
Facebook	0.951 (20.38)	0.029 (2.06)	0.951 (27.33)	0.029 (1.21)	-0.001 (-0.04)
Amazon	0.946 (71.35)	0.052 (4.47)	0.948 (59.21)	0.042 (1.52)	0.014 (0.47)
Apple	0.914 (60.56)	0.077 (6.75)	0.914 (69.36)	0.059 (2.34)	0.029 (0.84)
Google	0.967 (64.94)	0.028 (3.08)	0.964 (71.51)	0.021 (1.95)	0.010 (1.21)
Microsoft	0.945 (88.26)	0.051 (5.95)	0.859 (9.74)	0.016 (1.50)	0.142 (1.69)
Bitcoin	0.960 (74.04)	0.036 (3.56)	0.960 (21.68)	0.021 (0.19)	0.025 (0.12)

Note: The estimated parameters are $\theta = (\beta, \gamma)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)})$ for the asymmetric specification of λ_t . The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

4.8.2.2 Estimation: i.i.d. error term with parametric density

We estimate the model using an alternative approach – the semiparametric MLE approach – where we assume i.i.d. error terms. The conditional distribution of the error term ζ_t can be

freely chosen within the class of distributions satisfying the required characteristics, namely the density having non-negative support with unit mean and variance σ_ζ^2 . In this application, we assume that the error term follows a Weibull($\Gamma(1 + \varphi)^{-1}, \varphi$) distribution, with the parameter φ controlling the shape of the distribution.⁶ Based on the local linear estimator of the $g(t/T)$ function, we first obtain a consistent estimator of the λ_t process parameters via the Quasi-Maximum Likelihood (QML) estimation approach. This is achieved by maximizing the log likelihood assuming $g(t/T)$ is known and ζ_t is parametrically specified with a Weibull density. The fully efficient estimates can then be obtained by using a one-step update approach using the efficient scores based on the initial consistent estimators as introduced in Section 4.4.2.1.

We report the estimated parameters with the corresponding t-statistics in Table 4.4. The estimates for the parameters of the λ_t process are significant and all illiquidity series exhibit a high degree of persistence as the β coefficients are close to one. In addition, the estimated shape parameter of the Weibull distribution for the error terms are ranging from 1.14 to 1.40, indicating that the volatility of ζ_t is ranging from 0.73 to 0.88. This is consistent with the observation that the five tech stocks have comparable volatility levels while the Bitcoin asset has much higher volatility.

Furthermore, we provide diagnostics on the validity of our assumptions for the error term ζ_t . Concerning the i.i.d. assumption, we plot the autocorrelation function (ACF) of ζ_t in Figure 4.8 and Figure 4.9 of Appendix 4.E.2 respectively for the classic and asymmetric model specifications of λ_t . Similarly, we plot the ACF of ζ_t^2 under the two specifications in Figure 4.10 and Figure 4.11 of Appendix 4.E.2. We observe that in most of the cases, there is no evidence suggesting autocorrelation in the residual or squared residual series. Furthermore, we use the probability integral transformation (PIT) to check how well the assumed Weibull conditional distribution fits the data. The histogram plots of the PITs shown in Figure 4.12 and Figure 4.13 of Appendix 4.E.2 are quite close to a uniform distribution. All assets exhibit a common pattern where the error term has ticker tail on the left-hand side and thinner tail on the right-hand side compared to a Weibull distribution. This issue can be addressed using another more flexible distribution, for example the generalized Pareto distribution. However, we do not pursue this

⁶We consider Exponential, Gamma, and Weibull distributions in our analysis as they are commonly adopted in the literature. To save space, we focus on the Weibull distribution in the paper as it provides slightly better performance in terms of fit compared to the other two alternative distributions.

Table 4.4: Fully efficient estimates of the parameters for the λ_t process under the assumption that the error term ζ_t follows a Weibull distribution.

	Classic				Asymmetric				
	β	γ	φ	σ	β	γ	$\gamma^{(-)}$	φ	σ
Facebook	0.861 (5.98)	0.046 (4.94)	1.378 (804.81)	0.734	0.921 (33.88)	-0.028 (-5.77)	0.106 (9.06)	1.398 (720.95)	0.725
Amazon	0.916 (285.12)	0.080 (28.00)	1.367 (2204.62)	0.740	0.928 (346.05)	0.038 (13.18)	0.065 (21.72)	1.379 (2131.88)	0.734
Apple	0.876 (322.41)	0.095 (68.19)	1.309 (3151.33)	0.771	0.891 (360.82)	0.045 (49.13)	0.077 (27.71)	1.310 (2434.88)	0.770
Google	0.908 (31.83)	0.046 (8.18)	1.284 (5624.95)	0.785	0.932 (70.03)	-0.011 (-3.21)	0.088 (12.27)	1.302 (5351.82)	0.775
Microsoft	0.924 (361.55)	0.067 (33.24)	1.366 (1332.79)	0.740	0.927 (360.10)	0.038 (17.58)	0.052 (20.17)	1.370 (1217.23)	0.739
Bitcoin	0.892 (38.07)	0.061 (8.49)	1.137 (4774.77)	0.881	0.892 (36.48)	0.039 (6.64)	0.041 (7.15)	1.143 (4895.14)	0.877

Note: The estimated parameters are $\theta = (\beta, \gamma, \varphi)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)}, \varphi)$ for the asymmetric specification of λ_t . φ is the shape parameter of the Weibull distribution which has mean 1 and standard deviation σ_ζ of

$\sqrt{\frac{\Gamma(1+\frac{2}{\varphi})}{\Gamma^2(1+\frac{1}{\varphi})}} - 1$. The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

direction in our analysis and focus instead on investigating whether using a nonparametric density help improve the model fit.

We further improve the estimation of $g(t/T)$ by maximizing the local likelihood based on the estimated $\hat{\lambda}_t$ process and the error density. The log transforms of the initial and updated estimates of the trend function, i.e. $\log g(t/T)$, are plotted in [Figure 4.14](#) of [Appendix 4.E.2](#) together with the log illiquidity series. As in the GMM case (see [Section 4.8.2.1](#)), we observe that the updated trend function estimates are different from the initial estimate but only to a minor extent.

4.8.2.3 Estimation: i.i.d. error term with nonparametric density

We consider whether replacing the parametric assumption for the error density f with a nonparametric kernel estimator can further improve the fit of our model to empirical data. We plot in [Figure 4.15](#) and [Figure 4.16](#) of [Appendix 4.E.3](#) the estimated nonparametric density against the Weibull density using the shape parameter estimates from [Section 4.8.2.2](#). We observe that the estimated nonparametric density curves do not fall between the two standard deviation bands of the estimated Weibull densities, suggesting that the difference between the estimated parametric and nonparametric densities are statistically significant.

The estimated nonparametric density allows us to further improve the maximum likelihood estimation results for the λ_t process. We can obtain the fully efficient estimates in the nonparametric density case using the one-step update approach based on the efficient scores introduced in [Section 4.4.2.2](#). The estimates with associated t-statistics are reported in [Table 4.5](#) and the parameters are all statistically significant. Comparing the estimated values for the λ_t parameters reported in [Table 4.4](#) and [Table 4.5](#), we observe that the difference in the estimated parameter values between the parametric and nonparametric cases are overall quite small. This indicates that the QML estimation approach, combined with the one-step update based on the efficient scores to improve efficiency, provides rather accurate parameter estimates.

We further present the log likelihood computed using the parameter estimates obtained in the parametric and nonparametric cases in [Table 4.6](#). We conclude that the ML estimation approach assuming a Weibull distribution for the error term provides good estimation performance, but using a nonparametric estimator for the error density can further improve performance in terms of the likelihood (the only exception is Microsoft).

Table 4.5: Fully efficient estimates of the parameters for λ_t process when using the nonparametric estimates of the density of the error term ζ_t .

	Classic		Asymmetric		
	β	γ	β	γ	$\gamma^{(-)}$
Facebook	0.858 (7.71)	0.049 (5.39)	0.904 (104.54)	-0.019 (-5.42)	0.102 (11.51)
Amazon	0.916 (294.21)	0.079 (24.97)	0.928 (384.91)	0.039 (14.50)	0.058 (20.19)
Apple	0.895 (175.42)	0.083 (26.28)	0.907 (238.28)	0.039 (14.98)	0.067 (19.60)
Google	0.911 (49.55)	0.046 (9.07)	0.924 (210.14)	-0.007 (-3.72)	0.086 (15.67)
Microsoft	0.929 (425.62)	0.066 (33.04)	0.931 (504.54)	0.042 (22.95)	0.044 (18.30)
Bitcoin	0.902 (66.59)	0.064 (11.04)	0.900 (67.85)	0.050 (8.93)	0.025 (3.78)

Note: The estimated parameters are $\theta = (\beta, \gamma)$ for the classic specification and $\theta = (\beta, \gamma, \gamma^{(-)})$ for the asymmetric specification of λ_t . The numbers in parentheses are the t-statistics of the corresponding parameter estimates.

Table 4.6: Log-likelihood comparison between models using the parametric (Weibull) and nonparametric estimates of the ζ_t density.

	Weibull		Nonparametric		Difference	
	classic	asymmetric	classic	asymmetric	classic	asymmetric
Facebook	-2171.12	-2138.88	-2148.38	-2106.34	22.74	32.55
Amazon	-4947.87	-4898.92	-4908.41	-4848.90	39.46	50.01
Apple	-8860.69	-8790.35	-8664.33	-8579.91	196.36	210.44
Google	-4020.91	-3969.86	-3969.77	-3892.27	51.14	77.59
Microsoft	-7440.48	-7398.32	-7455.04	-7404.63	-14.56	-6.32
Bitcoin	-2397.82	-2389.82	-2384.46	-2379.91	13.36	9.91

Note: The numbers reported are in terms of $\log LL$. The difference is computed as $\log LL$ in nonparametric density case minus $\log LL$ in the parametric Weibull density case.

4.8.3 Testing for permanent shifts: discontinuity in g function

We test for a potential discontinuity at a given time u_0 by estimating the $g^\pm(u_0)$ functions via the local linear approach. We then construct the test statistics $\tau(u_0)$ to detect whether there is a permanent shift in the illiquidity level occurring at time u_0 . To facilitate the computation of the asymptotic variance of $\hat{g}^\pm(u_0)$, we work with the improved estimator obtained by smoothing out ℓ_t , i.e. $\ell_t/\hat{\lambda}_t$. We plot in [Figure 4.5](#) and [Figure 4.6](#) the test statistics $\tau(u_0)$ for the five tech stocks and the Bitcoin asset over the corresponding sample period.

We start with a typical stock specific-event, a stock split, and test for permanent shifts in the liquidity dynamics arising after stock splits. The five tech stocks we consider have quite different corporate policies regarding shareholders and in particular their propensity to split their stock differs. In our study, Facebook never split its stock, Amazon split its stock three times but the last time being before 2000 (perhaps coincidentally these were all in the pre-decimal era). Microsoft split its stock 9 times in our sample period but the last time was in 2002. Google split its stock twice in our sample in 2014 and 2015, but not before that or since. Apple is a regular splitter with 5 splits in our sample fairly evenly spaced in time. Each split is marked as a red dot on the curves in [Figure 4.5](#) and [Figure 4.6](#). The majority of the statistics on stock split dates is outside of the 5% critical value bands, suggesting an overall significance of the stock split events.

In addition, [Table 4.7](#) provides the average test statistics for each stock on their stock split dates together with the average across all stock split events for the four considered stocks. Firstly, we should note that the average test statistic τ is positive in all cases, indicating an increase in stock illiquidity and thus a corresponding decrease in stock liquidity after the splits. Secondly, we observe that the average statistic indicates a significant difference between pre- and post-split long-term trends of the illiquidity series. This suggests that the decrease in liquidity after stock splits is permanent and significant.

Table 4.7: Average statistics for testing permanent breaks in the liquidity series.

	Amazon	Apple	Google	Microsoft	Average
τ	6.710	4.109	3.904	3.328	4.128

To test for temporary effects of stock splits on the liquidity level, we need to normalize the illiquidity series using the estimated one-sided trend functions $\hat{g}^\pm(u)$. Once the detrended

illiquidity series are obtained, we use the consistent test developed in Section 4.6 to test the null hypothesis that $\alpha_1 = \alpha_2 = \dots = \alpha_J = 0$ against the general alternative semiparametric model with the assumption that the error terms follow a Weibull distribution. Here, we consider a five-day window, namely from two days before until two days after the stock split date. We report in Table 4.8 and Table 4.9 the test statistic values for the permanent (τ_{LR}) and temporary (τ_{SR}) shifts together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.⁷ We observe that the effect of stock splits on the short-term dynamics of liquidity is almost always not significant. The only two exceptions are for Microsoft stock splits in 1991 and 1996 which were associated with a significant short-term shift in liquidity dynamics. Therefore, our empirical evidence suggest overall that stock splits of tech companies had a significant permanent effect on the long-run trend of their illiquidity process but not on the short-run dynamics.

⁷Note that for the split events preceded by another one we only consider the period after the first stock split event for the computation of the τ_{SR} quantiles.

Table 4.8: Test statistics for detecting permanent and temporary breaks in the liquidity series of Amazon and Apple stocks.

		Split 1	Split 2	Split 3		
		Date	Date	Date		
Amazon	Splits	02:01	03:01	02:01		
	τ_{LR}	3.89	15.01	1.23		
	τ_{LR}^{AAPL}	5.17	7.75	1.33		
	τ_{SR}	13.56	8.39	-21.57		
	$Q_{SR}^{2.5\%}$	-30.50	-32.43	-39.32		
	$Q_{SR}^{97.5\%}$	50.62	67.06	3.82		
		Split 1	Split 2	Split 3	Split 4	Split 5
Apple	Date	1987-06-16	2000-06-21	2005-02-28	2014-06-09	2020-08-31
	Splits	02:01	02:01	02:01	07:01	04:01
	τ_{LR}	5.21	2.89	9.29	1.08	2.07
	τ_{LR}^{MSFT}	-8.51	2.51	8.65	1.95	0.09
	τ_{SR}	1.42	8.81	-2.24	3.85	11.24
	$Q_{SR}^{2.5\%}$	-17.42	-14.91	-11.43	-13.56	-10.88
	$Q_{SR}^{97.5\%}$	21.56	17.76	18.05	12.18	13.92

Note: We report the test statistic values for the permanent (τ_{LR}) and temporary (τ_{SR}) shifts together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.

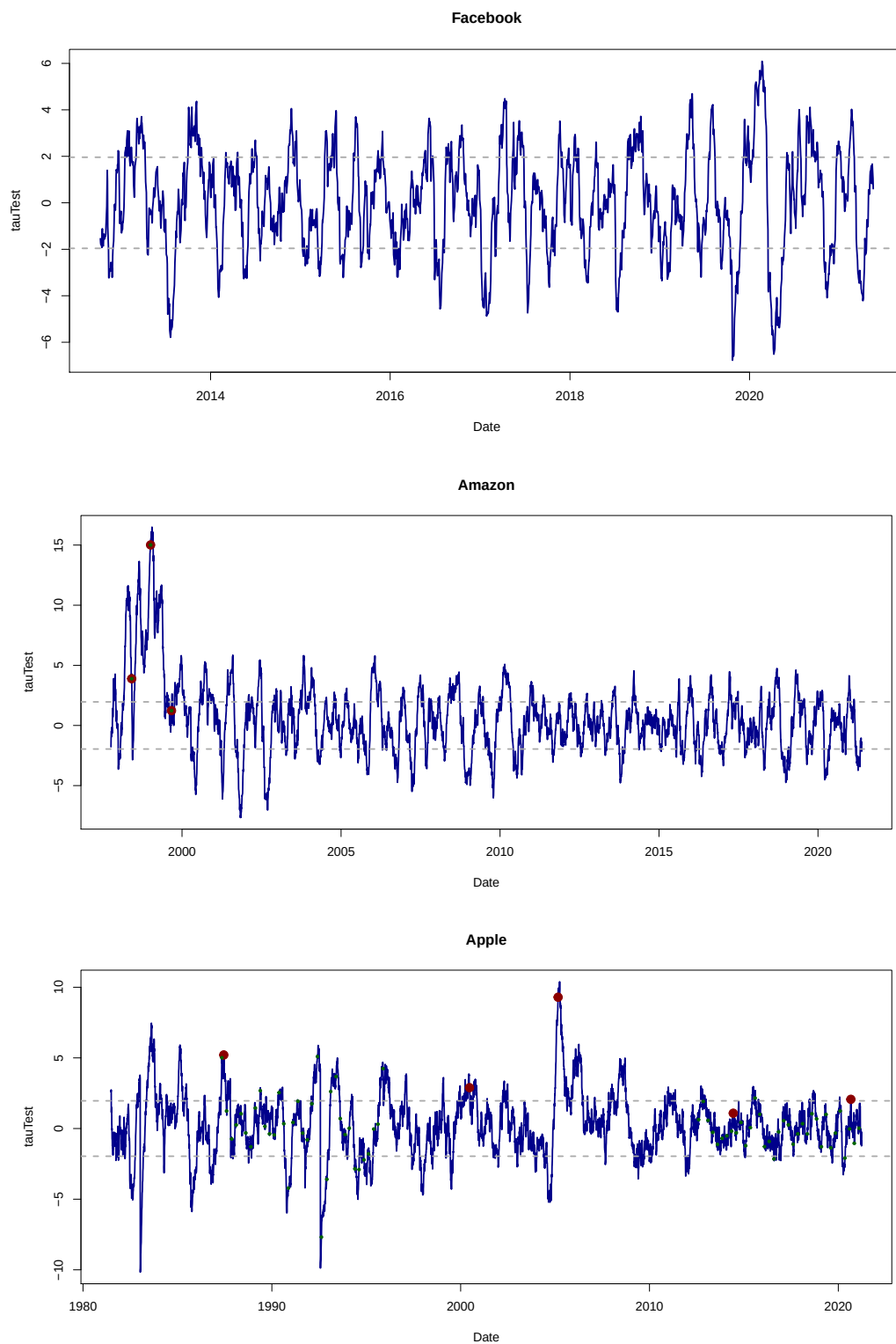


Figure 4.5: Test statistics for detecting permanent breaks in the illiquidity series – Facebook, Amazon, Apple.

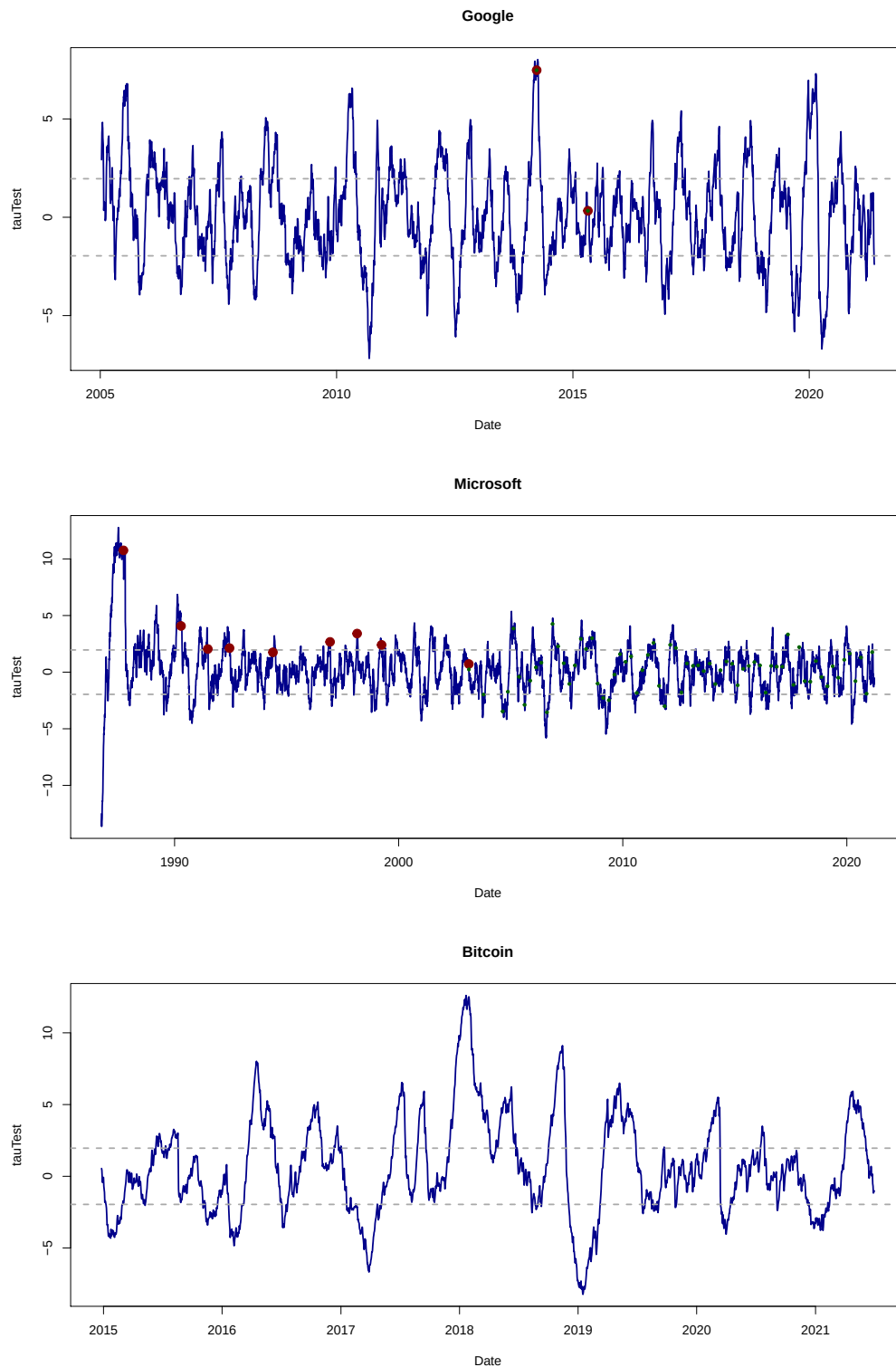


Figure 4.6: Test statistics for detecting permanent breaks in the illiquidity series – Google, Microsoft, Bitcoin.

Table 4.9: Test statistics for detecting permanent and temporary breaks in the liquidity series of Google and Microsoft stocks.

Google	Split 1		Split 2			
	Date	2014-03-27	2015-04-27			
	Splits	2002:1000	10027455:10000000			
	τ_{LR}	7.48	0.33			
	τ_{LR}^{MSFT}	7.59	-0.30			
	τ_{SR}	-2.04	9.16			
	$Q_{SR}^{2.5\%}$	-21.64	-14.37			
$Q_{SR}^{97.5\%}$	21.70	23.26				
Microsoft	Split 1		Split 2	Split 3	Split 4	Split 5
	Date	1987-09-21	1990-04-16	1991-06-27	1992-06-15	1994-05-23
	Splits	02:01	02:01	03:02	03:02	02:01
	τ_{LR}	10.76	4.08	2.04	2.12	1.75
	τ_{LR}^{AAPL}	7.93	2.92	-0.49	-5.53	2.78
	τ_{SR}	10.06	0.29	33.70	7.41	-4.09
	$Q_{SR}^{2.5\%}$	-34.27	-21.75	-16.82	-17.95	-17.87
	$Q_{SR}^{97.5\%}$	32.86	21.34	26.39	17.86	20.46
	Split 6		Split 7	Split 8	Split 9	
	Date	1996-12-09	1998-02-23	1999-03-29	2003-02-18	
	Splits	02:01	02:01	02:01	02:01	
	τ_{LR}	2.67	3.41	2.41	0.72	
	τ_{LR}^{AAPL}	-0.72	1.59	2.00	-0.21	
	τ_{SR}	26.11	6.05	10.08	12.43	
	$Q_{SR}^{2.5\%}$	-15.61	-16.43	-18.37	-20.18	
	$Q_{SR}^{97.5\%}$	17.46	17.70	16.61	25.56	

Note: We report the test statistic values for the permanent (τ_{LR}) and temporary (τ_{SR}) shifts together with the 2.5% and 97.5% quantiles of τ_{SR} which are estimated based on past data.

4.8.4 Risk premium

Amihud (2002) studies how illiquidity, captured by his illiquidity measure A_t introduced in Section 4.2, relates to stock excess returns in both the time series and cross-sectional dimensions. We build on this analysis to investigate the effect of each component of the S&P 500 index illiquidity process – i.e. the expected long-term and short-term components (respectively $g(t/T)$ and λ_t) and illiquidity shocks (ζ_t) – on the stock market index excess returns (the market “risk premium”). We consider three frequencies in our analysis – daily, weekly and monthly. The S&P 500 index illiquidity and log illiquidity series together with the stock market index return data for the three considered frequencies are plotted respectively in Figure 4.18, Figure 4.19 and Figure 4.20 of Appendix 4.E.4. We note that there exists a strong downward trend in the illiquidity process while the return series is somewhat stationary. This suggests that the relationship between the long-run trend of market liquidity and the stock excess return would be less significant.⁸ Therefore, we focus on detrended illiquidity and market excess return series to study the effect of expected short-run illiquidity variations and unexpected illiquidity shocks on the market risk premium.

We consider the specification from Equation (4.31) in Section 4.7 for the regression of detrended risk premium on illiquidity components.⁹ The estimation results for the three sampling frequencies considered are provided in Table 4.10. We observe that the estimated γ coefficients for the short-run expected illiquidity component λ_t are positive and significant which indicates that the expected market excess return is an increasing function of the short-run expected illiquidity process. This observation is consistent with the intuition that higher expected market illiquidity would make investors demand higher excess returns on stocks as a compensation for gaining exposure to this source of risk. Moreover, the estimated δ coefficients for the shock term ζ_t are negative and significant, suggesting that the unexpected market illiquidity has a negative effect on the stock excess return. This can be explained by the fact that stock prices would likely fall when illiquidity unexpectedly rises, thus decreasing expected returns.

⁸This is confirmed by regression results based on Equation (4.30) introduced in Section 4.7. The coefficient estimates for the parameter b associated with the long-run trend illiquidity component are not significant. Results are available from the authors upon request.

⁹The time-varying unconditional equity premium $m(t/T)$ is obtained via a local linear estimator.

Table 4.10: Coefficient estimates for regressions using daily, weekly and monthly observations.

	Daily	Weekly	Monthly
α	−0.0007 (0.0005)	−0.0044 (0.0032)	−0.0513** (0.0176)
γ	0.0017*** (0.0004)	0.0066* (0.0032)	0.0561** (0.0177)
δ	−0.0010*** (0.0001)	−0.0023*** (0.0006)	−0.0045* (0.0022)
R^2	0.0068	0.0073	0.0187
Adj. R^2	0.0066	0.0066	0.0161
Num. obs.	15387	2893	741

Note: We estimate the regression based on Equation (4.31): $R_{mt} - R_{ft} - m(t/T) = \alpha + \gamma \times \lambda_t + \delta \times \zeta_t + \epsilon_t$, where $m(t/T)$ is the time-varying unconditional equity premium. The significance level is indicated by *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

4.9 Conclusions

The motivation for this paper stems from the observation that financial market illiquidity dynamics across various asset classes are driven by both low-frequency and higher-frequency variations, which makes the stationarity assumption unreasonable for illiquidity modelling. We develop a class of a dynamic autoregressive models that captures the slow-varying long-term trend with a nonparametric component and the short-run variations in the illiquidity series with an autoregressive component. We provide estimation theory for the GMM estimator and the efficient semiparametric ML estimator for the parametric and nonparametric error density cases. An empirical application – using the five largest US technology stocks and the Bitcoin asset – demonstrates the good performance of our framework in capturing the salient features of illiquidity dynamics.

We further develop a methodology to detect the occurrence of permanent and temporary shifts in the illiquidity process at a given point in time. We apply this framework to study how stock splits affect liquidity dynamics. Clearly stock splits are only one of many events that seem to permanently shift the stock price, quarterly earnings announcements, new product releases,

and macroeconomic news are all known to have big effects on the prices and trading volumes of these stocks in particular. Nevertheless, we do find a significant negative effect of stock splits on the long-run trend level of liquidity around the time of the stock splits themselves, while the effect on the short-run illiquidity dynamics is not significant. Our results are broadly consistent with [Copeland \(1979\)](#).

Lastly, we investigate the link between stock market excess returns and the different components of illiquidity for the S&P 500 stock market index. We show that, while excess returns are an increasing function of the expected illiquidity component, unexpected illiquidity shocks decrease stock prices and returns. Our finding is consistent with the findings of [Amihud \(2002\)](#) based on his cruder methodology.

4.A Lemmas

Lemma 1. *We have $\widehat{g}(u) - g(u) = V_T(u) + B_T(u)$, where $B_T(u)$ is deterministic and*

$$\sup_{u \in [0,1]} |V_T(u)| = O_P \left(\sqrt{\frac{\log T}{Th}} \right), \quad \sup_{u \in [0,1]} |B_T(u)| = O(h^2).$$

Proof of Lemma 1. We write

$$\widehat{g}(u) - g(u) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) v_t + \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) - g(u) = V_T(u) + B_T(u), \quad (4.32)$$

where B_T is a deterministic bias term and V_T is a mean zero stochastic term. The result follows by standard arguments.

4.B Proof of main results

Proof of Theorem 1. From the expansion in Equation (4.32), we have $V_T(u) = \sum_{t=1}^T K_h(t/T - u) v_t/T$, and we may show that

$$\sqrt{Th} V_T(u) \implies N(0, \|K\|_2^2 g(u)^2 \sigma_v^2),$$

where σ_v^2 is the long run variance.

Proof of Theorem 2. First, note that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} - g(u) &= \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) \zeta_t - g(u) \\ &= \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) (\zeta_t - 1) \\ &\quad + \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) g(t/T) - g(u) \\ &= V_T^+(u) + B_T^+(u), \end{aligned}$$

where $V_T^+(u)$ is a mean zero stochastic term, whereas $B_T^+(u) = B_T(u)$ is the deterministic bias term. The term $V_T^+(u)$ has a MDS error term and satisfies the CLT

$$\sqrt{Th} V_T^+(u) \implies N \left(0, \|K\|_2^2 g(u)^2 \sigma_\zeta^2 \right).$$

We next show that this is the leading term.

We have

$$\begin{aligned}
\left| \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\widehat{\lambda}_t} - \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} \right| &= \left| -\frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} \frac{\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t}{\lambda_t(\widehat{\theta}, \widehat{g})} \right| \\
&\leq \max_{1 \leq t \leq T} \left| \frac{1}{\lambda_t(\widehat{\theta}, \widehat{g})} \right| \left| \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} (\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t) \right| \\
&\leq O_P(1) \times \left| \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{\ell_t}{\lambda_t} (\lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t) \right| \\
&= o_P(T^{-1/2} h^{1/2}),
\end{aligned}$$

because

$$\begin{aligned}
\lambda_t(\widehat{\theta}, \widehat{g}) &= \lambda_t(\theta_0, g_0) - \left| \lambda_t(\widehat{\theta}, \widehat{g}) - \lambda_t(\theta_0, g_0) \right| \\
&\geq \lambda_t(\theta_0, g_0) - o_P(1)
\end{aligned}$$

by the triangle inequality and the uniform convergence of $\widehat{g}$ given in Lemma 1.

Proof of Theorem 3. We apply [Chen et al. \(2003\)](#). The key thing is to determine the contribution of the nonparametric estimation to the variance of the score function.

We have

$$M_T(\theta, \widehat{g}) = \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, g_0) + \frac{1}{T} \sum_{t=1}^T z_{t-1} (\ell_t^* - \ell_t^* - (\widehat{\lambda}_t(\theta) - \lambda_t(\theta))).$$

We consider

$$\begin{aligned}
M_T(\theta, g) &= \frac{1}{T} \sum_{t=1}^T \rho_t(\theta, g), \quad \rho_t(\theta, g) = z_{t-1} \left(\frac{\ell_t}{g(t/T)} - \lambda_t(\theta, g) \right) \\
\rho_t(\theta_0, g_0) &= z_{t-1} \lambda_t (\zeta_t - 1) \\
\lambda_t(\theta, g) &= 1 - \beta - \gamma + \beta \lambda_{t-1} + \gamma \frac{\ell_{t-1}}{g((t-1)/T)} = \frac{1 - \beta - \gamma}{1 - \beta} + \gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{\ell_{t-j}}{g((t-j)/T)}.
\end{aligned}$$

We have

$$M(\theta, g) = \lim_{T \rightarrow \infty} E(M_T(\theta, g)).$$

We define

$$\Gamma_2(\theta, g_0) \circ (g - g_0) = \frac{\partial}{\partial \tau} M(\theta, g_0 + \tau(g - g_0)).$$

We have

$$\frac{\lambda_t(\theta, g_0 + \tau(g - g_0)) - \lambda_t(\theta, g_0)}{\tau} \simeq -\gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{\ell_{t-j}}{g_0((t-j)/T)} \frac{g((t-j)/T) - g_0((t-j)/T)}{g_0((t-j)/T)}$$

and so

$$\begin{aligned} \lim_{\tau \rightarrow 0} E \left[\frac{\lambda_t(\theta, g_0 + \tau(g - g_0)) - \lambda_t(\theta, g_0)}{\tau} \right] &= -\gamma \sum_{j=1}^{\infty} \beta^{j-1} \frac{g((t-j)/T) - g_0((t-j)/T)}{g_0((t-j)/T)} \\ &\simeq -\frac{g(t/T) - g_0(t/T)}{g_0(t/T)} \frac{\gamma}{1 - \beta} \end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{\frac{\ell_t}{g(t/T)} - \frac{\ell_t}{g_0(t/T)}}{\tau} &\simeq -\frac{\ell_t}{g_0(t/T)} \frac{g(t/T) - g_0(t/T)}{g_0(t/T)} \\ E \left(\frac{\frac{\ell_t}{g(t/T)} - \frac{\ell_t}{g_0(t/T)}}{\tau} \right) &\simeq -\frac{g(t/T) - g_0(t/T)}{g_0(t/T)} \end{aligned}$$

Therefore,

$$\begin{aligned} M_T(\theta, \hat{g}) &= M_T(\theta, g_0) + \Gamma_2(\theta_0, g_0) \circ (\hat{g} - g_0) \\ &= \frac{1}{T} \sum_{t=1}^T \left(\rho_t(\theta_0, g_0) + \frac{1 - \beta - \gamma}{1 - \beta} z_{t-1} \frac{\hat{g}(t/T) - g_0(t/T)}{g_0(t/T)} \right). \end{aligned}$$

We have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T z_{t-1} \frac{\hat{g}(t/T) - g_0(t/T)}{g_0(t/T)} &= \frac{1}{T} \sum_{t=1}^T z_{t-1} \frac{1}{T} \sum_{s=1}^T K_h(s/T - t/T) (\lambda_s \zeta_s - 1) + O(h^2) \\ &= \frac{1}{T} \sum_{s=1}^T (\lambda_s \zeta_s - 1) \frac{1}{T} \sum_{t=1}^T z_{t-1} K_h(s/T - t/T) \\ &\simeq \frac{1}{T} \sum_{s=1}^T (\lambda_s \zeta_s - 1) E(z_{s-1}). \end{aligned}$$

It follows that

$$M_T(\theta, \hat{g}) = \frac{1}{T} \sum_{t=1}^T \left(\lambda_t(\zeta_t - 1) z_{t-1} + \frac{1 - \beta - \gamma}{1 - \beta} (\lambda_t \zeta_t - 1) E(z_{t-1}) \right) + o_P(T^{-1/2}).$$

Proof of Theorem 4. Let $\ell_t^{**} = \ell_t / \lambda_t$ then

$$\ell_t^{**} = g(t/T) \zeta_t.$$

The local likelihood is apart from a constant

$$L(g; u) = \sum_{t=1}^T K_h(t/T - u) \left(-\log g + \log f \left(\frac{\ell_t^{**}}{g} \right) \right).$$

Let $s_1(\zeta) = -f'_0(\zeta)/f_0(\zeta)$ and $s_2(\zeta) = -((f'_0(\zeta)/f_0(\zeta))\zeta + 1)$. We have

$$\frac{\partial L(g; u)}{\partial g} = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \left(-\frac{1}{g} - \frac{1}{g} \frac{f' \left(\frac{\ell_t^{**}}{g} \right)}{f \left(\frac{\ell_t^{**}}{g} \right)} \frac{\ell_t^{**}}{g} \right) = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{1}{g} s_2 \left(\frac{\ell_t^{**}}{g} \right)$$

$$\frac{\partial^2 L(g; u)}{\partial g^2} = \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \left(\frac{-1}{g^2} s_2 \left(\frac{\ell_t^{**}}{g} \right) - \frac{1}{g^2} s_2' \left(\frac{\ell_t^{**}}{g} \right) \frac{\ell_t^{**}}{g} \right)$$

We have

$$\begin{aligned} \frac{\partial L(g_0(u); u)}{\partial g} &= \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{1}{g(u)} s_2(\zeta_t) \\ \frac{\partial^2 L(g_0(u); u)}{\partial g^2} &\simeq \frac{1}{T} \sum_{t=1}^T K_h(t/T - u) \frac{1}{g(u)^2} s_2'(\zeta_t) \zeta_t. \end{aligned}$$

We have $s_2(\zeta) = 1 + \zeta f'(\zeta) / f(\zeta)$

$$E(s_2'(\zeta_t)) = \int s_2'(\zeta) \zeta f(\zeta) d\zeta = - \int s_2(\zeta) \zeta f'(\zeta) d\zeta - \int s_2(\zeta) f(\zeta) d\zeta = - \int s_2^2(\zeta) f(\zeta) d\zeta = -I_2(f).$$

4.C Semiparametric efficiency

4.C.1 Known f

Suppose that

$$\ell_t = g_\delta(t/T) \lambda_t(\theta) \zeta_t$$

$$\lambda_t = 1 - \beta - \gamma + \beta \lambda_{t-1} + \gamma \lambda_{t-1} \zeta_{t-1}$$

where ζ_t is i.i.d. with mean one and density f supported on $\mathbb{R}_+$, so that $E(\lambda_t) = 1$ and $E(\zeta_t) = 1$.

We suppose that g is unknown but we consider the parameterization by δ . We first suppose that f is known. Let $s_2(\zeta) = -(1 + \zeta f'(\zeta) / f(\zeta))$.

Consider the log likelihood

$$\begin{aligned} L(\theta, \delta | \ell_1, \dots, \ell_T) &= - \sum_{t=1}^T \log \lambda_t(\theta, \delta) - \sum_{t=1}^T \log g_\delta(t/T) + \sum_{t=1}^T \log f(\zeta_t(\theta, \delta)) \\ \lambda_t(\theta, \delta) &= 1 - \beta - \gamma + \beta \lambda_{t-1}(\theta, \delta) + \gamma \frac{\ell_{t-1}}{g_\delta((t-1)/T)}, \end{aligned} \tag{4.33}$$

$$\zeta_t(\theta, \delta) = \frac{\ell_t}{\lambda_t(\theta, \delta) g_\delta(t/T)}. \tag{4.34}$$

Note that λ_t depends implicitly on δ .

We have (at the true values)

$$\frac{\partial \zeta_t(\theta, \delta)}{\partial \theta} = \frac{-\ell_t}{\lambda_t(\theta, \delta) g_\delta(t/T)} \frac{\partial \log \lambda_t}{\partial \theta} = -\zeta_t \frac{\partial \log \lambda_t}{\partial \theta}.$$

$$\begin{aligned}\frac{\partial \zeta_t(\theta, \delta)}{\partial \delta} &= \frac{-\ell_t}{\lambda_t(\theta, \delta)g_\delta(t/T)} \left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta} \right) \\ &= -\zeta_t \left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta} \right).\end{aligned}$$

The score functions are

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^T \frac{f'}{f}(\zeta_t) \frac{\partial \zeta_t}{\partial \theta} - \frac{\partial \log \lambda_t}{\partial \theta} = \sum_{t=1}^T s_2(\zeta_t) \frac{\partial \log \lambda_t}{\partial \theta}.$$

Furthermore,

$$\frac{\partial \lambda_t(\theta, \delta)}{\partial \beta} = \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \beta} + \lambda_t(\theta, \delta) - 1$$

$$(1 - \beta L) \frac{\partial \lambda_t(\theta, \delta)}{\partial \beta} = \lambda_t - 1 = \beta(\lambda_{t-1} - 1) + \gamma u_{t-1}, \quad u_{t-1} = \lambda_{t-1} \zeta_{t-1} - 1$$

$$\frac{\partial \log \lambda_t(\theta, \delta)}{\partial \beta} = \frac{(1 - \beta L)^{-1} \beta(\lambda_{t-1} - 1) + \gamma (1 - \beta L)^{-1} u_{t-1}}{\lambda_t}$$

$$\frac{\partial \lambda_t(\theta, \delta)}{\partial \gamma} = \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \gamma} + u_{t-1}$$

$$(1 - \beta L) \frac{\partial \lambda_t(\theta, \delta)}{\partial \gamma} = u_{t-1}$$

$$\frac{\partial \log \lambda_t(\theta, \delta)}{\partial \gamma} = \frac{(1 - \beta L)^{-1} u_{t-1}}{\lambda_t}.$$

Here, L is the lag operator. We next consider the score wrt δ ,

$$\frac{\partial \lambda_t(\theta, \delta)}{\partial \delta} = \beta \frac{\partial \lambda_{t-1}(\theta, \delta)}{\partial \delta} - \gamma \frac{\ell_{t-1}}{g_\delta((t-1)/T)} \frac{\partial g_\delta((t-1)/T)/\partial \delta}{g_\delta((t-1)/T)}.$$

Therefore,

$$(1 - \beta L) \frac{\partial \lambda_t(\theta, \delta)}{\partial \delta} = -\gamma \lambda_{t-1} \zeta_{t-1} \frac{\partial \log g_\delta((t-1)/T)}{\partial \delta}$$

$$\frac{\partial \log \lambda_t(\theta, \delta)}{\partial \delta} \simeq -\gamma \frac{\partial \log g_\delta(t/T)}{\partial \delta} \frac{(1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t}.$$

The latter argument follows essentially because for a summable sequence $\{\psi_j\}$ and smooth function g we have

$$\sum_{j=1}^T \psi_j g\left(\frac{t-j}{T}\right) = g\left(\frac{t}{T}\right) \sum_{j=1}^T \psi_j - \frac{1}{T} \sum_{j=1}^T \psi_j j g'\left(\frac{s^*(t,j)}{T}\right) \simeq g\left(\frac{t}{T}\right) \sum_{j=1}^T \psi_j.$$

Therefore,

$$\begin{aligned} \frac{\partial L}{\partial \delta} &= \sum_{t=1}^T \frac{f'}{f}(\zeta_t) \frac{\partial \zeta_t}{\partial \delta} - \frac{\partial \log \lambda_t}{\partial \delta} - \frac{\partial \log g_\delta(t/T)}{\partial \delta} \\ &= \sum_{t=1}^T s_2(\zeta_t) \left(\frac{\partial \log \lambda_t}{\partial \delta} + \frac{\partial \log g_\delta(t/T)}{\partial \delta} \right) \end{aligned} \quad (4.35)$$

$$\begin{aligned} &= \sum_{t=1}^T s_2(\zeta_t) \frac{\partial \log g_\delta(t/T)}{\partial \delta} \left(1 - \gamma \frac{(1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t} \right) \\ &= \frac{1 - \beta - \gamma}{1 - \beta} \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \frac{\partial \log g_\delta(t/T)}{\partial \delta}, \end{aligned} \quad (4.36)$$

since

$$\begin{aligned} 1 - \gamma \frac{(1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t} &= \frac{\lambda_t - \gamma (1 - \beta L)^{-1} \lambda_{t-1} \zeta_{t-1}}{\lambda_t} \\ &= \frac{\lambda_t - \frac{\gamma}{1 - \beta} - \gamma (1 - \beta L)^{-1} u_{t-1}}{\lambda_t} \\ &= \frac{1}{\lambda_t} \times \frac{1 - \beta - \gamma}{1 - \beta} \end{aligned}$$

$$(1 - \beta L) \lambda_t = 1 - \beta - \gamma + \gamma \lambda_{t-1} \zeta_{t-1} = 1 - \beta + \gamma u_{t-1}$$

$$\lambda_t = 1 + \gamma (1 - \beta L)^{-1} u_{t-1}.$$

Therefore, the tangent space for g consists of functions of the form

$$\mathbb{T}_g = \left\{ \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} h(t/T) : h \in L_2[0, 1] \right\}. \quad (4.37)$$

That is, the score wrt g is of the form

$$\frac{\partial L}{\partial \delta} = \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} h(t/T)$$

for some function $h(\cdot)$. Because of the presence of $1/\lambda_t$ this is different from our volatility paper...

The efficient score function L_θ^* for θ in the presence of unknown g (but known φ) is the residual from the projection of L_θ onto the tangent space $\mathbb{T}_g$, this is

$$\begin{aligned} L_\theta^* &= \sum_{t=1}^T s_2(\zeta_t) \left(\frac{\partial \log \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \frac{1}{\lambda_t} \right) \\ &= \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right). \end{aligned} \quad (4.38)$$

This can be verified as for any element of $\mathbb{T}_g$ (indexed by $h(\cdot)$) we have

$$\begin{aligned}
& E \left[s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \times s_2(\zeta_t) \frac{1}{\lambda_t} h(t/T) \right] \\
&= E \left[s_2^2(\zeta_t) \right] E \left[\frac{1}{\lambda_t^2} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right] h(t/T) \\
&= E \left[s_2^2(\zeta_t) \right] \left[E \left(\frac{1}{\lambda_t^2} \frac{\partial \lambda_t}{\partial \theta} \right) - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} E \left(\frac{1}{\lambda_t^2} \right) \right] h(t/T) \\
&= 0
\end{aligned}$$

for any $h(\cdot)$.

4.C.2 Parametric f

Now suppose that $f = f_\varphi$, where φ is unknown and has to be estimated. Then

$$L(\theta, \varphi, \delta | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \log \lambda_t(\theta, \delta) - \sum_{t=1}^T \log g_\delta(t/T) + \sum_{t=1}^T \log f_\varphi(\zeta_t(\theta, \delta)),$$

where f_φ is a density function that imposes through its parameterization the unit mean assumption.

We have

$$\frac{\partial L}{\partial \varphi} = \sum_{t=1}^T \frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)}.$$

We have

$$E \left(\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \delta} \right) = \sum_{t=1}^T E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right) h(t/T).$$

We have

$$\begin{aligned}
E \left(\frac{\partial f_\varphi(\zeta_t) / \partial \varphi}{f_\varphi(\zeta_t)} s_2(\zeta_t) \right) &= \int \frac{\partial f_\varphi(\zeta) / \partial \varphi}{f_\varphi(\zeta)} s_2(\zeta) f_\varphi(\zeta) d\zeta \\
&= \int s_2(\zeta) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta \\
&= - \int \left(1 + \zeta \frac{f'_\varphi(\zeta)}{f_\varphi(\zeta)} \right) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta \\
&= - \int \zeta \left(\frac{f'_\varphi(\zeta)}{f_\varphi(\zeta)} - 1 \right) \frac{\partial f_\varphi(\zeta)}{\partial \varphi} d\zeta.
\end{aligned}$$

Therefore,

$$E \left(\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \delta} \right) = \sum_{t=1}^T E \left(\frac{\partial f_{\varphi}(\zeta_t) / \partial \varphi}{f_{\varphi}(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right) h(t/T) \neq 0$$

for any parameterization of g and so the efficient score function for φ in the presence of unknown g is

$$L_{\varphi}^* = \sum_{t=1}^T \left(\frac{\partial f_{\varphi}(\zeta_t) / \partial \varphi}{f_{\varphi}(\zeta_t)} - \frac{E \left(\frac{\partial f_{\varphi}(\zeta_t) / \partial \varphi}{f_{\varphi}(\zeta_t)} s_2(\zeta_t) \right) E \left(\frac{1}{\lambda_t} \right)}{I_2(f) E \left(\frac{1}{\lambda_t^2} \right)} s_2(\zeta_t) \frac{1}{\lambda_t} \right). \quad (4.39)$$

A special case where φ is a scale parameter, i.e.,

$$f_{\varphi}(\zeta) = \frac{1}{\varphi} f_0 \left(\frac{\zeta}{\varphi} \right).$$

We have

$$\begin{aligned} f'_{\varphi}(\zeta) &= \frac{1}{\varphi^2} f'_0 \left(\frac{\zeta}{\varphi} \right), \\ s_2(\zeta) &= - \left(1 + \frac{f'_{\varphi}(\zeta)}{f_{\varphi}(\zeta)} \zeta \right) = - \left(1 + \frac{f'_0 \left(\frac{\zeta}{\varphi} \right)}{f_0 \left(\frac{\zeta}{\varphi} \right)} \left(\frac{\zeta}{\varphi} \right) \right) = - \left(1 + \frac{f'_0(z)}{f_0(z)} z \right), \end{aligned}$$

where $z = \zeta/\varphi$. In that case

$$\begin{aligned} \frac{\partial f_{\varphi}(\zeta_t)}{\partial \varphi} &= \frac{-1}{\varphi^2} f_0 \left(\frac{\zeta}{\varphi} \right) + \frac{1}{\varphi} f'_0 \left(\frac{\zeta}{\varphi} \right) \frac{\zeta}{\varphi^2} \\ \frac{\partial f_{\varphi}(\zeta_t) / \partial \varphi}{f_{\varphi}(\zeta_t)} &= \frac{\frac{-1}{\varphi^2} f_0 \left(\frac{\zeta}{\varphi} \right) + \frac{1}{\varphi} f'_0 \left(\frac{\zeta}{\varphi} \right) \frac{\zeta}{\varphi^2}}{\frac{1}{\varphi} f_0 \left(\frac{\zeta}{\varphi} \right)} \\ &= \frac{-1}{\varphi} \left(1 + \frac{f'_0 \left(\frac{\zeta}{\varphi} \right) \frac{\zeta}{\varphi}}{f_0 \left(\frac{\zeta}{\varphi} \right)} \right) \\ &= \frac{-1}{\varphi} \left(1 + \frac{f'_0(z) z}{f_0(z)} \right) \\ &= \frac{1}{\varphi} s_2(\zeta). \end{aligned}$$

Therefore,

$$\frac{\partial L}{\partial \varphi} = \sum_{t=1}^T \frac{1}{\varphi} s_2(\zeta_t)$$

and

$$\begin{aligned} E \left[\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \theta} \right] &= \frac{1}{\varphi} I_2(f) E \left[\frac{\partial \log \lambda_t}{\partial \theta} \right] \\ E \left[\frac{\partial L}{\partial \varphi} \frac{\partial L}{\partial \delta} \right] &= \frac{1}{\varphi} I_2(f) E \left[\frac{1}{\lambda_t} \right]. \end{aligned}$$

In this case, the efficient score function for φ is

$$L_{\varphi}^* = \sum_{t=1}^T \frac{1}{\varphi} s_2(\zeta_t) \left(1 - \frac{E\left(\frac{1}{\lambda_t}\right)}{E\left(\frac{1}{\lambda_t^2}\right)} \frac{1}{\lambda_t} \right).$$

4.C.3 Unknown f

We next consider the semiparametric case where f is of unknown form but has unit mean. In this case for any parameterization φ

$$\begin{aligned} \int f_{\varphi}(\zeta) d\zeta = 1 &\implies \int \frac{\partial f_{\varphi}(\zeta)/\partial \varphi}{f_{\varphi}(\zeta)} f_{\varphi}(\zeta) d\zeta = 0 \\ \int \zeta f_{\varphi}(\zeta) d\zeta = 1 &\implies \int \zeta \frac{\partial f_{\varphi}(\zeta)/\partial \varphi}{f_{\varphi}(\zeta)} f_{\varphi}(\zeta) d\zeta = 0. \end{aligned}$$

According to Drost and Werker, the tangent space for f consists of functions τ that satisfy

$$\mathbb{T}_f = \left\{ \tau(\cdot) : \int \zeta^j \tau(\zeta) f(\zeta) d\zeta = 0, \quad j = 0, 1. \right\}$$

If we project

$$L_{\theta}^* = \sum_{t=1}^T s_2(\zeta_t) \frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E\left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)} \right)$$

on to the tangent space for f we have

$$L_{\theta}^{**} = \sum_{t=1}^T \left(s_2(\zeta_t) + \frac{1 - \zeta_t}{\sigma_{\zeta}^2} \right) E \left[\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E\left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2}\right]}{E\left(\frac{1}{\lambda_t^2}\right)} \right) \right], \quad (4.40)$$

because

$$s_2(\zeta) + \frac{1 - \zeta}{\sigma_{\zeta}^2} \in \mathbb{T}_f$$

since

$$E \left(s_2(\zeta) + \frac{1 - \zeta}{\sigma_{\zeta}^2} \right) = 0, \quad E \left(\zeta s_2(\zeta) + \frac{\zeta - \zeta^2}{\sigma_{\zeta}^2} \right) = 0.$$

Since

$$\begin{aligned} - \int \zeta s_2(\zeta) f(\zeta) d\zeta &= \int \zeta f(\zeta) d\zeta + \int \zeta^2 \frac{f'(\zeta)}{f(\zeta)} f(\zeta) d\zeta \\ &= 1 + \int \zeta^2 f'(\zeta) d\zeta \\ &= 1 - 2 \int \zeta f(\zeta) d\zeta \\ &= -1 \end{aligned}$$

by integration by parts. Furthermore, $E(\zeta^2) = 1 + \sigma_\zeta^2$ and the result follows.

Therefore, the efficient score function in the model where both g, f are unknown is

$$\begin{aligned} L_\theta^* &= \sum_{t=1}^T s_2(\zeta_t) \left(\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) - E \left[\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right] \right) \\ &\quad + \sum_{t=1}^T \frac{\zeta_t - 1}{\sigma_\zeta^2} E \left[\frac{1}{\lambda_t} \left(\frac{\partial \lambda_t}{\partial \theta} - \frac{E \left[\frac{\partial \lambda_t}{\partial \theta} \frac{1}{\lambda_t^2} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right] \\ &= \sum_{t=1}^T s_2(\zeta_t) \left(\left(\frac{\partial \log \lambda_t}{\partial \theta} - E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \left(\frac{1}{\lambda_t} - E \left(\frac{1}{\lambda_t} \right) \right) \frac{E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right) \right) \\ &\quad + \sum_{t=1}^T \frac{\zeta_t - 1}{\sigma_\zeta^2} \left(E \left(\frac{\partial \log \lambda_t}{\partial \theta} \right) - \frac{E \left(\frac{1}{\lambda_t} \right) E \left[\frac{\partial \log \lambda_t}{\partial \theta} \frac{1}{\lambda_t} \right]}{E \left(\frac{1}{\lambda_t^2} \right)} \right). \end{aligned}$$

Let $\theta = (\beta, \gamma)$ and $\eta = (\theta^\top, \sigma_\zeta^2)^\top$. The log likelihood function for η based on known f, g is

$$\begin{aligned} L(\eta | \ell_1, \dots, \ell_T) &= - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \ln g(t/T) - \frac{T}{2} \ln \sigma_\zeta^2 - \sum_{t=1}^T \ln f_0(z_t) \\ z_t &= \frac{\frac{\ell_t}{\lambda_t(\theta)g(t/T)} - 1}{\sigma_\zeta} = \frac{\ell_t - \lambda_t(\theta)g(t/T)}{\sigma_\zeta \lambda_t(\theta)g(t/T)}. \end{aligned}$$

Let $s_1(z) = -f'_0(z)/f_0(z)$ and $s_2(z) = -((f'_0(z)/f_0(z))z + 1)$, then

$$\begin{aligned} \frac{\partial L}{\partial \theta} &= \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta)}{\partial \theta} s_2(z_t) \\ \frac{\partial L}{\partial \sigma_\zeta^2} &= \sum_{t=1}^T s_2(z_t). \end{aligned}$$

We have

$$E \left(\frac{\partial L}{\partial \eta} \frac{\partial L}{\partial \eta^\top} \right) = I_2(f)M, \quad M = \begin{pmatrix} E \left(\frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} \right) & E \left(\frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \right) \\ E \left(\frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} \right) & 1 \end{pmatrix}.$$

The efficient score for θ in the presence of unknown σ_ζ^2 but known g, f_0 is

$$\frac{\partial L^*}{\partial \theta} = \sum_{t=1}^T \left(\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} - E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] \right) s_2(z_t).$$

It follows that the asymptotic variance of the MLE procedure based on known f, g is

$$\sqrt{T}(\hat{\eta} - \eta) \implies N \left(0, I_2(f)^{-1} M^{-1} \right).$$

We can estimate $M, I_2(f)$ by

$$\widehat{M} = \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} & \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta} \\ \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta_0)}{\partial \theta^\top} & 1 \end{pmatrix}$$

$$I_2(f) = \frac{1}{T} \sum_{t=1}^T \widehat{s}_2^2(z_t), \quad \widehat{s}_2(z) = - \left(\frac{\widehat{f}'(z)}{\widehat{f}(z)} z + 1 \right).$$

We next show that the score function $\partial L^*/\partial \theta$ is orthogonal to the scores for f, g so that asymptotically, the semiparametric procedure with estimated f, g is first order equivalent to the MLE with known f, g .

We suppose that

$$\ln g(u) = \sum_{j=0}^{\infty} \psi_j H_j(u)$$

$$\ln f_0(v) = \sum_{j=0}^{\infty} \alpha_j \Gamma_j(v).$$

for some basis functions H_j, Γ_j .

Then

$$L(\theta, \eta; \psi, \alpha | \ell_1, \dots, \ell_T) = - \sum_{t=1}^T \ln \lambda_t(\theta) - \sum_{t=1}^T \sum_{j=0}^M \psi_j H_j(t/T) - \sum_{t=1}^T \sum_{j=0}^N \alpha_j \Gamma_j(z_t)$$

$$z_t = \frac{\ell_t}{\lambda_t(\theta) \exp \left(\sum_{j=0}^M \psi_j H_j(t/T) \right)}.$$

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta)}{\partial \theta} s_2(z_t)$$

$$\frac{\partial L}{\partial \psi} = \sum_{t=1}^T H(t/T) s_2(z_t)$$

$$\frac{\partial L}{\partial \alpha} = - \sum_{t=1}^T \Gamma(z_t),$$

where $H = (H_0, \dots, H_M)^\top$ and $\Gamma = (\Gamma_0, \dots, \Gamma_N)^\top$ and $s_2(z_t) = \sum_{j=0}^N \alpha_j \Gamma_j'(z_t) z_t - 1$.

We have

$$\frac{1}{T} E \left[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \theta^\top} \right] = E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \frac{\partial \ln \lambda_t(\theta)}{\partial \theta^\top} \right] I_2(f)$$

$$\frac{1}{T} E \left[\frac{\partial L}{\partial \psi} \frac{\partial L}{\partial \psi^\top} \right] = \int H(u) H(u)^\top du I_2(f) = I_M \times I_2(f)$$

$$\frac{1}{T} E \left[\frac{\partial L}{\partial \alpha} \frac{\partial L}{\partial \alpha^\top} \right] = E \left[\Gamma(z_t) \Gamma(z_t)^\top \right]$$

$$\begin{aligned}
\frac{1}{T} E \left[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \psi^\top} \right] &= E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] \int H(u)^\top du I_2(f) = I_2(f) E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] (1, 0, \dots, 0) \\
\frac{1}{T} E \left[\frac{\partial L}{\partial \theta} \frac{\partial L}{\partial \alpha^\top} \right] &= E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] E \left[s_2(z_t) \Gamma(z_t)^\top \right] \\
\frac{1}{T} E \left[\frac{\partial L}{\partial \psi} \frac{\partial L}{\partial \alpha^\top} \right] &= (1, 0, \dots, 0)^\top E \left[s_2(z_t) \Gamma(z_t)^\top \right].
\end{aligned}$$

The efficient score of θ wrt ψ is

$$s_{\theta|\psi}^* = \sum_{t=1}^T \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} - E \left(\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right) \right] s_2(z_t)$$

and this score function is orthogonal to $\partial L / \partial \alpha$.

The efficient score of θ wrt α is

$$s_{\theta|\alpha}^* = \sum_{t=1}^T \frac{\partial \ln \lambda_t(\theta)}{\partial \theta} s_2(z_t) - E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] E \left[s_2(z_t) \Gamma(z_t)^\top \right] \left(E \left[\Gamma(z_t) \Gamma(z_t)^\top \right] \right)^{-1} \times \sum_{t=1}^T \Gamma(z_t)$$

and this score function is orthogonal to $\partial L / \partial \alpha$.

It follows that the efficient score for θ in the presence of unknown σ_ζ^2 but known g, f_0 is

$$\frac{\partial L^*}{\partial \theta} = \sum_{t=1}^T \left(\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} - E \left[\frac{\partial \ln \lambda_t(\theta)}{\partial \theta} \right] \right) s_2(z_t).$$

4.D Risk premium regressions

We consider the population regression model

$$R_{mt} - g(t/T) = a + b\lambda_t + c\zeta_t + \varepsilon_t,$$

where in practice we replace $g(\cdot)$ by $\hat{g}(\cdot)$ and λ_t, ζ_t by $\lambda_t(\hat{\theta}), \zeta_t(\hat{\theta})$. This does not affect consistency but does affect the limiting distribution and hence standard errors. The dependent variable effect takes care of itself, the main issue is around the estimated covariate. We argue as follows. Suppose that

$$y_t = \beta^\top x_t(\theta_0) + \varepsilon_t.$$

By a Taylor expansion we have

$$x_t(\hat{\theta}) \simeq x_t(\theta_0) + \frac{\partial x_t(\theta_0)}{\partial \theta} (\hat{\theta} - \theta_0).$$

We also have an expansion for our estimators of the form

$$\hat{\theta} - \theta_0 = \frac{1}{T} \sum_{t=1}^T \psi_t(\theta_0) + e_t,$$

where $\psi_t(\theta_0)$ are mean zero and the sum satisfies a CLT, while e_t is of smaller order. Let

$$x_t^* = x_t(\hat{\theta}) - \frac{\partial x_t(\hat{\theta})}{\partial \theta} \frac{1}{T} \sum_{t=1}^T \psi_t(\hat{\theta}).$$

Then we regress $\hat{y}_t$ on x_t^* and use the linear regression standard errors.

4.E Other tables and figures

4.E.1 Estimation based on conditional moment restrictions

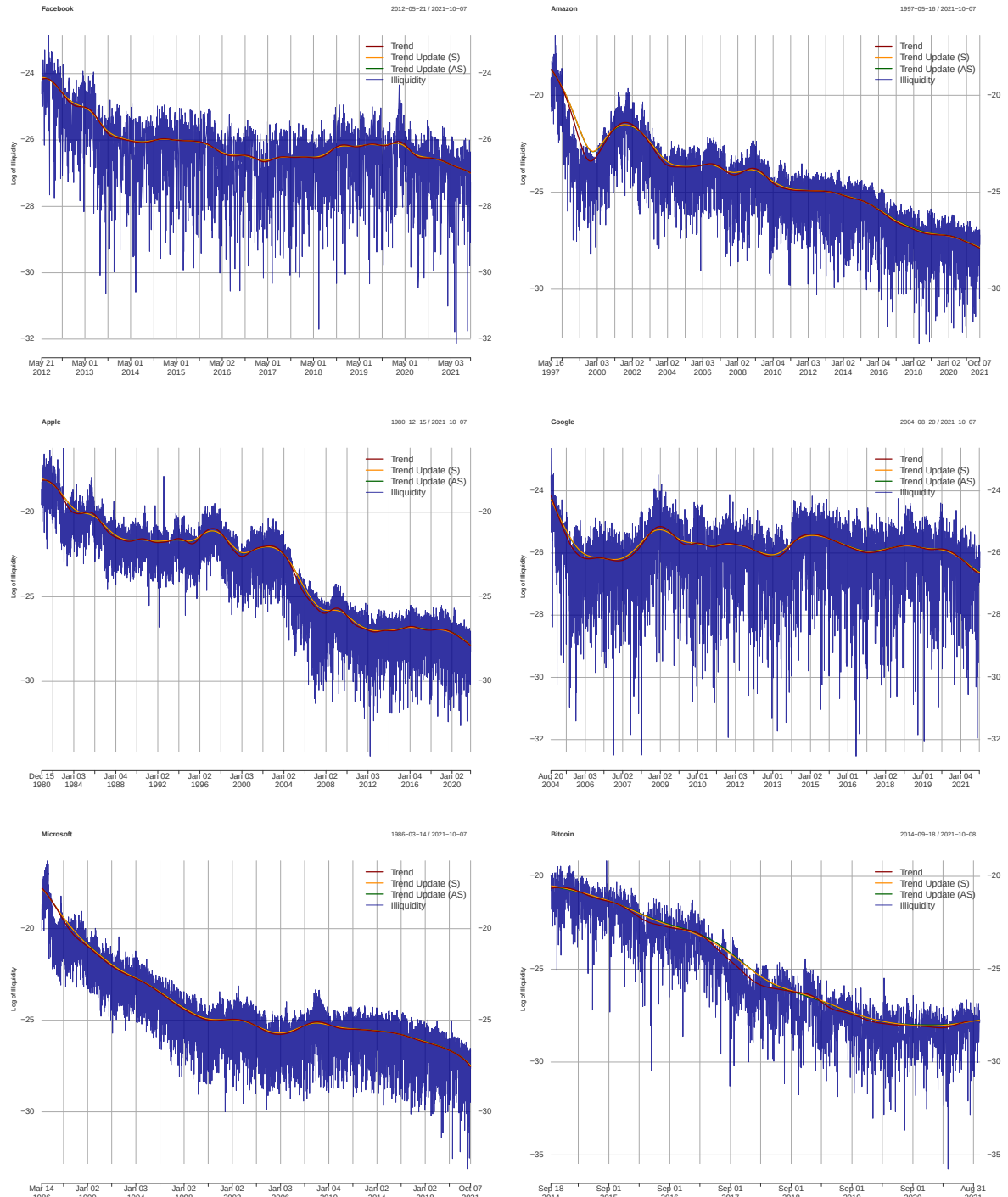
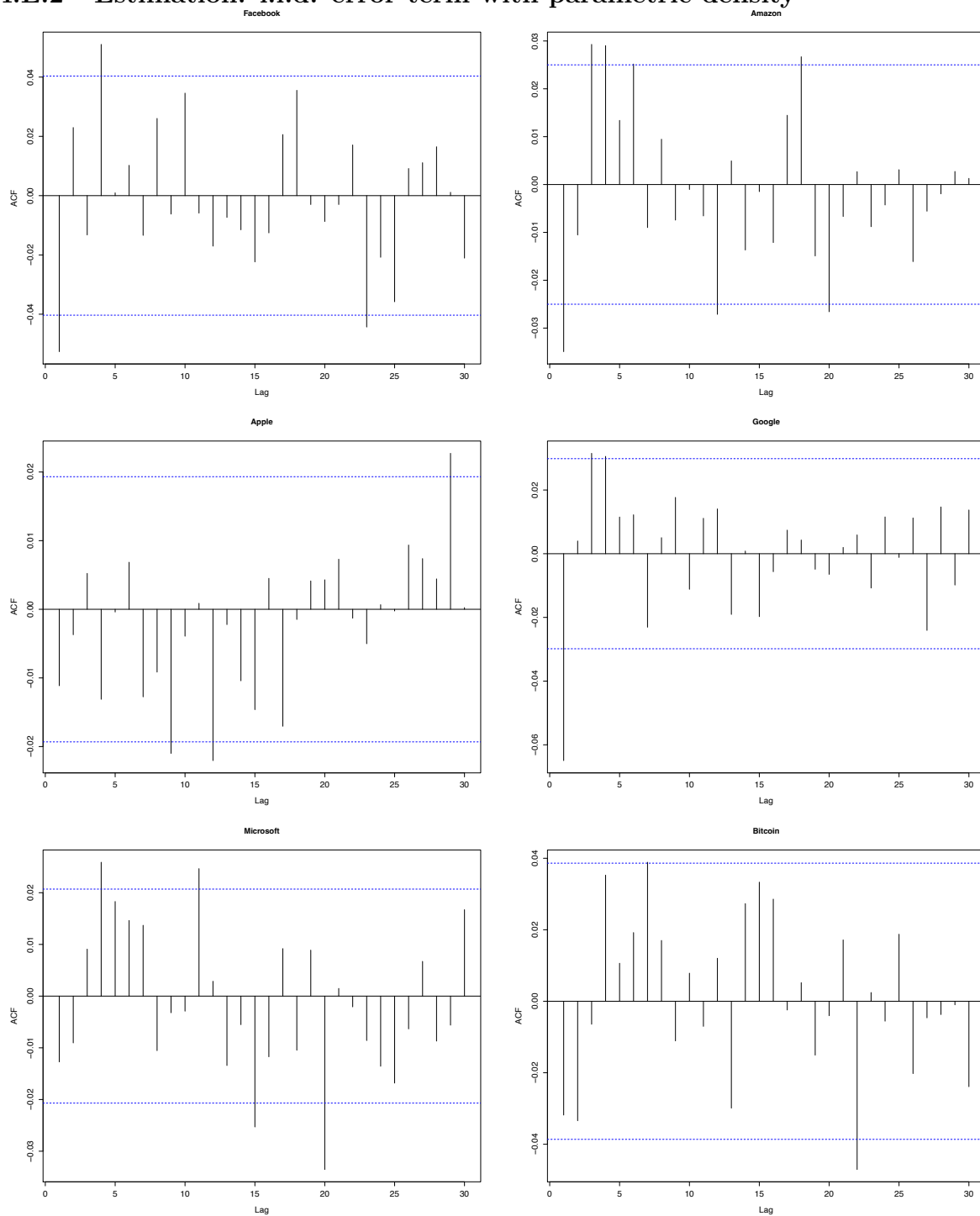


Figure 4.7: Fab 5 and Bitcoin log illiquidity and updated trend function based on the GMM estimator of λ_t parameters. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t .

4.E.2 Estimation: i.i.d. error term with parametric density

Figure 4.8: ACF of ζ_t under the symmetric specification for λ_t .

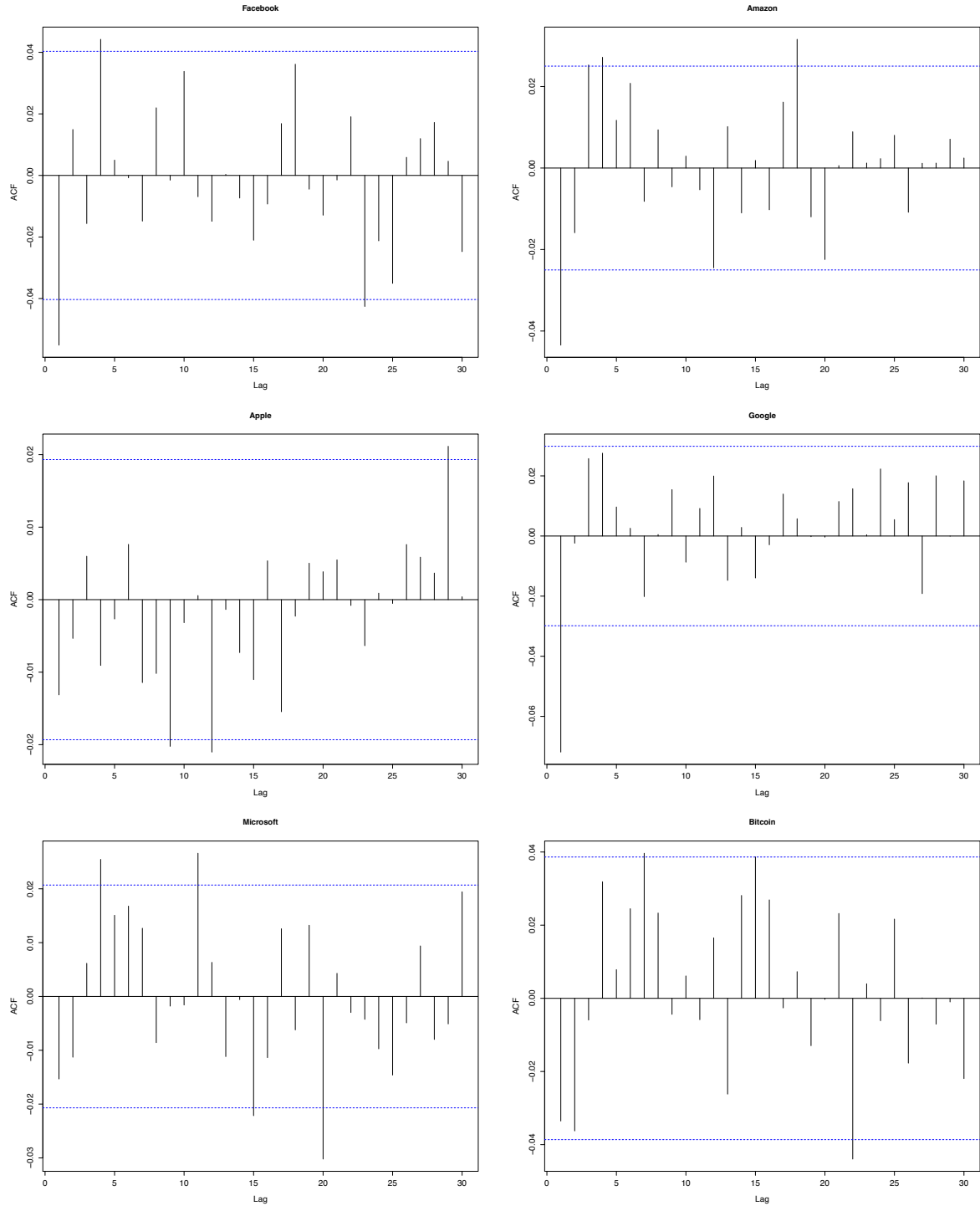
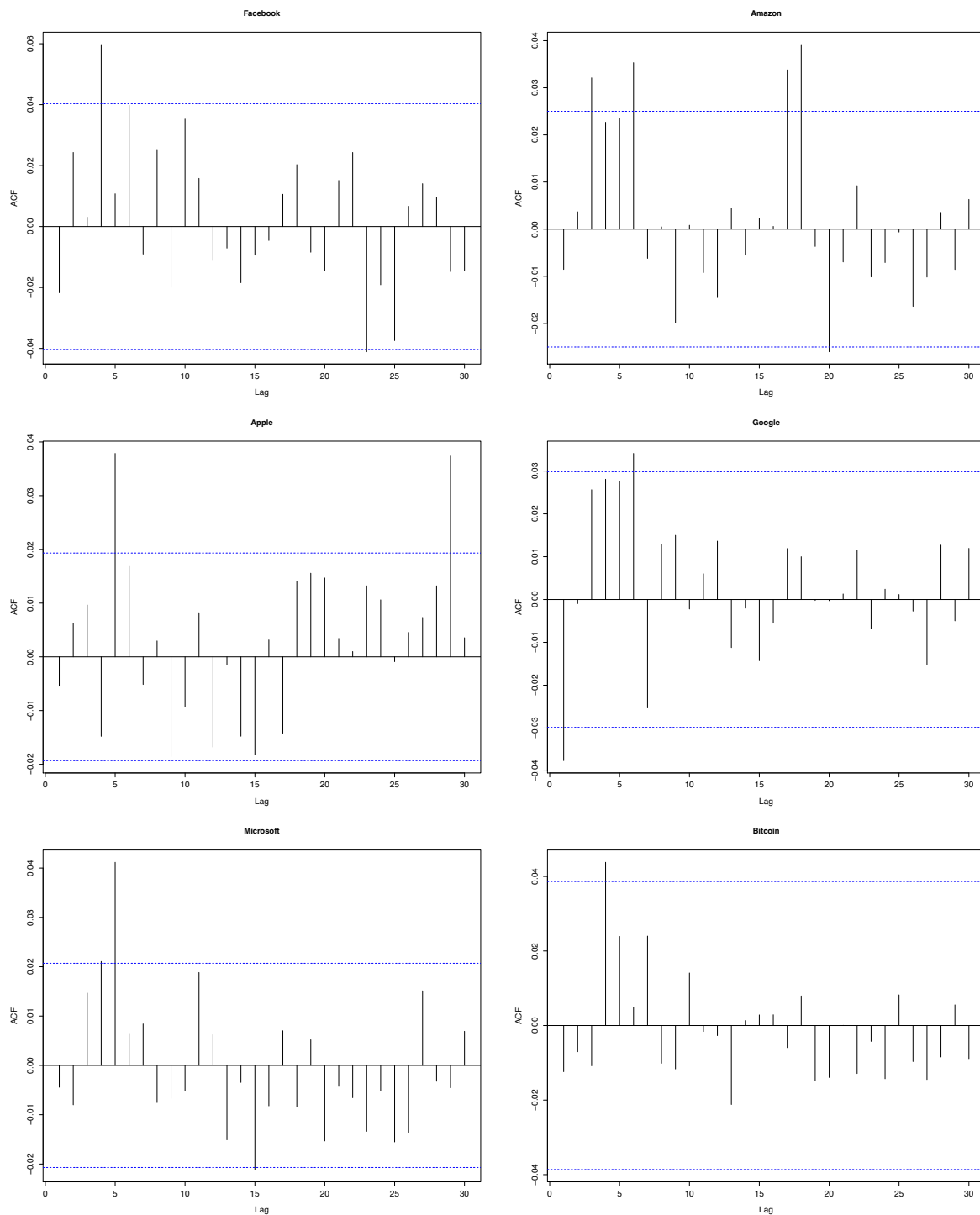


Figure 4.9: ACF of ζ_t under the asymmetric specification for λ_t .

Figure 4.10: ACF of ζ_t^2 under the symmetric specification for λ_t .

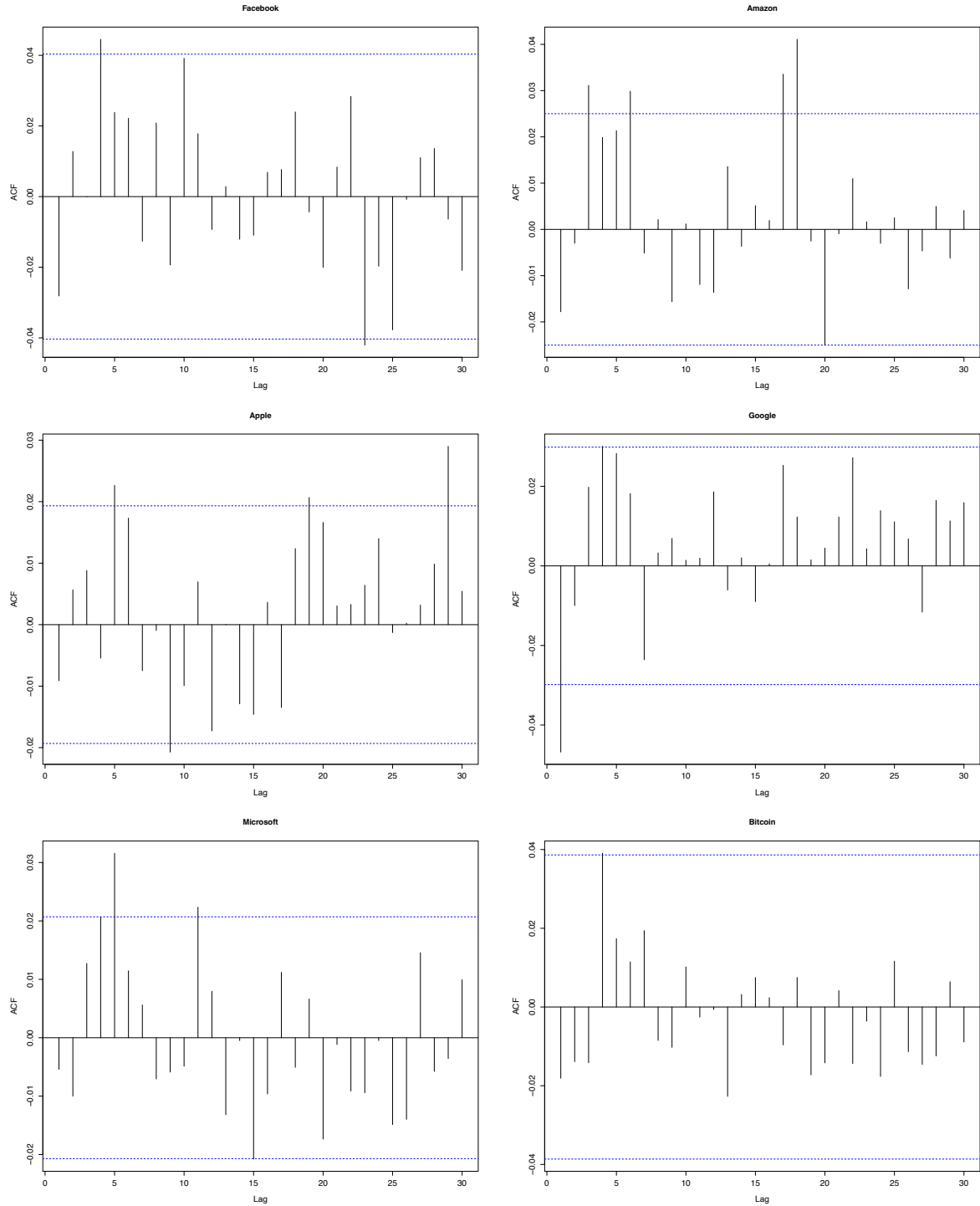
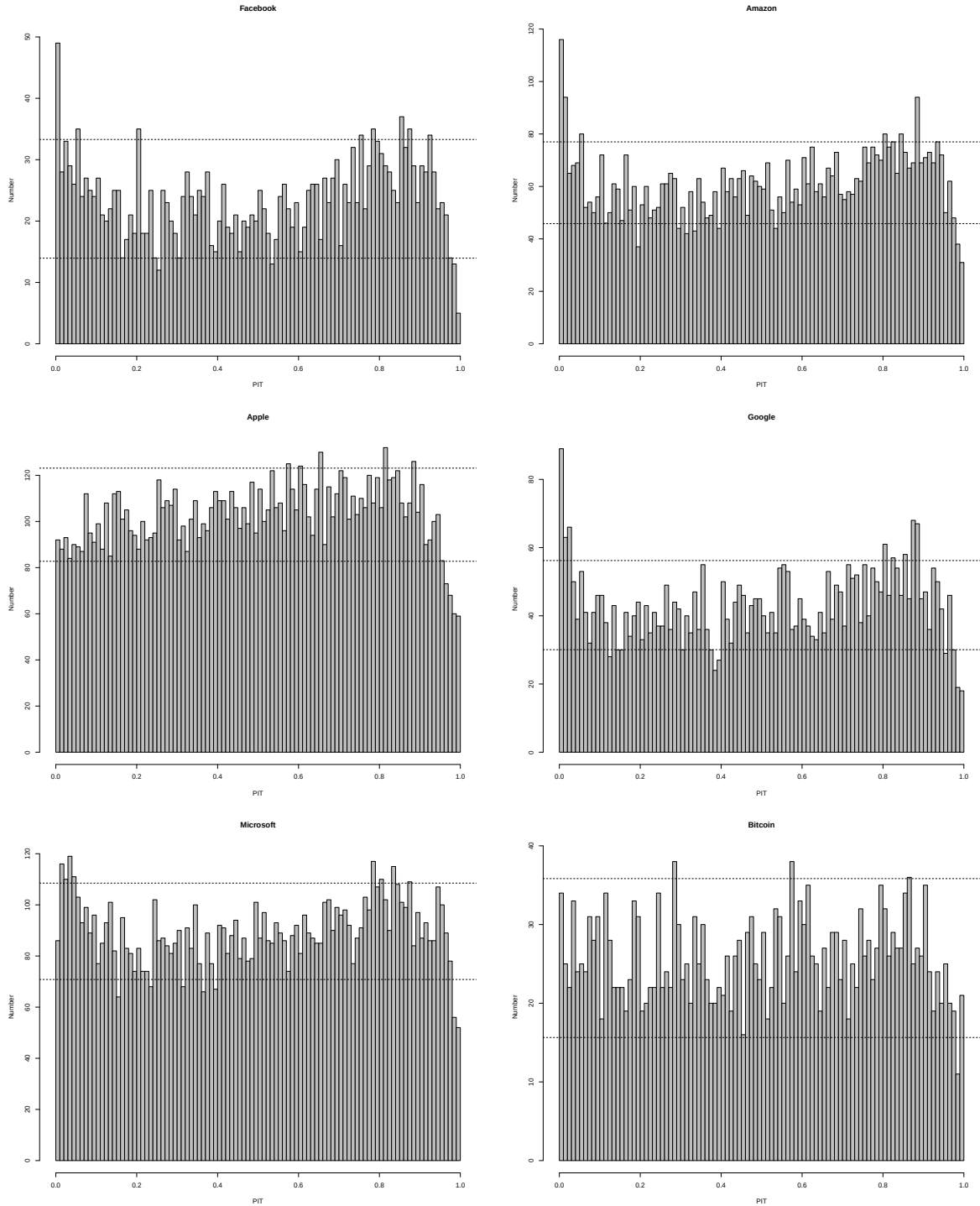


Figure 4.11: ACF of ζ_t^2 under the asymmetric specification for λ_t .

Figure 4.12: Probability integral transform (PIT) of ζ_t under the symmetric specification for λ_t .

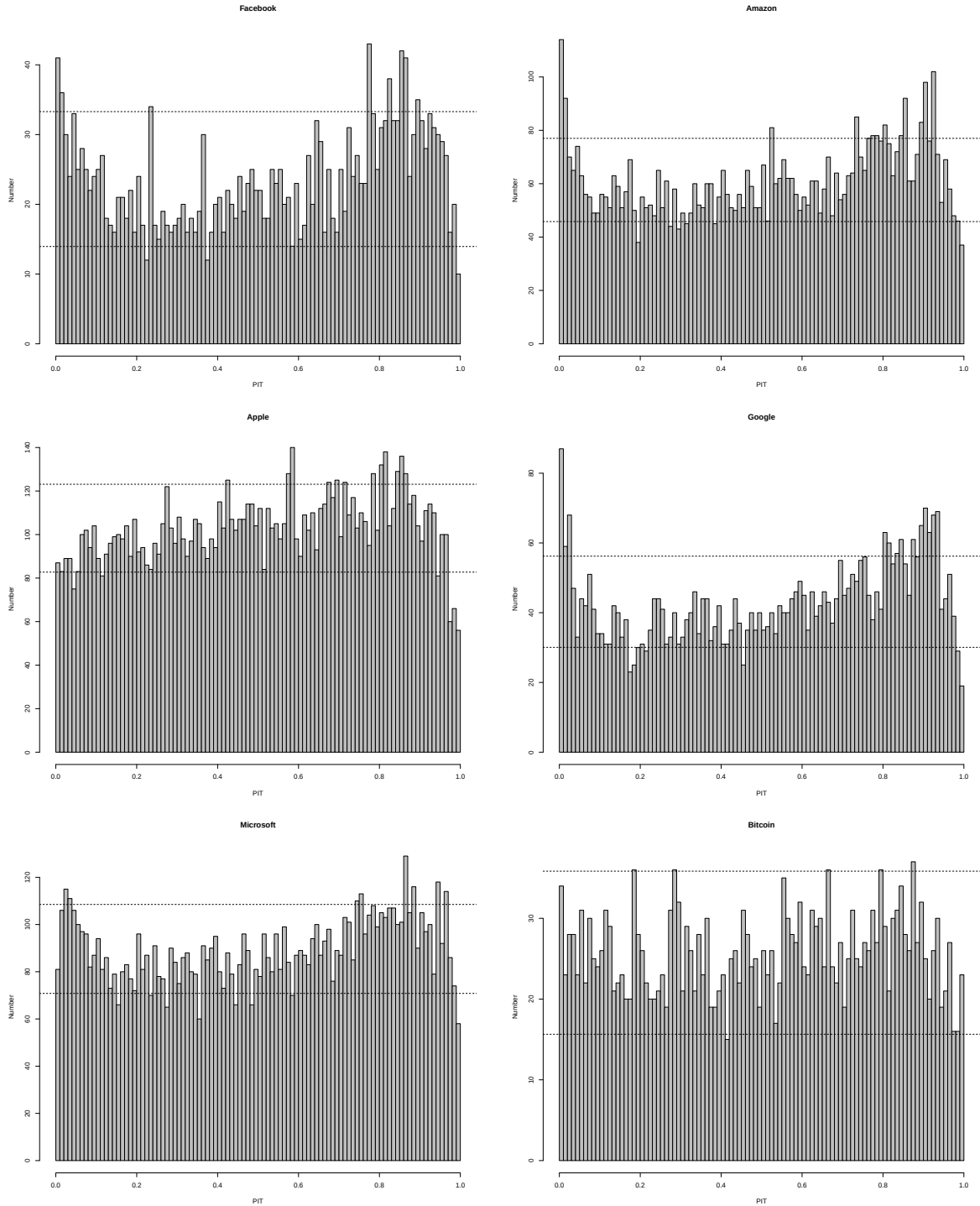


Figure 4.13: Probability integral transform (PIT) of ζ_t under the asymmetric specification for λ_t .

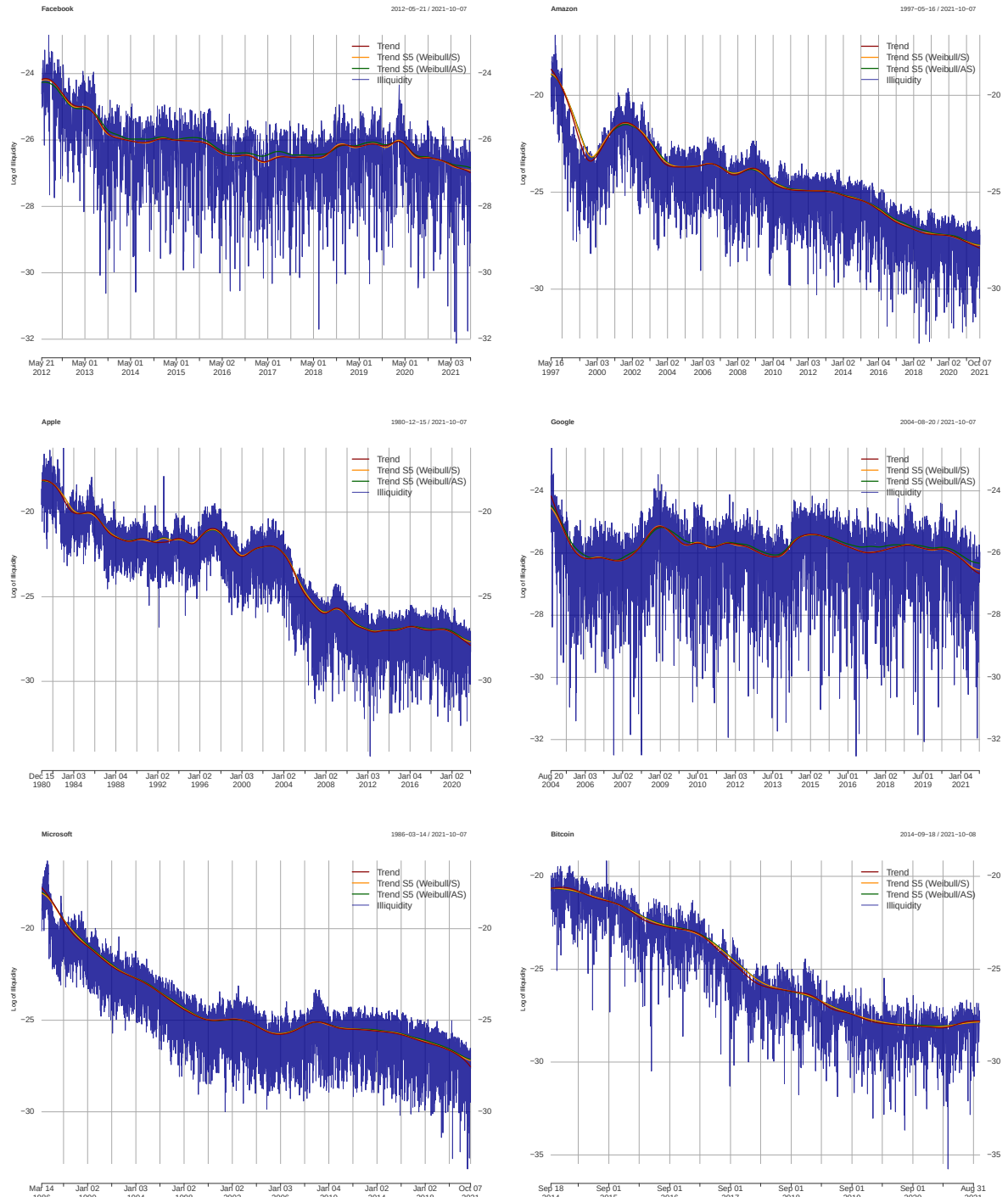


Figure 4.14: Fab 5 and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the error term ζ_t follows a Weibull distribution. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t .

4.E.3 Estimation: i.i.d. error term with nonparametric density

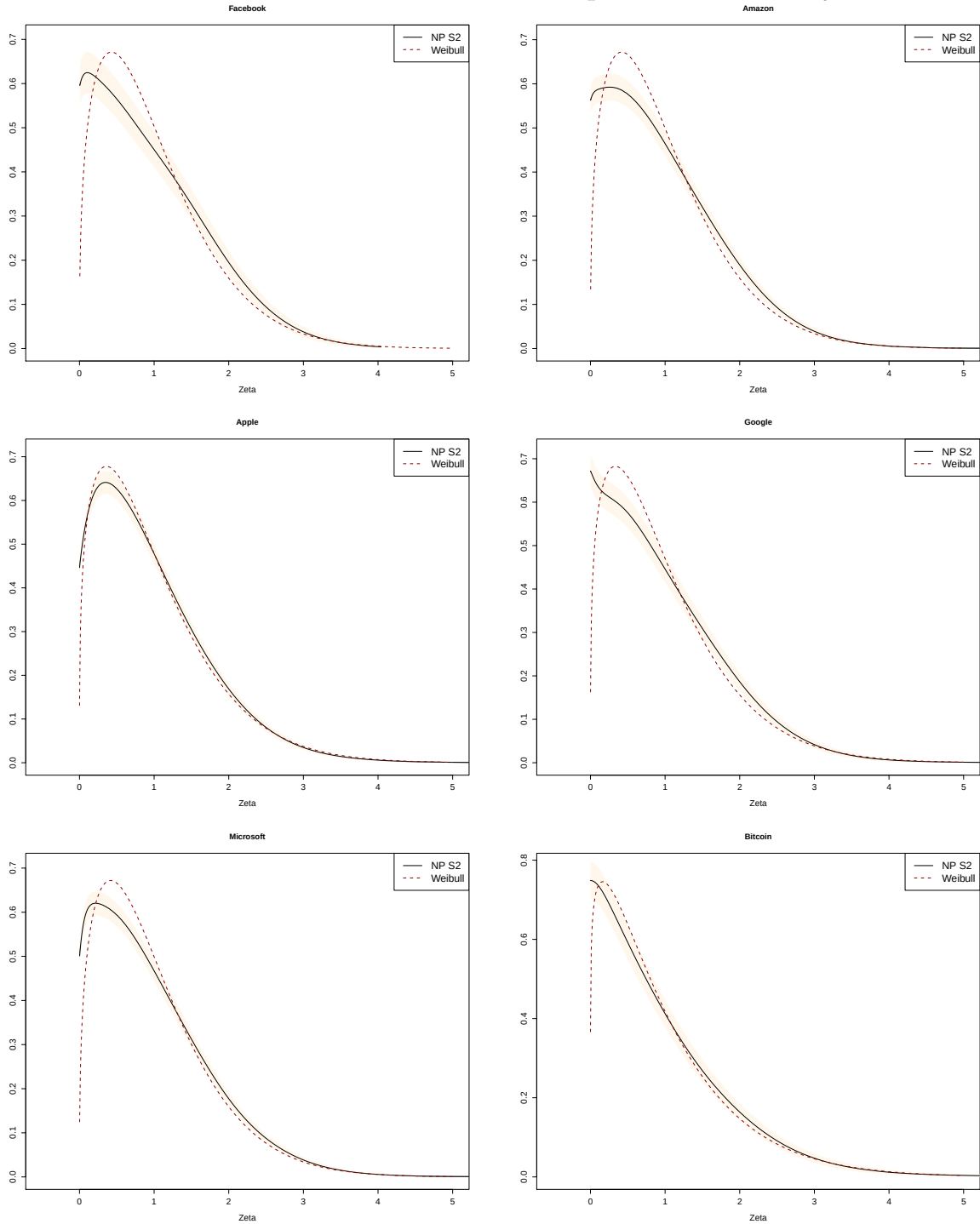


Figure 4.15: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the symmetric specification for λ_t .

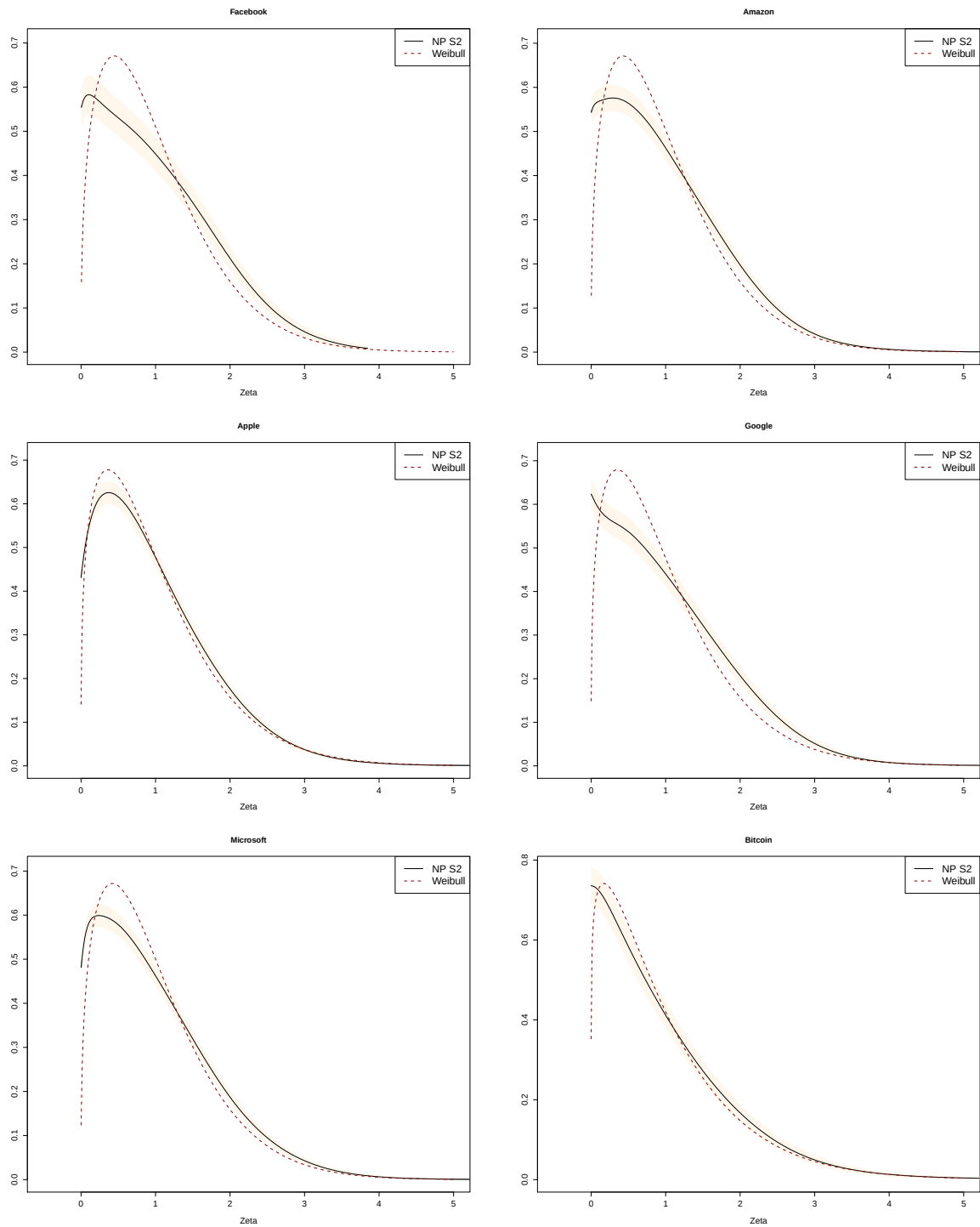


Figure 4.16: Comparison between the kernel density estimate of ζ_t (solid line) and the Weibull density (dashed line) under the asymmetric specification for λ_t .

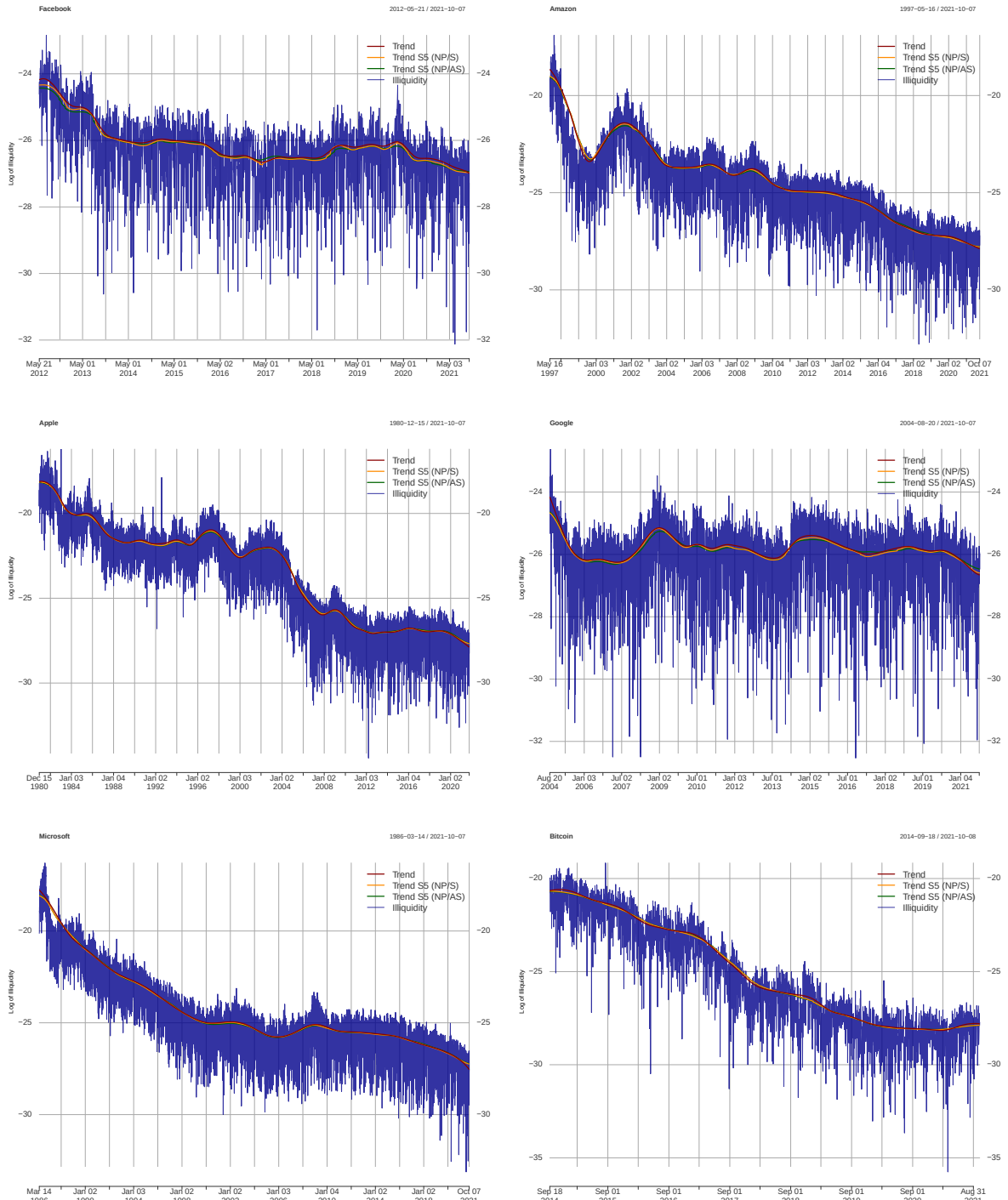


Figure 4.17: Facebook and Bitcoin log illiquidity and updated trend function based on the semiparametric ML estimator of λ_t parameters where the density of the error term ζ_t is estimated nonparametrically. The red curve corresponds to the initial estimate of the trend function and the yellow and green curves correspond to the updated estimates in, respectively, the symmetric and asymmetric specifications of λ_t .

4.E.4 Risk premium

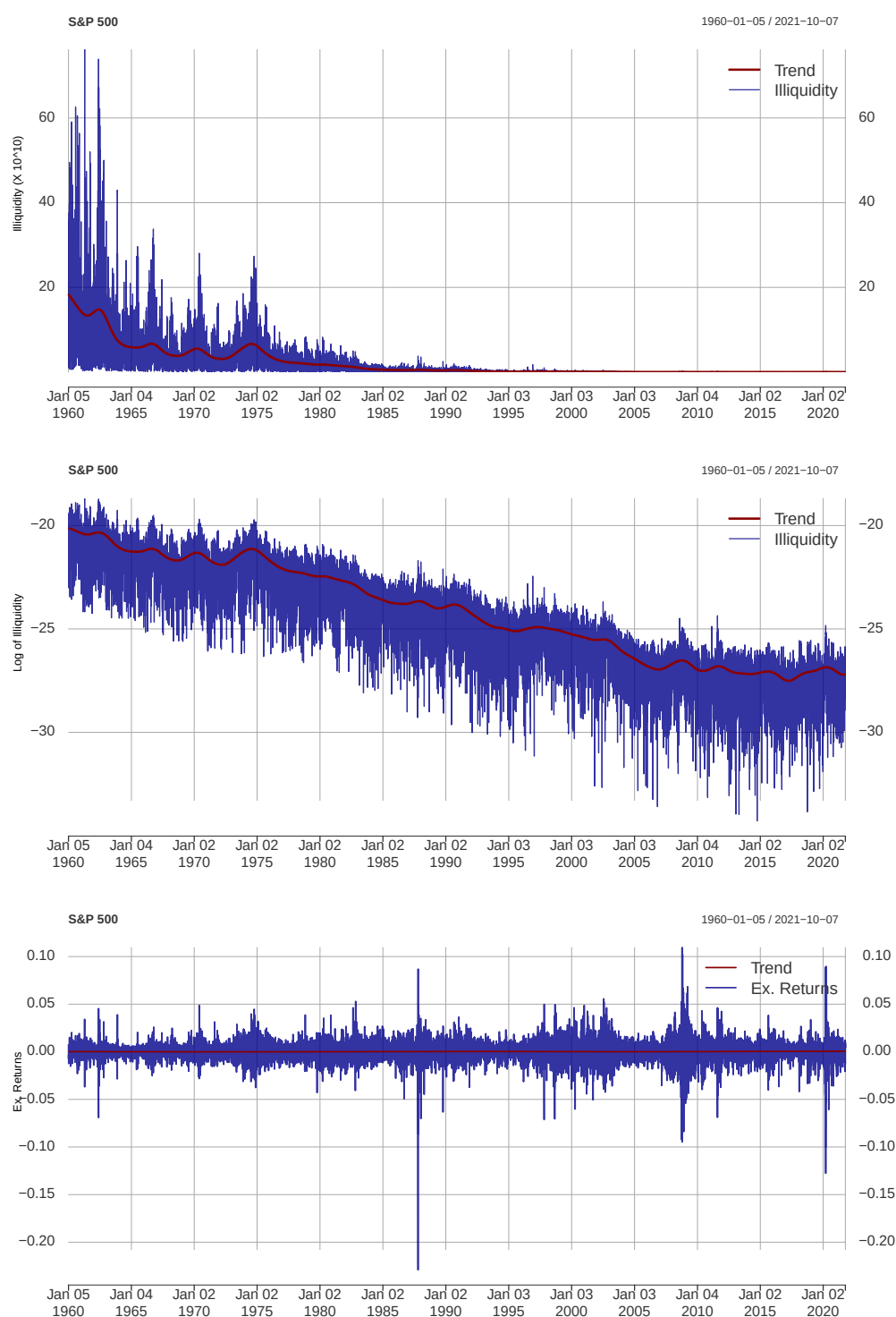


Figure 4.18: S&P 500 index daily (log) illiquidity series and return data.

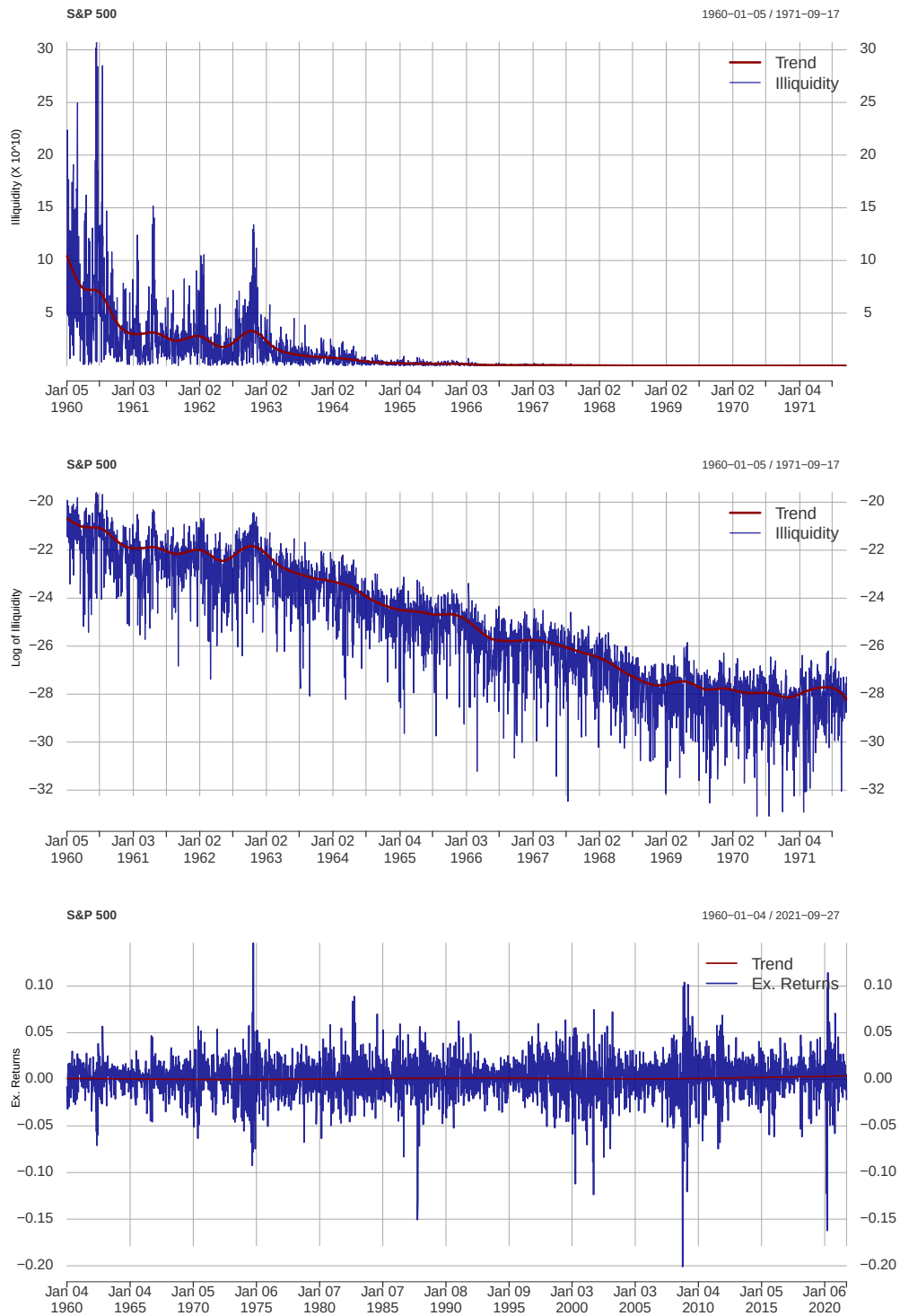


Figure 4.19: S&P 500 index weekly (log) illiquidity series and return data.

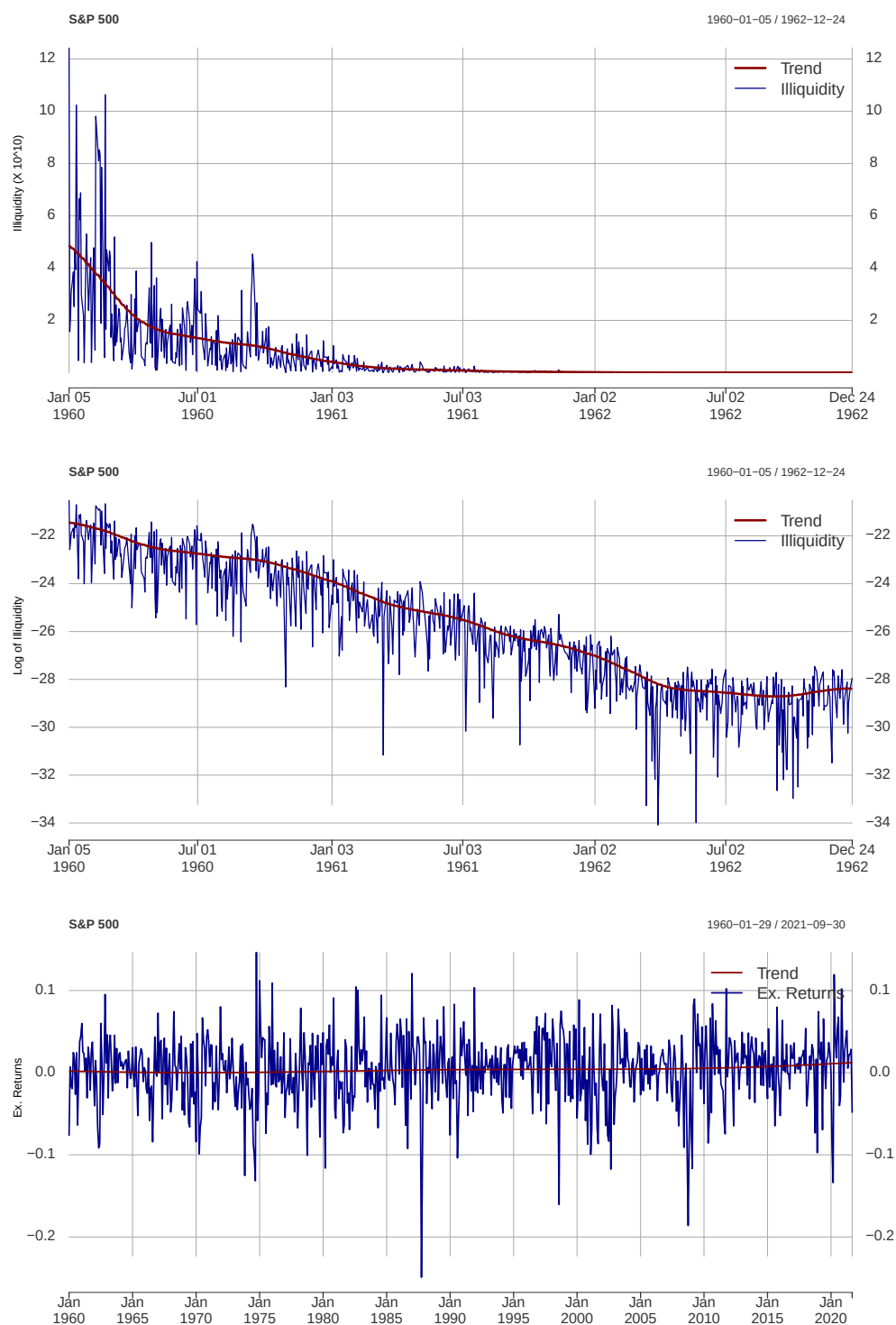


Figure 4.20: S&P 500 index monthly (log) illiquidity series and return data.

Bibliography

- Acharya, V. V., T. Eisert, C. Eufinger, and C. Hirsch (2019): “Whatever it takes: The real effects of unconventional monetary policy”. *The Review of Financial Studies* 32.9, pp. 3366–3411.
- Adrian, T., M. Fleming, O. Shachar, and E. Vogt (2017): “Market liquidity after the financial crisis”. *Annual Review of Financial Economics* 9, pp. 43–83.
- Aielli, G. P. (2013): “Dynamic conditional correlation: on properties and estimation”. *Journal of Business & Economic Statistics* 31.3, pp. 282–299.
- Aït-Sahalia, Y. and J. Duarte (2003): “Nonparametric option pricing under shape restrictions”. *Journal of Econometrics* 116.1-2, pp. 9–47.
- Amihud, Y. (2002): “Illiquidity and stock returns: Cross-section and time-series effects”. *Journal of Financial Markets* 5.1, pp. 31–56.
- Amihud, Y. and H. Mendelson (2015): “The pricing of illiquidity as a characteristic and as risk”. *Multinational Finance Journal* 19.3, pp. 149–168.
- Ao, M., L. Yingying, and X. Zheng (2019): “Approaching mean-variance efficiency for large portfolios”. *The Review of Financial Studies* 32.7, pp. 2890–2919.
- Archakov, I. and P. R. Hansen (2020): “A canonical representation of block matrices with applications to covariance and correlation matrices”. *arXiv preprint arXiv:2012.02698*.
- Archakov, I. and P. R. Hansen (2021): “A new parametrization of correlation matrices”. *Econometrica* 89.4, pp. 1699–1715.
- Archakov, I., P. R. Hansen, and A. Lunde (2020): “A multivariate realized GARCH model”. *arXiv preprint arXiv:2012.02708*.
- Babii, A., E. Ghysels, and J. Striaukas (2020): “High-dimensional Granger causality tests with an application to VIX and news”. *Available at SSRN 3615718*.

- Babii, A., E. Ghysels, and J. Striaukas (2021): “Machine learning time series regressions with an application to nowcasting”. *Journal of Business & Economic Statistics*, pp. 1–23.
- Baillie, R. T. and T. Bollerslev (1992): “Prediction in dynamic models with time-dependent conditional variances”. *Journal of Econometrics* 52.1-2, pp. 91–113.
- Basseville, M. (2013): “Divergence measures for statistical data processing — An annotated bibliography”. *Signal Processing* 93.4, pp. 621–633.
- Bauer, M. D. and M. Chernov (2021): “Interest rate skewness and biased beliefs”. *National Bureau of Economic Research, Working Paper* 28954.
- Bauwens, L., C. M. Hafner, and S. Laurent (2012): *Handbook of volatility models and their applications*. Vol. 3. John Wiley & Sons.
- Bauwens, L. and S. Laurent (2005): “A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models”. *Journal of Business & Economic Statistics* 23.3, pp. 346–354.
- Bauwens, L., S. Laurent, and J. V. Rombouts (2006): “Multivariate GARCH models: A survey”. *Journal of Applied Econometrics* 21.1, pp. 79–109.
- Beddington, J., C. Furse, P. Bond, D. Cliff, C. Goodhart, K. Houstoun, O. Linton, J.-P. Zigrand, et al. (2012): *Foresight: The future of computer trading in financial markets: Final project report*. Tech. rep. London School of Economics and Political Science, LSE Library.
- Best, M. J. and R. R. Grauer (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, pp. 315–342.
- Blasques, F., C. Francq, and S. Laurent (2022): “Quasi score-driven models”. *Journal of Econometrics*.
- Bollerslev, T. (1986): “Generalized autoregressive conditional heteroskedasticity”. *Journal of Econometrics* 31.3, pp. 307–327.
- Breeden, D. T. and R. H. Litzenberger (1978): “Prices of state-contingent claims implicit in option prices”. *The Journal of Business*, pp. 621–651.
- Brigo, D. and F. Mercurio (2001): “A deterministic-shift extension of analytically-tractable and time-homogeneous short-rate models”. *Finance and Stochastics* 5.3, pp. 369–387.
- Brigo, D. and F. Mercurio (2007): *Interest rate models-theory and practice: With smile, inflation and credit*. Springer Science & Business Media.

- Brito, D., M. C. Medeiros, and R. Ribeiro (2018): “Forecasting large realized covariance matrices: The benefits of factor models and shrinkage”. *Available at SSRN 3163668*.
- Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009): “Sparse and stable Markowitz portfolios”. *Proceedings of the National Academy of Sciences* 106.30, pp. 12267–12272.
- Brody, D. C. and L. P. Hughston (2002): “Entropy and information in the interest rate term structure”. *Quantitative Finance* 2.1, p. 70.
- Brogaard, J. (2010): “High frequency trading and its impact on market quality”.
- Brown, B. M. (1971): “Martingale central limit theorems”. *The Annals of Mathematical Statistics* 42.1, pp. 59–66.
- Candelon, B., C. Hurlin, and S. Tokpavi (2012): “Sampling error and double shrinkage estimation of minimum variance portfolios”. *Journal of Empirical Finance* 19.4, pp. 511–527.
- Caporin, M. and M. McAleer (2012): “Do we really need both BEKK and DCC? A tale of two multivariate GARCH models”. *Journal of Economic Surveys* 26.4, pp. 736–751.
- Chan, L. K., J. Karceski, and J. Lakonishok (1999): “On portfolio optimization: Forecasting covariances and choosing the risk model”. *The Review of Financial Studies* 12.5, pp. 937–974.
- Chang, J., Y. Qiu, Q. Yao, and T. Zou (2018): “Confidence regions for entries of a large precision matrix”. *Journal of Econometrics* 206.1, pp. 57–82.
- Chen, J., G. Dai, and N. Zhang (2020): “An application of sparse-group lasso regularization to equity portfolio optimization and sector selection”. *Annals of Operations Research*, pp. 1–20.
- Chen, X. (2007): “Large sample sieve estimation of semi-nonparametric models”. *Handbook of Econometrics* 6, pp. 5549–5632.
- Chen, X., O. Linton, and I. Van Keilegom (2003): “Estimation of semiparametric models when the criterion function is not smooth”. *Econometrica* 71.5, pp. 1591–1608.
- Copeland, T. E. (1979): “Liquidity changes following stock splits”. *The Journal of Finance* 34.1, pp. 115–141.
- Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985): “A theory of the term structure of interest rates”. *Econometrica* 53.2, pp. 385–407.

- Creal, D., S. J. Koopman, and A. Lucas (2011): “A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations”. *Journal of Business & Economic Statistics* 29.4, pp. 552–563.
- Creal, D., S. J. Koopman, and A. Lucas (2013): “Generalized autoregressive score models with applications”. *Journal of Applied Econometrics* 28.5, pp. 777–795.
- De Nard, G., O. Ledoit, and M. Wolf (2021): “Factor models for portfolio selection in large dimensions: The good, the better and the ugly”. *Journal of Financial Econometrics* 19.2, pp. 236–257.
- Delgado, M. A. and J. Hidalgo (2000): “Nonparametric inference on structural breaks”. *Journal of Econometrics* 96.1, pp. 113–144.
- Demarta, S. and A. J. McNeil (2005): “The t copula and related copulas”. *International Statistical Review* 73.1, pp. 111–129.
- DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009a): “A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms”. *Management Science* 55.5, pp. 798–812.
- DeMiguel, V., L. Garlappi, and R. Uppal (2009b): “Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?” *The Review of Financial Studies* 22.5, pp. 1915–1953.
- Drost, F. C. and C. A. Klaassen (1997): “Efficient estimation in semiparametric GARCH models”. *Journal of Econometrics* 81.1, pp. 193–221.
- Engle, R. (2002): “Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models”. *Journal of Business & Economic Statistics* 20.3, pp. 339–350.
- Engle, R. F. (1982): “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”. *Econometrica: Journal of the Econometric Society*, pp. 987–1007.
- Engle, R. F. and G. M. Gallo (2006): “A multiple indicators model for volatility using intra-daily data”. *Journal of Econometrics* 131.1-2, pp. 3–27.
- Engle, R. F., G. M. Gallo, and M. Velucchi (2012): “Volatility spillovers in East Asian financial markets: a MEM-based approach”. *Review of Economics and Statistics* 94.1, pp. 222–223.

- Engle, R. F., O. Ledoit, and M. Wolf (2019): “Large dynamic covariance matrices”. *Journal of Business & Economic Statistics* 37.2, pp. 363–375.
- Engle, R. F. and J. R. Russell (1998): “Autoregressive conditional duration: A new model for irregularly spaced transaction data”. *Econometrica*, pp. 1127–1162.
- Engle, R. F. and K. Sheppard (2001): “Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH”. *National Bureau of Economic Research, Working Paper* 8554.
- Fama, E. F. and K. R. French (1993): “Common risk factors in the returns on stocks and bonds”. *Journal of Financial Economics* 33, pp. 3–56.
- Fan, J. and J. Chen (1999): “One-step local quasi-likelihood estimation”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.4, pp. 927–943.
- Fan, J., A. Furger, and D. Xiu (2016): “Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data”. *Journal of Business & Economic Statistics* 34.4, pp. 489–503.
- Fan, J., Y. Liao, and M. Mincheva (2011): “High dimensional covariance matrix estimation in approximate factor models”. *The Annals of Statistics* 39.6, 3320–3356.
- Fan, J., J. Zhang, and K. Yu (2012): “Vast portfolio selection with gross-exposure constraints”. *Journal of the American Statistical Association* 107.498, pp. 592–606.
- Fastrich, B., S. Paterlini, and P. Winker (2015): “Constructing optimal sparse portfolios using regularization methods”. *Computational Management Science* 12.3, pp. 417–434.
- Fengler, M. R. (2009): “Arbitrage-free smoothing of the implied volatility surface”. *Quantitative Finance* 9.4, pp. 417–428.
- Filipović, D., M. Larsson, and A. B. Trolle (2017): “Linear-rational term structure models”. *The Journal of Finance* 72.2, pp. 655–704.
- Filipović, D. and A. B. Trolle (2013): “The term structure of interbank risk”. *Journal of Financial Economics* 109.3, pp. 707–733.
- Filipović, D. and Y. Kitapbayev (2018): “On the American swaption in the linear-rational framework”. *Quantitative Finance* 18.11, pp. 1865–1876.
- Fiorentini, G., E. Sentana, and G. Calzolari (2003): “Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations”. *Journal of Business & Economic Statistics* 21.4, pp. 532–546.

- Fisher, J. D., D. W. Puelz, and C. M. Carvalho (2020): “Monotonic effects of characteristics on returns”. *Annals of Applied Statistics* 14.4, pp. 1622–1650.
- Francq, C. and J.-M. Zakoian (2019): *GARCH models: Structure, statistical inference and financial applications*. John Wiley & Sons.
- Friedman, J., T. Hastie, and R. Tibshirani (2010): “A note on the group lasso and a sparse group lasso”. *arXiv preprint arXiv:1001.0736*.
- Fu, A., B. Narasimhan, and S. Boyd (2020): “CVXR: An R Package for Disciplined Convex Optimization”. *Journal of Statistical Software* 94.1, pp. 1–34.
- Gorgi, P., P. Hansen, P. Janus, and S. Koopman (2018): “Realized Wishart-GARCH: A score-driven multi-asset volatility model”. *Journal of Financial Econometrics* 17.1, pp. 1–32.
- Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009): “Do liquidity measures measure liquidity?” *Journal of Financial Economics* 92.2, pp. 153–181.
- Hafner, C. M. and O. Linton (2010): “Efficient estimation of a multivariate multiplicative volatility model”. *Journal of Econometrics* 159.1, pp. 55–73.
- Hafner, C. M., O. B. Linton, and H. Tang (2020): “Estimation of a multiplicative correlation structure in the large dimensional case”. *Journal of Econometrics* 217.2, pp. 431–470.
- Hafner, C. M., O. B. Linton, and L. Wang (2022): “Dynamic Autoregressive Liquidity (DARLiQ)”. *Cambridge Working Papers in Economics* CWPE2214.
- Hafner, C. M. and O. Reznikova (2012): “On the estimation of dynamic conditional correlation models”. *Computational Statistics & Data Analysis* 56.11, pp. 3533–3545.
- Hafner, C. M. and L. Wang (2021): “A dynamic conditional score model for the log correlation matrix”. *Journal of Econometrics*. Forthcoming.
- Hafner, C. M. and L. Wang (2022): “Dynamic portfolio selection with sector-specific regularization”. *Econometrics and Statistics*. Forthcoming.
- Hagan, P. and M. Konikov (2004): “Interest rate volatility cube: Construction and use”. *Bloomberg Technical Report* 62.
- Hansen, P. R., A. Lunde, and V. Voev (2014): “Realized beta GARCH: A multivariate GARCH model with realized measures of volatility”. *Journal of Applied Econometrics* 29.5, pp. 774–799.
- Harvey, A. and R.-J. Lange (2017): “Volatility modeling with a generalized t distribution”. *Journal of Time Series Analysis* 38.2, pp. 175–190.

- Harvey, A. and G. Sucarrat (2014): “EGARCH models with fat tails, skewness and leverage”. *Computational Statistics & Data Analysis* 76, pp. 320–338.
- Harvey, A. C. (2013): *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Cambridge University Press.
- Hasbrouck, J. (2009): “Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data”. *The Journal of Finance* 64.3, pp. 1445–1477.
- Hautsch, N. and S. Voigt (2019): “Large-scale portfolio allocation under transaction costs and model uncertainty”. *Journal of Econometrics* 212.1, pp. 221–240.
- Heider, F., F. Saidi, and G. Schepens (2019): “Life below zero: Bank lending under negative policy rates”. *The Review of Financial Studies* 32.10, pp. 3728–3761.
- Hendershott, T., C. M. Jones, and A. J. Menkveld (2011): “Does algorithmic trading improve liquidity?” *The Journal of Finance* 66.1, pp. 1–33.
- Higham, N. J. (2008): *Functions of matrices: Theory and computation*. Society for Industrial and Applied Mathematics.
- Hong, H., W. Torous, and R. Valkanov (2007): “Do industries lead stock markets?” *Journal of Financial Economics* 83.2, pp. 367–396.
- Hou, K. (2007): “Industry information diffusion and the lead-lag effect in stock returns”. *The Review of Financial Studies* 20.4, pp. 1113–1138.
- Hull, J. and A. White (1990): “Pricing interest-rate-derivative securities”. *The Review of Financial Studies* 3.4, pp. 392–573.
- Jagannathan, R. and T. Ma (2003): “Risk reduction in large portfolios: Why imposing the wrong constraints helps”. *The Journal of Finance* 58.4, pp. 1651–1683.
- Jensen, S. T. and A. Rahbek (2004): “Asymptotic inference for nonstationary GARCH”. *Econometric Theory* 20.6, pp. 1203–1226.
- Joe, H. (1997): *Multivariate models and multivariate dependence concepts*. Chapman & Hall.
- Joe, H. (2005): “Asymptotic efficiency of the two-stage estimation method for copula-based models”. *Journal of Multivariate Analysis* 94, pp. 401–419.
- Kwakatsu, H. (2006): “Matrix exponential GARCH”. *Journal of Econometrics* 134.1, pp. 95–128.
- Kurose, Y. and Y. Omori (2020): “Multiple-block dynamic equicorrelations with realized measures, leverage and endogeneity”. *Econometrics and Statistics* 13, pp. 46–68.

- Lassance, N., V. DeMiguel, and F. Vrms (2022): “Optimal portfolio diversification via independent component analysis”. *Operations Research* 70.1, pp. 55–72.
- Ledoit, O. and M. Wolf (2004): “A well-conditioned estimator for large-dimensional covariance matrices”. *Journal of Multivariate Analysis* 88.2, pp. 365–411.
- Ledoit, O. and M. Wolf (2011): “Robust performances hypothesis testing with the variance”. *Wilmott* 2011.55, pp. 86–89.
- Ledoit, O. and M. Wolf (2017): “Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks”. *The Review of Financial Studies* 30.12, pp. 4349–4388.
- Ledoit, O. and M. Wolf (2020): “Analytical nonlinear shrinkage of large-dimensional covariance matrices”. *The Annals of Statistics* 48.5, pp. 3043–3065.
- Li, H. and F. Zhao (2009): “Nonparametric estimation of state-price densities implicit in interest rate cap prices”. *The Review of Financial Studies* 22.11, pp. 4335–4376.
- Linton, O. and J. R. McCrorie (1995): “Differentiation of an exponential matrix function”. *Econometric Theory* 11.5, pp. 1182–1185.
- Linton, O. and Z. Xiao (2001): “Second-order approximation for adaptive regression estimators”. *Econometric Theory* 17.5, pp. 984–1024.
- Longstaff, F. A., S. Mithal, and E. Neis (2005): “Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market”. *The Journal of Finance* 60.5, pp. 2213–2253.
- Lu, Z. and O. Linton (2007): “Local linear fitting under near epoch dependence”. *Econometric Theory* 23.1, pp. 37–70.
- Magnus, J. R. and H. Neudecker (2019): *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons.
- Manganelli, S. (2005): “Duration, volume and volatility impact of trades”. *Journal of Financial Markets* 8.4, pp. 377–399.
- Markowitz, H. (1952): “Portfolio selection”. *The Journal of Finance* 7.1, pp. 77–91.
- Morana, C. (2019): “Regularized semiparametric estimation of high dimensional dynamic conditional covariance matrices”. *Econometrics and Statistics* 12, pp. 42–65.
- Moskowitz, T. J. and M. Grinblatt (1999): “Do industries explain momentum?” *The Journal of Finance* 54.4, pp. 1249–1290.

- Nelson, D. B. (1991): “Conditional heteroskedasticity in asset returns: A new approach”. *Econometrica* 59.2, pp. 347–370.
- O’Hara, M. and M. Ye (2011): “Is market fragmentation harming market quality?” *Journal of Financial Economics* 100.3, pp. 459–474.
- Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2018): “New HEAVY models for fat-tailed realized covariances and returns”. *Journal of Business & Economic Statistics* 36.4, pp. 643–657.
- Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle (2020): “Fitting vast dimensional time-varying covariance models”. *Journal of Business & Economic Statistics*, pp. 1–17.
- Puelz, D., C. M. Carvalho, and P. R. Hahn (2015): “Optimal ETF selection for passive investing”. *arXiv preprint arXiv:1510.03385*.
- Puelz, D., P. R. Hahn, and C. M. Carvalho (2020): “Portfolio selection for individual passive investing”. *Applied Stochastic Models in Business and Industry* 36.1, pp. 124–142.
- Recchioni, M. C. and Y. Sun (2016): “An explicitly solvable Heston model with stochastic interest rate”. *European Journal of Operational Research* 249.1, pp. 359–377.
- Recchioni, M. C., Y. Sun, and G. Tedeschi (2017): “Can negative interest rates really affect option pricing? Empirical evidence from an explicitly solvable stochastic volatility model”. *Quantitative Finance* 17.8, pp. 1257–1275.
- Simon, N., J. Friedman, T. Hastie, and R. Tibshirani (2013): “A sparse-group lasso”. *Journal of Computational and Graphical Statistics* 22.2, pp. 231–245.
- Straumann, D. and T. Mikosch (2006): “Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach”. *The Annals of Statistics* 34.5, pp. 2449–2495.
- Tibshirani, R. (1996): “Regression shrinkage and selection via the lasso”. *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.
- Trimborn, S. and W. K. Härdle (2018): “CRIX an Index for cryptocurrencies”. *Journal of Empirical Finance* 49, pp. 107–122.
- Trolle, A. B. and E. S. Schwartz (2014): “The swaption cube”. *The Review of Financial Studies* 27.8, pp. 2307–2353.
- Vasicek, O. (1977): “An equilibrium characterization of the term structure”. *Journal of Financial Economics* 5.2, pp. 177–188.

- Vassallo, D., G. Bucchieri, and F. Corsi (2021): “A DCC-type approach for realized covariance modeling with score-driven dynamics”. *International Journal of Forecasting* 37.2, pp. 569–586.
- Vogt, M. (2012): “Nonparametric regression for locally stationary time series”. *The Annals of Statistics* 40.5, pp. 2601–2633.
- Vrins, F. and L. Wang (2021): “Asymmetric short-rate model without lower bound”. *LIDAM Discussion Paper LFIN* 2021006.
- Yatchew, A. (2000): “Scale economies in electricity distribution: A semiparametric analysis”. *Journal of Applied Econometrics* 15.2, pp. 187–210.
- Yatchew, A. and W. Härdle (2006): “Nonparametric state price density estimation using constrained least squares and the bootstrap”. *Journal of Econometrics* 133.2, pp. 579–599.
- Yuan, M. and Y. Lin (2006): “Model selection and estimation in regression with grouped variables”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.1, pp. 49–67.
- Zhao, Z., O. Ledoit, and H. Jiang (2020): “Risk reduction and efficiency increase in large portfolios: leverage and shrinkage”. *University of Zurich, Department of Economics, Working Paper* 328.
- Zou, H. and T. Hastie (2005): “Regularization and variable selection via the elastic net”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.