Alma Mater Studiorum
Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE STATISTICHE

CICLO XXXII

Settore Concorsuale: 13/D2
Settore Scientifico Disciplinare: SECS-S/03

Essays in Robust and Nonlinear Time Series Models

Presentata da: Enzo D'Innocenzo

Coordinatrice Dottorato:
Prof.ssa Alessandra Luati

Supervisor:
Prof. Mario Mazzocchi
Co-Supervisor:
Prof.ssa Alessandra Luati

Esame Finale Anno 2020

Declaration of Authorship

I, Enzo D'INNOCENZO, declare that this thesis titled, "Essays in Robust and Nonlinear Multivariate Time Series Models" and the work presented in it are my own. I confirm that:

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

“ **Παθήματα** - *μαθήματα* ”

“ *Pathēmata - mathēmata* ”

The Dog and The Cook - Aesop’s Fables

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name
Department of Statistical Sciences “Paolo Fortunati”

Doctor of Philosophy in Statistical Sciences

Essays in Robust and Nonlinear Multivariate Time Series Models

by Enzo D’INNOCENZO

This PhD dissertation deals with the world of multivariate time series models where the behaviour of the observed process is described by using a time-varying parameter. In particular, this thesis explore three different dynamic multivariate nonlinear models which are able to deal with multivariate time series gathered from heavy-tailed phenomena.

Although the popularity of linear and univariate time series models, empirical evidences have shown that variables generated from complex phenomena are typically inter-related both contemporaneously and across time. This is the case for several fields of science such as economics, finance, biology or physics, where it is widely accepted that with a univariate approach it is difficult to obtain a satisfactory representation of the reality or to make good predictions about the future. For these reasons, the literature of linear multivariate Gaussian time series models has received increasing attention. However, these models are known for their unsatisfactory performances when the collected data are contaminated by outliers, yielding biased estimates and unreliable forecasts. In fact, when departure from the hypothesis of normality is confirmed by the observed data, it is reasonable to switch into the realm of nonlinear or non-Gaussian time series models.

Unfortunately, despite the development of recent technologies, the estimation of nonlinear time series models might be really challenging, since they require simulation-based and computer-intensive methods. In addition, statistical properties of such estimators are not always easy to be derived. This thesis contributes to the literature by defining dynamic multivariate and heavy-tailed models that are relatively simple. The emphasis is models which are analytically tractable and can be easily estimated by means of maximum likelihood. For each of the models, a very detailed statistical and asymptotic analysis it is provided. Their practical usefulness is highlighted with several simulation studies and empirical applications.

Acknowledgements

The fascination of a PhD Thesis is that it reflects the way in which the author experienced the world during his/her PhD period. For this reason, I would like to dedicate this acknowledgments to the people who have heavily influenced the birth and the development of this work.

I wish to express my gratitude to my supervisors. To Professor Mario Mazzocchi, for his combination of professional skills and human warmth, which surely helped me overcome several challenges I was confronted with when starting scientific research. To professor Alessandra Luati, who since my bachelor degree program at the Department of Statistical Sciences of the University of Bologna, taught me time series. Her invaluable and stimulating lectures provided me the means to do research, and this dissertation work heavily leans on her given priceless background work. Moreover, I also owe to her the most significant periods of my PhD duration. She made possible my experience as a visiting PhD student at the excellent Department of Econometrics and Operations Research of the Vrije Universiteit Amsterdam.

The opportunity to work and discuss with Professor Siem Jan Koopman, Professor Francisco Blasques and Professor André Lucas was a true inspiration that I will never forget. I would give a special mention to Professor Jan Magnus for having preciously dedicated an amazing amount of time to help me solve algebra problems. All the discussions I had in that department tremendously improved my personal skills as a researcher and thus, I would like to express a special thanks to all these professors for their extraordinary hospitality.

There is another Professor to which I would explicitely express my gratitude, that is, Professor Andrew C. Harvey from the University of Cambridge. I could never forget our meeting of may 2017, where the first chapter of this thesis was born.

Of course, I would certainly be remiss to not mention and thank my whole family for everything they gave me, from love and patient, to all the means necessary to live. They always have listened to me stress and complain but also shared in the joy of my achievements. In particular, my sister Roberta, brother Dante Luca and my parents, Giovanna and Marco. It is largely due to their efforts that I am as I am today. For this reason I dedicate my PhD thesis to them.

Contents

List of Figures

List of Tables

To my parents

Chapter 1

Introduction

1.1 A Tale in Robustness, Nonlinearity and Multivariate Time Series Models

In the recent decades, the interest in multivariate time series modelling has grown exponentially. The time series data gathering is becoming easier and faster on a everyday basis, motivating researchers and practitioners to build complex models that are able to analyze these series in univariate and multivariate environments. It is generally agreed that while the major concern in the univariate case is the study, forecast and modelling the dependence structure in the series via time domain or frequency domain approaches, in the multivariate case one is mostly involved in the studies of the relationships between variables. However, this is also the major task since the specification of a multivariate time series model can be really challenging to be established. In particular, from a theoretical point of view, since the mathematical tractability of a multivariate model can be easily lost. This was the objective of Hannan, 1970, who provides a comprehensive theoretical treatment of that subject. At the time, he already noted the following fact.

> –"There is one other characteristic that a modern book on time series must have and that is the development of the theory and methods for the case in which multiple measurements are made at each point, for this is usually the case."

There exist plenty of well understood linear and nonlinear univariate models which can be easily implemented in order to study and forecast time series that arise from several phenomena. When a multivariate system is our concern, the class of vector autoregressive moving average (VARMA) models has gained lot of experience and even today they are still at the first level of development. A detailed reference for that subject is the seminal book of Lütkepohl, 2007. Notably, these models have been succesfully employed in economics and finance and are nowadays considered as the standard approach by applied econometricians to address policy relevant economic problems.

In contrast, because of the inherent flexibility of nonlinear multivariate models, little learning has occurred, forcing time series analysts to be still stacked in linear frameworks, where stochastic properties of models and asymptotic theory of the estimators are well understood. However, despite their mathematical tractability, linear models are known to be quite restrictive and sometimes of little practical usage, especially for time series which derive heavy-tailed phenomena. For example, empirical studies in several scientific fields, such as finance, economics, physics or engineering, have strongly proved the fact that they are not suited for dealing with observed processes that present the so-called outliers. In general, outliers can be classified into two categories, recording or gross errors and outliers due to a non-fulfillment of the model. The first case are merely classified as measurement errors, while in the second case the arguments are more subtle. In this respect, we are following the formal definition given by Davies and Gather, 1993, where an outlying or influential observation is such that does not fit the postulated model. Hence, these observations might be regular observations gathered from a nonlinear model, and judged as outliers with respect to a linear model. Therefore, one must be careful when imposing a linear structure, because this may leads to misspecification problems and yields biased estimations. Inevitably, will also yields bad fitting and forecast performances, which are of fundamental importance in time series modeling.

To still benefit from features of a multivariate setup, various techniques as been proposed, such as the construction of robust procedure that are able to deal with the occurrence of such outliers in the series. Another approach may consists in the diagnostic analysis, where the interest is moved in first detecting outlying observations and after correcting them, making estimates that are more robust and reliable. Here, an estimate is thought as produced by a robust estimator if its value do not change too much when we add some extreme observation into the original series.

In this doctoral thesis I consider the former approach, thus my primary aim is to try to give interesting insights between multivariate linear and nonlinear time series models, by offering some nonlinear and robust alternative models which may be useful to study heavy-tailed and non-Gaussian variables. Of course, since there exist vast numbers of possible nonlinear features in the observed data, I consider a more restricted concept of nonliearity. The emphasis is more concentrated on the assessment of nonlinear robust multivariate time series models which may be useful to overcome the issues discussed above. Even though robustness and nonlinear multivariate time series modeling have been vastly researched, at present, there is not much interaction between the two topics. A possible reason for the lack of the interaction between the robust statistics and multivariate time series modeling is the underdevelopment of theoretical results, such as the stability conditions and asymptotic theory of the estimators that allows for standard testing procedure. Furthermore, a second reason can be given in terms of the computational burden, since nonlinear multivariate time series models are usually considered to be computer-intensive and the possibility of unfeasible waiting-time for parameter estimations is a serious problem. Thus, I will give the motivation, definition and properties of the proposed nonlinear models.

The observations that I shall consider here are adequately described by a stochastic vector time series process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$, that is a collection of random variables. Consider the $N \times 1$ dimensional vector

$$\boldsymbol{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix}, \qquad t = 1, \dots, T.$$

Then, the general mathematical nonlinear model considered here for this stochastic vector process is

$$\boldsymbol{y}_{t+1} = \boldsymbol{\mu}(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_{t-m+1}) + \boldsymbol{\sigma}(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_{t-m+1})\boldsymbol{\epsilon}_{t+1}, \tag{1.1}$$

where the vector sequence $\boldsymbol{\epsilon}_t$ is called white noise process if its mean and its variance are constant, while all covariance components are zero, i.e., we consider the case where $\mathbb{E}[\boldsymbol{\epsilon}_t] = \boldsymbol{0}_N$, that is an $N \times 1$ vector of zeroes and $\mathbb{V}[\boldsymbol{\epsilon}_t] = \boldsymbol{I}_N$ is the identity matrix. Also, the vector-valued function $\boldsymbol{\mu}(\,\cdot\,)$ models the conditional mean of $\boldsymbol{y}_t$ given the past $\{\boldsymbol{y}_s, s < t\}$, whereas $\boldsymbol{\sigma}(\,\cdot\,)$ models the conditional deviation of $\boldsymbol{y}_t$ given the past $\{\boldsymbol{y}_s, s < t\}$. Of course, it is plausible that the lags of the vector process that enter into the conditional mean function are different from those that appears in the conditional variance function.

Although its simple and intuitive definition, models such as (1.1) can be extremely complex, since there is no general rule which define the actual form of the functions $\boldsymbol{\mu}(\,\cdot\,)$ and $\boldsymbol{\sigma}(\,\cdot\,)$. Furthermore, even more complicated model can be generated if exogenous variables are considered to be part of the respective dynamics in addition to their own lags. For sake of applicability, it is necessary to concentrate on special class of system which are fairly parsimonious, by keeping the level of complexity into a feasible region, while still providing satisfactory reflection of the reality.

Our journey in time series modeling can now enter in a more technical and restricted world, in order to provide the right tools that I hope will help the reader to understand the major aims and goals of the present work.

1.1.1 State-Space Models

A widely applied approach to analyze non linear multivariate dynamics in time series data is the general state space modeling, see e.g. Harvey, 1989 and Durbin and Koopman, 2012. This approach involves sophisticated techniques which are able to deal with non-Gaussian or nonlinear data. However, it is generally agreed that the aforementioned techniques could be too demanding in terms of computational burden, since computer-intensive methods, such as Monte Carlo Markov Chain or Importance sampling are required to estimate the parameters and then fit the model. Furthermore, the statistical properties of such models are not easy to establish and the developing of a proper asymptotic theory of the estimator is often not discussed. The motivations for the lack of a detailed theoretical analysis of nonlinear state space models can be summarized as follows.

It is well-known that in a linear framework with Gaussian random errors, state-space models are efficiently handled by the Kalman filter. This methodology was developed by Kalman, 1960 and it basically consists on a efficient algorithm based on a set of recursions for computing conditional expectations. With this methodology, one can optimally solve the filtering problem. Specifically, suppose that the state $\boldsymbol{x}_t \in \mathbb{R}^N$ at time t of a multivariate system is given by a first-order Markov process, that is a stochastic difference equation known as the transition equation. Formally we have that

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}_t \boldsymbol{x}_t + \boldsymbol{\eta}_t, \qquad t = 1, \ldots, T, \tag{1.2}$$

where the disturbance term $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{Q}_t)$. $\boldsymbol{F}_t$ is the transition matrix associated to the Markov process. In addition, now suppose that for every value of t we observe a realization of the vector process $\boldsymbol{y}_t \in \mathbb{R}^N$, that might be described via the so called measurement equation,

$$\boldsymbol{y}_t = \boldsymbol{G}_t \boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \qquad t = 1, \ldots, T, \tag{1.3}$$

where $\boldsymbol{G}_t$ contains a set of parameters and sometimes is referred as the output matrix. Again, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{H}_t)$. Furthermore $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\eta}_t^\top] = \mathbf{0}_{N\times N}$ and hence they are uncorrelated. So the transition equation of the underlying process described in (1.2) is affected by a stochastic noise, thus, it is not observable and as a result it is impossible to measure exactly for a fixed t.

As it stands, equation (1.3) reveals that the observations are noisy representation of the underlying Markov process (1.2). Thus, the filtering problem might be subsumed into the following question.

▼ Given the observations $\{\boldsymbol{y}_s, s < t\}$ satisfying (1.3), what is the best estimate of the state $\boldsymbol{x}_t$?

More precisely, consider the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ is a probability space and $\mathcal{F}_t$ is the σ-field generated by $\{\boldsymbol{y}_s, \, 0 \le s \le t\}$. We say that $\hat{\boldsymbol{x}}_t$ is the best estimate in the sense that minimize the mean square error, i.e.

$$\mathbb{E}\Big[\|\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t\|_2\Big] = \underset{\boldsymbol{k}\in\mathcal{L}^2(\mathbb{P})}{inf} \Big\{\mathbb{E}\Big[\|\boldsymbol{x}_t - \boldsymbol{k}\|_2\Big]\Big\} \tag{1.4}$$

where with $\|\cdot\|_2$ we denote the $\mathcal{L}^2 - norm$ and with $\mathcal{L}^2(\mathbb{P}) = \{\boldsymbol{k} : \Omega \to \mathbb{R}^N; \|\boldsymbol{k}\|_2 < \infty$ and $\boldsymbol{k}$ is $\mathcal{F}_t$-measurable $\}$ the Hilbert space. To link this problem with the discussion of the previous section it is insightful to consider the following simple example.

Example 1.1.1. *Suppose that the* $N \times 1$ *vector* $\boldsymbol{y}_t$ *is composed by a set on a cross-section of time series variables gathered from some experiment or collected by observing real phenomena. In practice, the real structure of the data generating process might be to complex to be specified precisely. In addition, if the time series data are collected from some experiment, they are surely subject to measurement errors, due to the fact some kind of instrument is used to register the data. Thus, we hope to find a specification which mimic the actual source of variation. In other words, we* *seek to approximate that structure with a cheaper or a simplified version of it and this can be done by assuming that the observed variables are affected by random errors. If we believe that these random errors come from a multivariate Gaussian process, the notorious signal-plus-noise state-space model can give us the desired solution. Formally, we specify a multivariate Gaussian sigal-plus-noise model with the following system of equations. For* $t = 1, \ldots, T$ *we have*

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_\epsilon^{1/2} \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N), \tag{1.5}$$

where $\boldsymbol{\mu}_{t+1}$ *is an* $N \times 1$ *vector of stochastic mean components and it is assumed to follow a* VAR(1) *specification, such that*

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\Phi} \boldsymbol{\mu}_t + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta). \tag{1.6}$$

The $N \times N$ *matrix of autoregressive (unknown) coefficients* $\boldsymbol{\Phi}$ *is the analogue of the transition matrix discussed for the general transition equation defined in (1.2).* $\boldsymbol{\epsilon}_t$ *and* $\boldsymbol{\eta}_t$ *are* $N \times 1$ *vectors of multivariate Gaussian white noise with their respective covariance matrices. As discussed above, they are independent for every* t*. Therefore, the mean vector process is not observable and we can easily tackle this problem by using the Kalman filter recursions, which give us the proper statistical tool, able to filter out the stochastic conditional mean vector process* $\{\boldsymbol{\mu}_t\}_{t \in \mathbb{N}}$ *from the noisy cross-section of time series.*

To summarize, the Kalman filter give us the mathematically optimal solution to the aforementioned problem for linear and Gaussian state-space models, which means that is defined as being that filter, from the set of all possible filters which minimizes the mean square error in (1.4). Once we drop such important assumptions we lose optimality and we must rely on approximation based on computer-intensive methods as discussed earlier. Immediately, there are a couple of crucial questions which might be raised,

- What if the noise that influence the multivariate system is not normal, but instead comes from an heavy-tailed distribution?
- What if the variables to be predicted are approximated by nonlinear functions of some observed processes?

It is obvious that misspecification of this type will induce the time series analysts to wrongly judge some of the outcome provided by the fitted model. For example, one may judge as outliers some observations which may be perfectly regular in another model. Another example of misspecification is concerned with the linearity assumption, since the possible nonlinearity of the experiment or phenomena under study may give raise to outliers. In other words, the simple Gaussian signal-plus-noise model may be too simplistic and is not able to capture the complex nature of real world studies.

Once the gaussianity of the noise processes is dropped, the Kalman filter will lose its mathematical optimality. Therefore, approximation methods are required, and in state-space models cannot be applied other than with simulation-based and Monte Carlo methods. This thesis deal with approximate filtering methods for this kind of problems. More precisely, this thesis should be considered as an attempt to strike the complexities and nonlinearities inherent in some real problems by offering a valuable alternative, which requires less computational efforts and sophistication, and still feasible for a comprehensive statistical analysis. For the sake of completeness, nonlinear state-space models are briefly reviewed in the next subsection.

1.1.2 Nonlinear State-Space Models

These models are generally defined by the same state-space representation given by the transition equation 1.2 and the measurement equation 1.3. However, instead of the simple affine transformation, we allow for general functions which explain the dynamics of the variables of interest. Of course, the vector process may still depends on its lagged values, unobservable components and strong multivariate white noise. Formally, for the vector process $\boldsymbol{y}_t \in \mathbb{R}^N$, we have the nonlinear measurement equation

$$\boldsymbol{y}_t = \boldsymbol{a}(\boldsymbol{x}_t, \boldsymbol{\epsilon}_t; \boldsymbol{\theta}), \qquad t = 1, \ldots, T, \tag{1.7}$$

and the nonlinear transition equation

$$\boldsymbol{x}_{t+1} = \boldsymbol{b}(\boldsymbol{x}_t, \boldsymbol{\eta}_t; \boldsymbol{\theta}), \qquad t = 1, \ldots, T, \tag{1.8}$$

where $\boldsymbol{x}_t \in \mathbb{R}^N$ is the underlying state at time t while $\boldsymbol{a}(\,\cdot\,)$ and $\boldsymbol{b}(\,\cdot\,)$ are general sufficiently smooth functions which are known up to the parameter vector $\boldsymbol{\theta}$. Note that the disturbance terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are now assumed to be independent and identically distributed (*IID*) with known distributions which is not necessarily Gaussian. It is straightforward to see that the multivariate system can be represented by means of conditional distributions, since the conditional distribution of the measurement equation depends on the information set generated by past values of $\boldsymbol{x}_t$ and $\boldsymbol{y}_{t-1}$, while the markovianity of the transition equation implies that it depends on the past only through the previous state $\boldsymbol{x}_{t-1}$. This may be summarized formally with the following general framework,

$$\boldsymbol{y}_t \sim p(\boldsymbol{y}_t | \boldsymbol{x}_t; \boldsymbol{\theta}), \qquad \boldsymbol{x}_t \sim g(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \boldsymbol{\theta}), \qquad \boldsymbol{x}_1 \sim g(\boldsymbol{x}_1; \boldsymbol{\theta}). \tag{1.9}$$

Thus, recursively we obtain the predictive likelihood of the bivariate system $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ by noting that

$$\mathcal{L}_T(\boldsymbol{y}_t, \boldsymbol{x}_t; \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\boldsymbol{y}_t | \boldsymbol{x}_t; \boldsymbol{\theta}) g(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}; \boldsymbol{\theta}). \tag{1.10}$$

With the above specification, we might think that predictions of $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ can be made on the basis of the past information sets $\{\boldsymbol{x}_{t-1}\}$ and $\{\boldsymbol{y}_{t-1}, \boldsymbol{x}_{t-1}\}$, but, analogously to the linear framework, the state variables in $\boldsymbol{x}_t$ are unobservable, making the predictions unfeasible. Moreover, it is also possible to note that if the process $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ is assumed to be jointly Gaussian, we are just turning back to the already discussed linear and Gaussian state-space framework. Therefore, we handle the filtering problem with Kalman filter, since the updating of the predictive likelihood is obtained by updating the conditional mean vector and the conditional covariance matrix. However, it is widely accepted that the complex nature of various phenomena hardly verify this strong assumption. Therefore, although every model is in general incorrect, we may try to give a better model which provide a satisfactory representation of the behaviour of the observed data.

To conclude, I provide an example of one of the most famous class of non linear state-space models.

Example 1.1.2. *Suppose again that we observe the* $N \times 1$ *vector process* $\{\boldsymbol{y}_t\}$ *which is composed by return series of a financial asset and we are interested in modeling and forecasting the volatilities of these financial time series. It is well-known that one can approximate these variables by using squared returns and thus, the variables that we wish to predict are nonlinear functions of the asset prices. A further complication is due to the fact that when moving from a univariate to a multivariate perspective, it is also of interest the common movements of the volatilities across markets. This kind of problems may be handled by fitting a multivariate stochastic volatility model, which can helps to shed light and explain these unobservable factors. A multivariate stochastic volatility model can be introduced into* *the return system of equations*

$$y_{it} = \epsilon_{it}(\exp\{h_{it}\})^{1/2}, \qquad i = 1, \dots, N, \qquad t = 1, \dots, T, \tag{1.11}$$

where ϵ_{it} is the i-th components of the multivariate normal vector $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})^\top$ with zero mean vector $\mathbf{0}_N$ and the covariance matrix $\boldsymbol{\Sigma}_\epsilon$ may be defined as a $N \times N$ symmetric matrix with one on the main diagonal and with ρ_{ij} as the ij-th (static) correlation coefficients in the off-diagonal. Since the focus is in modeling the dynamics of the variances, a simple causal ARMA(1,1) *process, can help us to reach this aim. Thus,*

$$h_{it} = \delta_i + \phi_i h_{i,t-1} + \eta_{it} \qquad i = 1, \dots, N, \qquad t = 1, \dots, T, \tag{1.12}$$

where, similarly to ϵ_{it}, η_{it} is the i-th components of the multivariate normal vector $\boldsymbol{\eta}_t = (\eta_{1t}, \dots, \eta_{Nt})^\top$ with zero mean vector $\mathbf{0}_N$ and the covariance matrix $\boldsymbol{\Sigma}_\eta$. δ_i and $|\phi_i| < 1$ are the ARMA(1,1) *parameters Note that the restriction on the i-th autoregressive coefficient ϕ_i, for $i = 1, \dots, N$, not only ensure the stationarity of our* ARMA(1,1) *process, but it ensures also the stationarity of the $\{y_{it}\}$ processes, since h_{it} and ϵ_{it} are independent of each other. With the above specifications, estimations can still be carried out by using the Kalman filter recursions, but, in contrast to the linear model described by the observation equation* (1.5) *and the transition equation* (1.6)*, the model is nonlinear and the standard recursions does not provide the exact solution anymore. As a result, simulation-based methods are required for statistical analysis.*

During my studies, I always struggled with the fact that Gaussian models are (mathematically speaking) tremendously elegant. However, since my interest in nonlinear time series, heavy-tailed processes and so forth, trying to fit a model such as the signal-plus noise of example 1.1.1 to this data was hopeless. On the other hand, nonlinear state-space models often lack of comprehensive statistical analysis and asymptotic theory of these models is not always available, motivating me in starting a research for some alternative nonlinear time series model which retain a similar mathematical elegance.

After this short digression, I can move the discussions to the next section, which provides the main contribution and introduce the class of score-driven models, “*le fil rouge*" of my thesis.

1.2 What is this thesis about?

This thesis is mostly devoted to the analysis and modeling of multivariate nonlinear time series. I propose novel nonlinear time series models which are useful to uncover the nature and the dependence structure of complex and heavy-tailed phenomena. My approach is based on the recent class of dynamic nonlinear models introduced by Creal, Koopman, and Lucas, 2011 and Harvey, 2013, known as Generalized Autoregressive Score (GAS) models from the former, or Dynamic Conditional Score (DCS) models from the latter. As we will see later in this thesis, the peculiarity of this general class of nonlinear time series models is that the dynamics of the time-varying parameters, are driven by the conditional score of the predictive likelihood, which is known in closed form, even if we allow for heavy-tailed conditional distributions, nonlinearities in the updating equations and so forth. An introduction with a discussion of some properties of score-driven models is given below. See also Creal, Koopman, and Lucas, 2013 and the dedicated website `www.gasmodel.com`, where the most recent research on the topic is collected.

More precisely, I contribute to the existing literature on score-driven models by extending some existing univariate models to a more general multivariate framework, providing proper theoretical and empirical justifications. My particular attention on score-driven models is motivated by the fact that with this observation-driven framework, one is able to specify powerful model that tackle the difficulties of nonlinear time series modeling with high degree of generality and clarity. Observation-driven models are in direct contrast to the state-space modeling framework introduced in the previous section. In fact, following the terminology of Cox et al., 1981, state-space models are classified as parameter-driven models. The main difference between the two classes consists in the fact that the time-varying parameter in parameter-driven models is not observable, even if the real data generating process were known. On the other hand, with an observation-driven framework, after conditioned the observed data and estimated the required parameters, one is able to retrieve the time-varying parameter path. As a consequence, one is also able to obtain the predictive likelihood in closed form, making the most classical maximum likelihood estimation methods totally feasible, even in highly nonlinear and multivariate systems. This constitutes the main advantage of the latter class with respect to the former. For a recent review and comparison between the two class of models interested readers are referred to the insightful paper of Koopman, Lucas, and Scharth, 2016.

Finally, we are ready to introduce and formally specify the class of score-driven models.

1.2.1 Score-Driven Models

This thesis heavily leans from the successfully score-driven class of nonlinear observation-driven models and it should be thought as an attempt to contribute on the existing literature with the definition of novel multivariate models. For each model I provide a detailed statistical analysis enriched with theoretical and empirical results.

For the sake of easy notation, I present a univariate score-driven model in full generality: Consider model for the observations $Y_t = \{y_1, \ldots, y_t\}$, where the behaviour of the process $\{y_t\}$ is explained with the observation equation

$$y_t \sim p(y_t | Y_{t-1}, f_t, \boldsymbol{\theta}). \tag{1.13}$$

Analogously to nonlinear state space models, the behaviour of $\{y_t\}$ is explained using a time-varying parameter, denoted with $\{f_t\}$. However, score-driven models are observation-driven models, thus the update of f_{t+1} depends on the observed y_t, such that

$$f_{t+1} = \psi(y_t, f_t), \tag{1.14}$$

where $\psi(\,\cdot\,)$ is a general continuous map. Note that, (1.14) is in direct contrast with (1.8) since the latter depends on exogenous innovation, but likewise, this is a postulated structure, that means that it does not necessarily reflect the real data generating process.

In order to describe the dynamics of the time-varying $\{f_t\}$ with a concrete example, the updates of its values may be presented with a parametric map. The recursions are commonly specified by an affine transformation of the type

$$f_{t+1} = \omega + A s_t + B f_t, \tag{1.15}$$

where the innovation or driving force is given by $s_t = S_t \nabla_t$, and

$$\nabla_t = \frac{\partial \ln p(y_t | Y_{t-1}, f_t, \boldsymbol{\theta})}{\partial f_t} \quad \textit{and} \quad S_t = -\mathbb{E}\left[\frac{\partial^2 \ln p(y_t | Y_{t-1}, f_t, \boldsymbol{\theta})}{\partial^2 f_t}\right]^{-1}. \tag{1.16}$$

It is easy to see that $\mathbb{E}_{t-1}[\nabla_t] = 0$ is a martingale difference (MD) and the updating recursion of the time-varying parameter is a measurable function of the past scores and therefore of the observations. The static (scalar) parameters ω, A and B are unknown and need to be estimated. Therefore, $\boldsymbol{\theta}$ is defined as the vector of static parameters that involve the parameters of the assumed conditional distributions and the parameters needed for the functional form of f_t.

The interpretation of these parameters is similar to those of causal autoregressive moving average (*ARMA*) processes. Thus, ω is the intercept and B is the autoregressive parameter, which has to be restricted in order to ensure the stationarity of the recursion. As regard the A coefficient, there exists no direct interpretation. In general, one may note that it should be restricted in order to maintain the time-varying parameter in its appropriate domain, for example, positive values for variances. On the other hand, if a dynamic mean is of interest, the parameter A may (in principle) assume any value of the real line and therefore, a general rule is impossible to be defined. However, this problem can be easily avoided by allowing the time-varying f_t to move freely on the whole real line and then maps the assumed values into its proper space by using appropriate link functions. Consider again the volatility case, an exponential link function will do the right job.

In conclusion, the likelihood function is given by

$$\mathcal{L}_T(\boldsymbol{\theta}) = \prod_{t=1}^{T} p(y_t|Y_{t-1}, f_t, \boldsymbol{\theta}), \tag{1.17}$$

which depends on both the time-varying parameter and a vector of static parameters. It is possible to acknowledge the differences with the state-space counterpart in (1.10).

Given the specifications above, extension to multivariate framework is straightforward. One has only to note that, similar to the state-space set-up, where the underlying state vector $\boldsymbol{x}_t$ contains the unobserved factors, in score-driven models the time-varying parameters are stacked in the vector $\boldsymbol{f}_t$. Thus, instead of simple univariate coefficients ω, A and B, one may define the updating equation with matrix of coefficients with proper dimensions and the required restrictions.

As already argued, this thesis is intended to offer definitions and detailed discussions of several of these models and hence it will be redundant to offer a concrete example of some specific score-driven model in this part of our journey. For this reason, concrete examples and applications of the multivariate versions of the set-up described above, is postponed to the subsequent chapters. Therefore, I conclude the introduction with the next section, where it is summarized the structure of the present thesis.

1.3 Structure of the Thesis

I conclude the introduction with this section, where I give some detail about each chapter of the present thesis.

- *Chapter 2:* (based on joint work with Prof. A. Luati and Prof M. Mazzocchi)

  A novel multivariate score-driven model is proposed to extract signals from noisy vector processes. By assuming that the conditional location vector from a multivariate Student's t distribution changes over time, we construct a robust filter which is able to overcome several issues that naturally arise when modeling heavy-tailed phenomena and, more in general, vectors of dependent non-Gaussian time series. We derive conditions for stationarity and invertibility and estimate the unknown parameters by maximum likelihood. Strong consistency and asymptotic normality of the estimator are proved and the finite sample properties are illustrated by a Monte Carlo study. From a computational point of view, analytical formulae are derived, which consent to develop estimation procedures based on the Fisher scoring method. As an empirical illustration, we show how the model can be effectively applied to estimate consumer prices from home scanner data.

- *Chapter 3:* (based on joint work with Prof. F. Blasques and Prof S. J. Koopman)

  Factor volatility models which decompose the conditional volatility of a N-variate time-series into a common factor and N idiosyncratic components can offer a useful and parsimonious way of handling multivariate time-series and panels of data. While the common factor is naturally robust to fat-tailed innovations due to averaging of information over the N cross-sectional elements, the idiosyncratic components are typically sensitive to outliers. We propose a nonlinear dynamic factor model for conditional volatilities with score-robust updating equation for the idiosyncratic components with substantially improved fit on data sets with fat tailed innovations. We derive stochastic properties for the model, including bounded moments, stationarity, ergodicity, and filter invertibility. We also establish consistency and asymptotic normality of the maximum likelihood estimator in large samples. Additionally, we study the small sample properties of the estimator by means of a Monte Carlo study. Finally, we provide an empirical illustration using a panel of ten stocks from the S&P500 which highlights the advantages of the proposed dynamic factor structure, as well as the need for robust filtering techniques for the idiosyncratic component.

▼ *Chapter 4:*(based on joint work with Prof. F. Blasques and Prof A. Lucas)

It is well known that taking nonlinear dynamic filters from univariate to multivariate settings is challenging due to the likely failure of the sufficient contraction conditions used to ensure filter invertibility. Our main contribution is to show that only invertible filters may provide reliable and stable maximum likelihood estimators. We first review two newly introduced exponential multivariate score driven models for conditional volatilities and we also incorporate a time-varying conditional correlation coefficient, as a possible extension of the model. After that, we propose an empirical method to verify that the invertibility condition holds. Our method is simple and can be used in practice to find an empirical invertible domain. Therefore, constrained optimization routine can be performed in order to obtain reliable maximum likelihood estimators. This is crucial to ensure that the estimation and filtering results are not spurious. We show the practical relevance of our results by means of an empirical application to stock returns.

Chapter 2

A Robust Filter for Multivariate Time Series with an Application to Consumer Price Estimates from Homescan Data

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name
Department of Statistical Sciences "Paolo Fortunati"

Doctor of Philosophy in Statistical Sciences

Essays in Robust and Nonlinear Multivariate Time Series Models

by Enzo D'INNOCENZO

A novel multivariate score-driven model is proposed to extract signals from noisy vector processes. By assuming that the conditional location vector from a multivariate Student's *t* distribution changes over time, we construct a robust filter which is able to overcome several issues that naturally arise when modeling heavy-tailed phenomena and, more in general, vectors of dependent non-Gaussian time series. We derive conditions for stationarity and invertibility and estimate the unknown parameters by maximum likelihood. Strong consistency and asymptotic normality of the estimator are proved and the finite sample properties are illustrated by a Monte Carlo study. From a computational point of view, analytical formulae are derived, which consent to develop estimation procedures based on the Fisher scoring method. As an empirical illustration, we show how the model can be effectively applied to estimate consumer prices from home scanner data.

2.1 Introduction

The analysis of multivariate time series has a long history, due to the empirical evidence from most research fields that time series resulting from complex phenomena do not only depend on their own past, but also on the history of other variables. For this reason, from Hannan, 1970, the literature on multivariate time series has grown very fast. The leading example is the dynamic representation of the conditional mean of a vector process which gives rise to vector autoregressive processes (see Hamilton, 1994 and Lütkepohl, 2007).

Following the taxonomy proposed in Cox et al., 1981, two main classes of models can be considered when modeling multivariate phenomena, parameter-driven and observation-driven models. The first class consists of parameter-driven models. It is a broad class, which involves the widely applied state-space models, or unobserved component models (Harvey, 1989; West and Harrison, 1997). Within this framework, parameters are allowed to vary over time as dynamic processes driven by idiosyncratic innovations. Hence, likelihood functions are analytically tractable only in specific cases, notably linear Gaussian models, where inference can be handled by the Kalman filter. On the other hand, parameter-driven models are very sensitive to small deviations from the distributional assumptions. In addition, the Gaussian assumption often turns out to be too restrictive, and flexible distributions may be more appropriate. Thus, a fast growing field of research is dealing with nonlinear or non-Gaussian state-space models, resting on computer intensive simulation methods like the particle filter discussed in Durbin and Koopman, 2012. Although these methods provide extremely powerful instruments for estimating nonlinear and/or non-Gaussian models, they can be computationally too demanding. Furthermore, it may be difficult to derive the statistical properties of the implied estimators, due to the complexity of the joint likelihood function.

In contrast, in observation-driven models, the dynamics of time varying parameters depend on deterministic functions of lagged variables. This allows for a stochastic evolution of the parameters which becomes predictable given the past observations. Koopman, Lucas, and Scharth, 2016 assess the performances and optimality properties of the two classes of models, in terms of their predictive likelihood. The main advantage of observation-driven models is that the likelihood function is available in closed form, even in nonlinear and/or non-Gaussian cases. Thus, the asymptotic analysis of the estimators becomes feasible and computational costs are reduced drastically.

Within the class of observation-driven models, score-driven models are a valid option for modeling time series that do not fall in the category of linear Gaussian processes. Examples have been proposed in the context of volatility estimation and originally referred to as generalised autoregressive score models (GAS, Creal, Koopman, and Lucas, 2011), and as dynamic conditional scores (DCS, Harvey, 2013). The key feature of these models is that the dynamics of time-varying parameters are driven by the score of the conditional distribution, which needs not necessarily be Gaussian but can be heavy tailed. For example, it may follow a Student's t distribution (Harvey and Luati, 2014), an exponential generalized beta distribution (Caivano, Harvey, and Luati, 2016), a binomial distribution as in the vaccine example by Hansen and Schmidtblaicher, 2019, or represented by a mixture (Lucas, Schaumburg, and Schwaab, 2019). Further applications are discussed in Creal, Koopman, and Lucas, 2013. The optimality of the score as a driving force for time varying parameters in observation-driven models is discussed in Blasques, Koopman, and Lucas, 2015a. According to which conditional distribution is adopted, specific situations may be conveniently handled due to the properties of the score. As an example for the univariate case, if a heavy-tailed distribution is specified (e.g. Student's t), the resulting score-driven model yields a simple and natural model-based signal extraction filter which is robust to extreme observations, without any external interventions or diagnostics like outlier detection and dummy variables (Harvey and Luati, 2014).

In score-driven models, as in all observation-driven models, the time varying parameters are updated by filtering procedures, i.e. weighted sums of functions of past observations, given some initial conditions that can be fixed or estimated along with the static parameters. A robust filtering procedure should assign less weight to extreme observations in order to prevent biased inference of the signal and parameters. In particular, the work of Calvet, Czellar, and Ronchetti, 2015 provides a remarkable application of robust methods when dealing with contaminated observations. The authors show that a substantial efficiency gain can be achieved by huberizing the derivative of the log-observation density, then integrating it. As we show in the present study, the same holds if one considers an alternative robustification method, based on the specification of a conditional multivariate Student's *t* distribution. A similar approach can be found in Prucha and Kelejian, 1984 and Fiorentini, Sentana, and Calzolari, 2003, where the multivariate Student's *t* distribution provides a valid alternative to relax the normality assumption of the distribution.

In this paper, we specify a score driven model for the time-varying location of a multivariate Student's *t* distribution. We envisage three main contributions to the existing literature.

The first contribution is the full derivation of the probabilistic and asymptotic theory behind the multivariate dynamic score-driven model for conditional Student's *t* distributions, including the conditions of stationarity, ergodicity and invertibility. Also, we prove the strong consistency and asymptotic normality of the maximum likelihood estimators of the static parameters.

The second contribution is the development of an estimation scheme grounded on Fisher's scoring method, based on closed-form analytic expressions, which can be directly implemented into any statistical or matrix-friendly software. Our computations are made using the software `R` and all our files are available upon request.

The third contribution of the paper is an innovative application, dealing with estimation of regional consumer prices based on home scanner data. The use of scanner data to compute official consumer price indices (CPIs) is gaining popularity, because of their timeliness and a high level of product and geographical detail (Feenstra and Shapiro, 2003b). However, they also suffer from a variety of shortcomings, which make time series of scanner data prices (SDPs) potentially very noisy, especially when they are estimated for population sub-groups, or at the regional level (Silver, 1995). There is extensive research and a lively debate on the issues related to the computation and use of scanner data based CPIs. In a dedicated session of the 2019 meeting of the the Ottawa Group on Price Indices, it has been suggested [1] to adopt model-based filtering techniques to extract the signal from scanner-based time series of price data. These filtered estimates lose the classical price index formula interpretation, but are expected to deliver the same information content with a better signal-to-noise ratio. We show that our robust multivariate model, applied to SDPs, provides information on the dynamics of the time series and on their interrelations without being affected from outlying observations, which are naturally downweighted in the updating mechanism.

The paper is organized as follows. In Section 2.2 the model is specified. Section 4.3 deals with the stochastic properties: stationarity and invertibility conditions are derived along with bounds for the moments. In Section 2.4 maximum likelihood estimation is discussed. The asymptotic theory is derived in Section 2.4.1 and the computational aspects are discussed in Section 2.4.2. Finite sample properties are analysed in Section 2.5. The empirical analysis is reported in section 4.7. Some concluding remarks are drawn in Section 2.7. The main proofs are collected in Appendix A.1, while the relevant quantities for the implementation of the Fisher scoring algorithm as well as the proofs of the Lemmata are deferred to Appendix A.2 and Appendix A.3 respectively.

2.2 The Multivariate Student's *t* Location Model

Let us consider a vector of $N \geq 1$ stochastic processes $\boldsymbol{y}_t \in \mathbb{R}^N$ and let $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots\}$ be its filtration at time $t-1$. We assume that the process is generated by a conditional Student's t distribution with $\nu > 0$ degrees of

[1] See Jens Mehroff presentation at `https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/towards_a_new_paradigm_for_scanner_data_price_indices_0.pdf`

freedom,

$$f(\boldsymbol{y}_t|\mathcal{F}_{t-1}) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{N/2}}|\boldsymbol{\Omega}|^{-1/2}\left[1+\frac{(\boldsymbol{y}_t-\boldsymbol{\mu}_t)^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_t)}{\nu}\right]^{-(\nu+N)/2} \tag{2.1}$$

where $\boldsymbol{\mu}_t$ is a time varying location vector, $\boldsymbol{\Omega}^{1/2}$ is the scale matrix of $\boldsymbol{y}_t$, that we assume here to be static. A location-scale representation of $\boldsymbol{y}_t$ is the following,

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t, \qquad t = 1,\dots,T, \tag{2.2}$$

where $\boldsymbol{\epsilon}_t \sim \boldsymbol{t}_\nu(\boldsymbol{0}_N, \boldsymbol{I}_N)$, with $\boldsymbol{0}_N$ the null vector of $\mathbb{R}^N$ and $\boldsymbol{I}_N$ the $N \times N$ identity matrix. The well known relation holds between the scale matrix $\Omega^{1/2}$ and the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{y}_t$, $\boldsymbol{\Omega} = (\nu/(\nu-2))\boldsymbol{\Sigma}$. As we have adopted the parameterisation based on the scale matrix, the location vector always exists. In contrast, the conditional mean exists for $\nu > 1$. If the representation based on the covariance matrix is preferred, then the stronger restriction $\nu > 2$ has to be imposed.

Our interest is in recovering $\boldsymbol{\mu}_t$ based on a set of observed vector of time series from $\boldsymbol{y}_t$, for $t = 1,\dots,T$. With no distributional assumptions on the dynamics of $\boldsymbol{\mu}_t$, a filter can be specified, $\boldsymbol{\mu}_{t+1|t} = \phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})$, i.e. a stochastic recurrence equation that approximates the path of $\boldsymbol{\mu}_t$, based on some function ϕ of the past observations and a set of static parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, that possibly include a starting value, say $\boldsymbol{\mu}_{1|0}$. The subscript notation $t|t-1$ is used to emphasise the fact that $\boldsymbol{\mu}_{t|t-1}$ is an approximation of the dynamic location process at time t given the past, that is equivalent to say that $\boldsymbol{\mu}_{t|t-1}$ is $\mathcal{F}_{t-1}$-measurable. The specification of the mapping ϕ usually involves some autoregressive scheme for the evolution of the dynamic parameter combined with the specification of a driving force, usually a highly nonlinear function of the past observations that forms a martingale difference sequence, playing an analogous role of the error term in parameter-driven models.

In this paper, we approximate the temporal changes of the conditional location vector by relying on the score-driven framework of Creal, Koopman, and Lucas, 2011 and Harvey, 2013, as follows,

$$\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega} = \boldsymbol{\Phi}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\boldsymbol{u}_t, \tag{2.3}$$

where $\boldsymbol{\mu}_{t|t-1} = (\mu_{1,t|t-1}, \dots, \mu_{N,t|t-1})^\top$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N)^\top$ is the N-dimensional vector of unconditional means, $\boldsymbol{\Phi}$ and $\boldsymbol{K}$ are $N \times N$ matrices of coefficients and the driving force, $\boldsymbol{u}_t$, is a martingale difference sequence proportional to the score of the conditional likelihood of $\boldsymbol{\mu}_t$,

$$\boldsymbol{u}_t = \frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})}{1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu}.$$

By differentiating with respect to $\boldsymbol{\mu}_t$ the logarithm of the density in equation (2.1), one gets the claimed proportionality relation, between the score vector and the martingale difference sequence $\boldsymbol{u}_t$, namely

$$\frac{\partial \ln f(\boldsymbol{y}_t|\mathcal{F}_{t-1})}{\partial \boldsymbol{\mu}_{t|t-1}} = \boldsymbol{\Omega}^{-1}\frac{\nu+N}{\nu}\boldsymbol{u}_t.$$

The score as the driving force in an updating equation for a time varying parameter is the key feature of score driven models. The rationale behind the recursion (2.3) is very intuitive. Analogously to the Gauss-Newton algorithm, it improves the model fit by pointing in the direction of the greatest increase of the likelihood.

In the context of location estimation under the Student's t assumption, a further relevant motivation for the score driven methodology lies in the robustness of the implied filters. Indeed, one can write,

$$\boldsymbol{u}_t = (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/w_t, \tag{2.4}$$

where the positive scaling factors $w_t = 1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu$ are scalar weights that involve the Mahalanobis distance. They possess the role of re-weighting the large deviation from the mean produced by the difference $\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}$. The bulk of the robustness comes precisely from this component.

A formal proof of this statement follows from the following crucial lemma, which satisfies the sufficient conditions for a filter to be robust, see Calvet, Czellar, and Ronchetti, 2015.

Lemma 2.2.1 (Uniformly Boundedness and Moments of the Score). *For $0 < \nu < \infty$, the vector sequence $\{\boldsymbol{u}_t\}$ is uniformly bounded in a vector $\mathbf{u}_t$ which is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, that is* $\sup_t \mathbb{E}[\|\boldsymbol{u}_t\|] < \infty$. *Hence, it possess all the even moments*

$$\mathbb{E}\left[\overset{m}{\otimes}\boldsymbol{u}_t\right] = \left(\overset{m}{\otimes}\boldsymbol{\Omega}^{1/2}\right)\frac{\mathcal{B}\left(\frac{N+m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)}\left(\frac{\nu}{N}\right)^{m/2},$$

where $\left[\overset{m}{\otimes}\boldsymbol{x}\right]$ *denote* $\boldsymbol{x}\otimes\cdots\otimes\boldsymbol{x}$ *m-times and* $\otimes$ *is the Kronecker product,* $\mathcal{B}(\alpha,\beta)$ *is the usual Beta function and*

$$b_t = \frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu}{1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu}, \quad 0 \le b_t \le 1, \quad \text{with} \quad b_t \sim \mathcal{B}eta\left(\frac{N}{2},\frac{\nu}{2}\right). \tag{2.5}$$

Proof. See Appendix A.1.1. □

The moment structure reveals important features of the innovation vector $\boldsymbol{u}_t$. In fact, it forms a vector martingale differences sequence, with zero mean vector and (vec)-variance covariance matrix,

$$\mathbb{E}[\boldsymbol{u}_t \otimes \boldsymbol{u}_t] = \operatorname{vec}\mathbb{E}[\boldsymbol{u}_t\boldsymbol{u}_t^\top] = \frac{\nu^2}{(\nu+N)(\nu+N+2)}\operatorname{vec}\boldsymbol{\Omega}.$$

Furthermore, the $\boldsymbol{u}_t$ are identically distributed. Multi-step forecasts can be straightforwardly obtained as

$$\mathbb{E}[\boldsymbol{y}_{T+l}|\mathcal{F}_T] = \mathbb{E}[\boldsymbol{\mu}_{T+l|T+l-1}\mathcal{F}_T] = \boldsymbol{\omega} + \sum_{j=1}^{l-1}\boldsymbol{\Phi}^j(\boldsymbol{\mu}_{T+1|T} - \boldsymbol{\omega}).$$

2.3 Stochastic Properties

We begin this section by giving conditions under which the multivariate DCS-t Location Model produces stationary and ergodic paths, i.e. we derive the model stochastic properties as a data generating process from which we obtain the moments structure of the model. Then, we turn the discussion to the invertibility. This latter property is of fundamental importance for estimation and prediction, especially in nonlinear multivariate models, in that it ensures that one can consistently approximate the real path of the dynamic location vector by some measurable function of the past information.

2.3.1 Stationarity, Ergodicity and Moments of the Process

Let us express the dynamics of the signal in terms of the innovations, which, in our case, is equivalent to rewriting equation (2.3) as

$$\boldsymbol{\mu}_{t+1} - \boldsymbol{\omega} = \boldsymbol{\Phi}(\boldsymbol{\mu}_t - \boldsymbol{\omega}) + \boldsymbol{K}\frac{\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t}{1+\boldsymbol{\epsilon}_t^\top\boldsymbol{\epsilon}_t/\nu}, \qquad t \in \mathbb{Z} \tag{2.6}$$

where we have removed the subscript $t|t-1$ since we now interpret equation (2.6) as a Markov chain. With this specification, it is evident that the driving-force is independent of $\boldsymbol{\mu}_t$ because so is the prediction error $\{(\boldsymbol{y}_t - \boldsymbol{\mu}_t)\}_{t\in\mathbb{Z}}$ and the stationary ergodic sequence $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$. This implies that one can generate a stationary and ergodic vector process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$, which satisfies (2.1), (2.2) and (2.3), by drawing *IID* sequences from the assumed multivariate Student's t and then plugging this sequence into the transition mechanism in (2.6). Equation (2.6) allows us to derive the stochastic properties of our score-driven model, summarized in the next Lemma.

Lemma 2.3.1 (Stationarity, Ergodicity and Moments of the Dynamic Location). *Consider the recursion (2.6) and let $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic vector sequence. Assume that $\varrho(\boldsymbol{\Phi}) < 1$ and $\det \boldsymbol{K} \neq 0$, where $\varrho(\boldsymbol{X})$ denote the spectral radius of any $N \times N$-dimensional real matrix. Then (2.6) admits a unique strictly stationary solution $\{\boldsymbol{\mu}_t\}_{t\in\mathbb{Z}}$ with representation*

$$\boldsymbol{\mu}_{t+1} - \boldsymbol{\omega} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{K}\frac{\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t}{1+\boldsymbol{\epsilon}_{t-j}^\top\boldsymbol{\epsilon}_{t-j}/\nu}.$$

Furthermore, $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\mu}_t\|^m] < \infty$ for every $m > 0$.

Proof. See Appendix A.1.1. □

The stability condition $\varrho(\boldsymbol{\Phi}) < 1$, is a well-known condition in the theory of linear systems, see Hannan and Deistler, 1987, Hannan, 1970 or Lütkepohl, 2007. This condition, however, also extends to our nonlinear model.

As a consequence of Lemma 2.3.1, we derive the moment structure of the process (2.2).

Lemma 2.3.2 (Bounded Moments). *Consider model (2.2) and let $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic vector sequence. Assume that $\varrho(\boldsymbol{\Phi}) < 1$ and $\det \boldsymbol{K} \neq 0$. If $\mathbb{E}[\|\boldsymbol{\epsilon}_t\|^m] < \infty$, for all $m > \nu - \delta$, then $\mathbb{E}[\|\boldsymbol{y}_t\|^m] < \infty$.*

Proof. See Appendix A.1.1 □

2.3.2 Invertibility of the Filter

From a filtering point of view, it is convenient to expand equation the compact form of $\boldsymbol{u}_t$ in equation (2.3), such that

$$\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega} = \boldsymbol{\Phi}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})}{1+(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu}, \qquad t \in \mathbb{N}. \tag{2.7}$$

Starting at some initial value, $\boldsymbol{\mu}_{1|0}$, and using equation (2.7) for $t = 1, \ldots, T$ one can recover a unique filtered path $\{\widehat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. A desirable property is that the initial values used to start the whole process are asymptotically negligible, in the sense that as t increases, the impact of the chosen $\boldsymbol{\mu}_{1|0}$ eventually vanishes, i.e. the process will converge to a unique stationary and ergodic solution. Moreover, an invertible model allows one to consistently estimate the innovations, that are typically obtained from $\boldsymbol{v}_t = \boldsymbol{y}_t - \mathbb{E}[\boldsymbol{y}_t|\mathcal{F}_{t-1}]$.

The seminal paper of Bougerol, 1993 provides a comprehensive analysis on the contraction conditions and the stochastic behaviour of the Kalman filter with random coefficients. Lately, Straumann, 2005 and Straumann and Mikosch, 2006 provide an extensive discussion on the stationarity and invertibility conditions with several applications to different classes of *GARCH* models. Basically, these latter authors rely on Theorem 3.1 of Bougerol, 1993 in order to develop an unified asymptotic analysis of the mentioned models based on the theory of the stochastic recurrence equations. Furthermore, a detailed discussion about approximation concepts and asymptotic theory of general dynamic nonlinear models, with particular focus to econometric applications can be found in Pötscher and Prucha, 1997.

We follow the theory developed by Pötscher and Prucha, 1997 and give sufficient conditions for invertibility in the present setting.

Lemma 2.3.3 (Invertibility of the Dynamic Location Filter). *Consider the model specified by equations* (2.1), (2.2) *and* (2.3). *Let* $\{\epsilon_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence. Let the conditions of Lemma 2.3.1 hold true and consider the filtering equation* (2.7). *Assume that*

1. $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\prod_{k=1}^{j}\frac{\partial\mu_{t-k+1|t-k}}{\partial\mu_{t-k|t-k-1}^{\top}}\right\|\right] < 1$, *for some* $j \geq 1$ *large enough and*
2. $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\mu_{1|0}}{\partial\theta^{\top}}\right\|\right] < \infty$.

Then, the filtered location vector $\{\hat{\mu}_{t|t-1}\}_{t\in\mathbb{N}}$ *is invertible and converges exponentially almost surely to the unique stationary ergodic solution* $\{\mu_{t|t-1}\}_{t\in\mathbb{Z}}$ *for any initialization of the filtering recursion,* $(\mu_{1|0} - \omega)$.

Furthermore, $\mathbb{E}[\sup_{\theta\in\Theta}\|\hat{\mu}_{t|t-1}\|^m] < \infty$ *and* $\mathbb{E}[\sup_{\theta\in\Theta}\|\mu_{t|t-1}\|^m] < \infty$ *for every* $m > 0$.

Proof. See Appendix A.1.1. □

2.4 Maximum Likelihood Estimation

Consider the stationary ergodic process $\{y_t\}_{t\in\mathbb{Z}}$, that satisfies (2.1) and (2.2). The unconditional distribution of $(\mu_1^{\top}, \mu_2^{\top}, \ldots, \mu_T^{\top})^{\top}$ is not known and the same holds for the distribution of $(y_1^{\top}, y_2^{\top}, \ldots, y_T^{\top})^{\top}$. However, conditionally on some nonrandom starting value for the dynamic location, $\mu_{1|0}$, and recursively applying equation (2.3), the conditional distribution of $y_1, \ldots, y_T$, is characterized by the distribution of the *IID* random vector ϵ_t, implying that the log-likelihood function for a single observation has the form

$$\ell_t(\theta) = \ln\Gamma\left(\frac{\nu+N}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\ln(\pi\nu) - \frac{1}{2}\ln|\Omega| - \frac{\nu+N}{2}\ln\left[1 + \frac{(y_t-\mu_{t|t-1})^{\top}\Omega^{-1}(y_t-\mu_{t|t-1})}{\nu}\right], \tag{2.8}$$

where $\theta = (\xi^{\top}, \psi^{\top})^{\top} \in \Theta \subseteq \mathbb{R}^p$, $\xi = (\nu, (\mathrm{vech}(\Omega))^{\top}, \omega^{\top})^{\top}$ and $\psi = ((\mathrm{vec}\,\Phi)^{\top}, (\mathrm{vec}\,K)^{\top})^{\top}$. The dimension of the p-vector of unknown parameters θ is thus determined by the dimensions of $\xi \in \mathbb{R}^s$, with $s = N + \frac{1}{2}N(N+1) + 1$ and $\psi \in \mathbb{R}^d$, with $d = (N \times N) + (N \times N)$, hence $p = s + d$.

Lemma 2.3.3, ensures that the initial conditions for the function $\mu_{t|t-1}$ are asymptotically equivalent such that, once the choice for its starting value is being made, it is possible to obtain an approximated version of the conditional log-likelihood, $\hat{\ell}_t(\theta)$, by replacing $\mu_{t|t-1}$ in equation (2.8) by the approximated dynamic location $\hat{\mu}_{t|t-1}$. Hence, for the whole sample we obtain

$$\hat{\ell}_T(\theta) = \sum_{t=1}^{T}\hat{\ell}_t(\theta) \tag{2.9}$$

and the *MLE* of θ is

$$\widehat{\theta}_T = \arg\max_{\theta\in\Theta}\hat{\ell}_T(\theta).$$

2.4.1 Asymptotic Theory

It is worth noting that even if we restrict ourselves to the case of diagonal scale matrix $\mathbf{\Omega}$, the random vectors $\boldsymbol{\epsilon}_t$ are uncorrelated, but they still maintain their tail-dependence. Features of these kind may be further appreciated under the following stochastic representations (see also Fang, Kotz, and Ng, 1990)

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{\Omega}^{1/2}\frac{\boldsymbol{z}_t}{\sqrt{s_t/\nu}}, \qquad \boldsymbol{z}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N), \qquad s_t \sim \chi^2_\nu, \tag{2.10}$$

and

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{\Omega}^{1/2}\mathbf{u}_t\sqrt{\frac{r_t}{s_t/\nu}}, \ \mathbf{u}_t \sim \textit{Uniformly Distributed on a Unit Sphere in } \mathbb{R}^N,$$
$$s_t \sim \chi^2_\nu, \ \textit{and} \ r_t \sim \chi^2_N. \tag{2.11}$$

We present our version of strong consistency and asymptotic normality of the ML estimator of the proposed model when $T \to \infty$ and N is fixed.

Assumption 1. *1.* $\varrho(\mathbf{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,

2. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}\left\|\prod_{k=1}^{j}\frac{\partial\boldsymbol{\mu}_{t-k+1|t-k}}{\partial\boldsymbol{\mu}_{t-k|t-k-1}^\top}\right\|\right] < 1$, *for some* $j \geq 1$ *large enough and*
3. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}\left\|\frac{\partial\boldsymbol{\mu}_{1|0}}{\partial\boldsymbol{\theta}^\top}\right\|\right] < \infty$,
4. *the parameter space* $\mathbf{\Theta}$ *is compact with* $2 < \nu < \infty$,
5. $\forall\boldsymbol{\theta} \in \mathbf{\Theta}$, *if* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *then* $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}) \neq \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0)$ *almost surely and* $\forall t \geq 1$.

The next results is concerned with the consistency of $\hat{\boldsymbol{\theta}}_T$.

Theorem 2.4.1. *Consider model (2.1), (2.2) and (2.3). Let* $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence and, further, suppose that Assumption 1 hold. Then,*

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \qquad \textit{as} \qquad T \to \infty.$$

Proof. See Appendix A.1.2. □

We now turn to asymptotic normality.

Theorem 2.4.2. *Consider model (2.1), (2.2) and (2.3). Let* $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence and, further, suppose that the conditions under witch the strong consistency of the maximum likelihood estimator holds. Then,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \Rightarrow \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1}),$$

where,

$$\mathcal{I}(\boldsymbol{\theta}_0) = -\mathbb{E}\left[\left.\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right],$$

is the Fisher's Information matrix evaluated at the true parameter vector $\boldsymbol{\theta}_0$.

Proof. See Appendix A.1.2. □

In light of the consistency of the ML estimator obtained with Theorem 2.4.1, we can consistently estimate $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ by

$$\widehat{\boldsymbol{\mathcal{I}}}(\hat{\boldsymbol{\theta}}_T) = -\frac{1}{T}\sum_{t=1}^{T}\left[\frac{d^2\hat{\ell}_t(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_T}\right]. \tag{2.12}$$

The general formula for the second partial derivatives in (2.12) has the form below

$$\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top} = \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} + \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)^\top \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right) + \frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}^\top}\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top}, \tag{2.13}$$

since the dynamic location and its derivatives, are nonlinear functions of the vector of parameters $\boldsymbol{\theta}$. However, we note that the assumption of correct specification implies that the score vector forms a martingale difference sequence. In addition, we know that the dynamic location (and its derivatives) are $\mathcal{F}_{t-1}$-measurable and therefore the last component of equation (2.13) will cancel out after the application of the conditional expectation operator $\mathbb{E}_{t-1}[\,\cdot\,]$.

By the law of iterated expectations we have

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\mathcal{I}}_t(\boldsymbol{\theta})] = \lim_{n\to\infty}\mathbb{E}_{t-n}\left\{\ldots\mathbb{E}_{t-2}\left[\mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top}\right]\right]\right\},$$

which, in turn, can be consistently estimated with

$$\widehat{\boldsymbol{\mathcal{I}}}_T(\boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{t-1}\left[\frac{d^2\hat{\ell}_t(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_T}\right].$$

As a matter of fact, the latter estimator $\widehat{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta})$ is strongly consistent, i.e.

$$\widehat{\boldsymbol{\mathcal{I}}}_T(\boldsymbol{\theta}) \xrightarrow{\text{a.s.}} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0) \qquad as \qquad T\to\infty,$$

and moreover, is easier and more stable to implement than $\widehat{\boldsymbol{\mathcal{I}}}(\boldsymbol{\theta})$, since it avoids the recursive evaluation of the second derivatives of the dynamic location vector.

ML estimation and inference is carried out by means of Fisher's scoring method. The development of a proper iterative procedure requires explicit formulas for the score vector and the Hessian matrix. In the following section, we discuss the computational aspects related to ML estimation.

2.4.2 Computational Aspects

A strongly reliable Fisher-scoring method based on analytical formulas (reported in Appendix A.2) is developed, which can be directly implemented in any statistical package through the following steps:

Step 1: Choose a starting value $\widehat{\boldsymbol{\theta}}_T^{(0)} = (\nu^{(0)}, (\text{vech}(\boldsymbol{\Omega})^{(0)})^\top), \boldsymbol{\omega}^{(0)\top}, (\text{vec}(\boldsymbol{\Phi})^{(0)})^\top), (\text{vec}(\boldsymbol{K})^{(0)})^\top))^\top$

Step 2: For $h>0$, update $\widehat{\boldsymbol{\theta}}_T^{(h)}$ using the scoring rule

$$\widehat{\boldsymbol{\theta}}_T^{(h+1)} = \widehat{\boldsymbol{\theta}}_T^{(h)} + \left[\widehat{\boldsymbol{\mathcal{I}}}_T(\boldsymbol{\theta})^{(h)}\right]^{-1}\widehat{\boldsymbol{s}}_T(\boldsymbol{\theta})^{(h)},$$

where here the score vector and the conditional information are, respectively,

$$\boldsymbol{s}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \boldsymbol{s}_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} \frac{d\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \quad \textit{and} \quad \boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}) = -\sum_{t=1}^{T} \mathbb{E}_{t-1}[\boldsymbol{\mathcal{H}}_t(\boldsymbol{\theta})] = -\sum_{t=1}^{T} \mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top}\right].$$

Step 3: Repeat until convergence, e.g.,

$$\frac{\left\|\widehat{\boldsymbol{\theta}}_T^{(h+1)} - \widehat{\boldsymbol{\theta}}_T^{(h)}\right\|}{\left\|\widehat{\boldsymbol{\theta}}_T^{(h)}\right\|} < \delta$$

for some fixed small $\delta > 0$.

The expressions for the score might be collected in a single vector,

$$\boldsymbol{s}_t(\boldsymbol{\theta}) = \left[\begin{array}{ccccc} \boldsymbol{s}_t^{(\nu)}(\boldsymbol{\theta}) & \boldsymbol{s}_t^{(\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) & \boldsymbol{s}_t^{(\boldsymbol{\omega})}(\boldsymbol{\theta}) & \boldsymbol{s}_t^{(\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) & \boldsymbol{s}_t^{(\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) \end{array}\right]^\top,$$

yielding the recursions for the static parameters

$$\begin{aligned} \boldsymbol{s}_t^{(\nu)}(\boldsymbol{\theta}) &= \frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{N}{\nu} + \frac{\nu+N}{\nu} b_t - \ln w_t\right] \\ &\qquad + \frac{\nu+N}{\nu}\frac{1}{w_t}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\nu}\right)^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}), \\ \boldsymbol{s}_t^{(\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\left[\frac{\nu+N}{\nu}\frac{1}{w_t}(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t) - \mathrm{vec}\, \boldsymbol{I}_N\right] \\ &\qquad + \frac{\nu+N}{\nu}\frac{1}{w_t}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top}\right)^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}), \end{aligned}$$

for the unconditional mean

$$\boldsymbol{s}_t^{(\boldsymbol{\omega})}(\boldsymbol{\theta}) = \frac{\nu+N}{\nu}\frac{1}{w_t}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\boldsymbol{\omega}^\top}\right)^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}),$$

and the remaining parameters for the dynamics of the location vector

$$\begin{aligned} \boldsymbol{s}_t^{(\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) &= \frac{\nu+N}{\nu}\frac{1}{w_t}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\, \boldsymbol{\Phi})^\top}\right)^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}), \\ \boldsymbol{s}_t^{(\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) &= \frac{\nu+N}{\nu}\frac{1}{w_t}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\, \boldsymbol{K})^\top}\right)^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}). \end{aligned}$$

Similarly, the conditional information matrix may then be represented as follows,

$$\mathcal{I}_t(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{I}_t^{(\nu)}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\nu,\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) & \underset{1\times N}{\mathbf{0}} & \mathcal{I}_t^{(\nu,\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\nu,\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) \\ \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Omega}),\nu)}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) & \underset{N^2\times N}{\mathbf{0}} & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Omega}),\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Omega}),\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) \\ \underset{N\times 1}{\mathbf{0}} & \underset{N\times N^2}{\mathbf{0}} & \mathcal{I}_t^{(\omega)}(\boldsymbol{\theta}) & \underset{N\times N^2}{\mathbf{0}} & \underset{N\times N^2}{\mathbf{0}} \\ \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}),\nu)}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}),\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) & \underset{N^2\times N}{\mathbf{0}} & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}),\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) \\ \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}),\nu)}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}),\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) & \underset{N^2\times N}{\mathbf{0}} & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}),\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) & \mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) \end{bmatrix}.$$

The four blocks of the matrix above have the following expansions: the first block is composed by

$$\begin{aligned}
\mathcal{I}_t^{(\nu)}(\boldsymbol{\theta}) &= \frac{1}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+N}{2}\right) - \frac{2N(\nu+N+4)}{\nu(\nu+N)(\nu+N+2)}\right] \\
&\qquad + \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right), \\
\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Omega}),\nu)}(\boldsymbol{\theta}) &= -\frac{1}{(\nu+N)(\nu+N+2)}\boldsymbol{\mathcal{D}}_N^{\top}(\mathrm{vech}(\boldsymbol{\Omega}^{-1})) \\
&\qquad + \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^{\top}}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right), \\
\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) &= \frac{\nu+N}{2(\nu+N+2)}\boldsymbol{\mathcal{D}}_N^{\top}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{\mathcal{D}}_N \\
&\qquad - \frac{1}{2(\nu+N+2)}\boldsymbol{\mathcal{D}}_N^{\top}(\mathrm{vech}(\boldsymbol{\Omega}^{-1}))(\mathrm{vech}(\boldsymbol{\Omega}^{-1}))^{\top}\boldsymbol{\mathcal{D}}_N \\
&\qquad + \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^{\top}}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^{\top}}\right).
\end{aligned}$$

The second,

$$\begin{aligned}
\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}),\nu)}(\boldsymbol{\theta}) &= \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^{\top}}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right), \\
\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}),\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) &= \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^{\top}}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^{\top}}\right), \\
\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}),\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) &= \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^{\top}}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^{\top}}\right), \\
\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}),\nu)}(\boldsymbol{\theta}) &= \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^{\top}}\right)^{\top}\boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right).
\end{aligned}$$

By virtue of its simmetry, the third and last block is composed by

$$\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}))}(\boldsymbol{\theta}) = \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top}\right)^\top \boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top}\right),$$

$$\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{\Phi}),\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) = \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top}\right)^\top \boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top}\right),$$

$$\mathcal{I}_t^{(\mathrm{v}(\boldsymbol{K}))}(\boldsymbol{\theta}) = \frac{\nu+N}{\nu+N+2}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top}\right)^\top \boldsymbol{\Omega}^{-1}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top}\right).$$

Notably, $\mathcal{I}_t^{(\omega,\xi)}(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathcal{I}_t^{(\omega,\psi)}(\boldsymbol{\theta}) = \mathbf{0}$, which means that $\boldsymbol{\omega}$ is asymptotically independent of the other parameters. Moreover, none of the terms of the conditional information matrix involves the second derivatives of the dynamic location. This result is a direct consequence of the asymptotic properties of the proposed ML estimator under the assumption of correct specification of the model and some regularity conditions.

To start the estimation procedure one can follows the approach suggested in Fiorentini, Sentana, and Calzolari, 2003. First, a consistent estimator of the restricted version of the parameter vector $\tilde{\boldsymbol{\theta}}_T$ is obtained by the Gaussian quasi-maximum likelihood procedure in Bollerslev and Wooldridge, 1992. Second, a consistent method of moment is adopted for the degrees of freedom ν, by making use of the empirical coefficient of excess kurtosis $\tilde{\kappa}$ on the standardized residuals and the relationship $\tilde{\nu} = (4\tilde{\kappa}+6)/\tilde{\kappa}$. Convergence is fast in that usually few iterations of that procedure are needed, which makes scoring methods particularly appealing for estimation purposes.

2.5 Simulation Study

The finite-sample properties of the ML estimator are investigated via Monte-Carlo simulations. We focus on: (a) the distribution of the ML estimator of the matrix $\boldsymbol{\Phi}_0$, which contains the autoregressive coefficients; (b) the matrix $\boldsymbol{K}_0$; and (c) the degrees of freedom ν. It well-known that estimating the degrees of freedom in multivariate Student's t distributions can be quite challenging, since the implied profile likelihood is remarkably flat (Breusch, Robertson, and Welsh, 1997). For this reason, we assume that the distribution of the heavy-tailed errors will be $\boldsymbol{\epsilon}_t \sim \boldsymbol{t}_{\nu_0}(\mathbf{0}_2, \boldsymbol{I}_2)$, where $\nu_0 \in \{3, 5, 10\}$, that is, a standard bivariate Student's t with three, five and ten degrees of freedom. With these three different specifications, we cover we a wide range of time series. However, it is important to note that the variables share the same degrees of freedom ν_0. In the next subsection, we illustrate the experiment for the bivariate case.

In practice, we simulate data from the different specifications of the standard bivariate Student's t and for each of the realized paths of the time series we consider the recursion (2.3) for $\boldsymbol{\mu}_{t|t-1}$, which satisfies the stationarity conditions of Lemma (2.3.1). During the process which generates the data, we use a burn-in period of $1,000$ replications and we store $T = 250$, 500 and $1,000$ observations. This ensures that the collected $\{\boldsymbol{y}_t\}$ are stationary ergodic.

With this simulated data at hand, we start the Fisher's scoring algorithm based on the analytical formulas described in Section 2.4.2. We stop the whole estimation process after a maximum of ten iterations in order to assess the rate of convergence and evaluate the precision of scoring rule. We repeat this simulation scheme $M = 1,000$ times for each case and we use the empirical measures of bias and root mean square error to quantify the accuracy of our proposed estimators.

Formally, we define the empirical bias measure for the estimated kurtosis parameter $\hat{\nu}$ and the $1,000$ replications as

$$Bias(\hat{\nu}) = \frac{1}{1000}\sum_{m=1}^{1000}(\hat{\nu}_m - \nu_0),$$

and the empirical root mean square error of $\hat{\nu}$ as

$$RMSE(\hat{\nu}) = \sqrt{\frac{1}{1000}\sum_{m=1}^{1000}(\hat{\nu}_m - \nu_0)^2}.$$

In general, given the considered DGPs, we expect the distributions of the estimators to be well approximated by a Gaussian distribution with low values of biases and root mean square error.

2.5.1 Bivariate Case

In the bivariate case, the vector of parameters assumes the following form

$$\boldsymbol{\theta} = (\nu, \Omega_{11}, \Omega_{21}, \Omega_{22}, \omega_1, \omega_2, \Phi_{11}, \Phi_{21}, \Phi_{12}, \Phi_{22}, \kappa_{11}, \kappa_{21}, \kappa_{12}, \kappa_{22})^\top,$$

thus $\boldsymbol{\theta} \in \mathbb{R}^{14}$, which means that we need to estimate 14 parameters, in order to obtain a complete bivariate system.

The true parameters of the considered DGP are

$$\nu_0 \in \{3,5,10\},\ \boldsymbol{\Omega}_0 = \begin{bmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \end{bmatrix},\ \boldsymbol{\omega}_0 = \begin{bmatrix} -3 & 5 \end{bmatrix},$$

$$\boldsymbol{\Phi}_0 = \begin{bmatrix} 0.85 & 0.00 \\ 0.00 & 0.80 \end{bmatrix},\ \boldsymbol{K}_0 = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.90 \end{bmatrix}.$$

Thee results are reported in tables 2.1, 2.2 and 2.3 for the three different considered values of the degrees of freedom $\nu_0 \in \{3,5,10\}$, respectively. Each table contains three columns which are associated with the time series dimensions, that is $T = 250$, 500 and $1,000$.

The first result that one could notice from the simulation results detailed in tables 2.1, 2.2 and 2.3 is that as the time series dimension increases, the values of the empirical *Bias* and *RMSE* tend to reduce sharply, which is line with the consistency Theorem 2.4.1. In particular, we note that even if the value of ν_0 are very low, such as $\nu_0 = 3$, the results are still satisfactory, which shows that our model and estimation procedure are robust against heavy-tailed data and potential outliers.

In general, estimation of the number of degrees of freedom is rather accurate. Moreover, the decreasing bias and *RMSE* patterns may be due to the fixed initial value of the dynamic location vector $\boldsymbol{\mu}_{1|0}$ which was used to start the filter recursions. However, the invertibility conditions introduced in Lemma 2.3.3 ensure that for $T \to \infty$, this initial estimation bias will eventually tapers off.

In conclusion, the *ML* estimations deliver satisfactory results in terms of bias and root mean square error, hence the reliability of the Fisher-scoring method.

TABLE 2.1: Monte-Carlo Simulation results for $\nu_0 = 10$.

	$T = 250$			$T = 500$			$T = 1000$		
	Estimate	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*
ν	10.529	-0.529	4.727	10.383	-0.384	2.232	10.226	-0.226	1.631
Ω_{11}	0.989	0.011	0.097	0.995	0.005	0.075	0.995	0.004	0.057
Ω_{12}	-0.001	0.002	0.067	0.000	0.000	0.045	-0.000	0.002	0.035
Ω_{22}	0.991	0.008	0.108	0.991	0.009	0.074	0.993	0.006	0.057
ω_1	-3.014	-0.015	0.365	-2.994	-0.006	0.234	-2.996	-0.004	0.189
ω_2	5.013	-0.013	0.287	4.994	0.006	0.204	4.997	0.002	0.129
Φ_{11}	0.834	0.016	0.052	0.838	0.011	0.032	0.845	0.005	0.023
Φ_{12}	-0.005	0.006	0.064	0.002	-0.003	0.040	-0.002	0.002	0.027
Φ_{21}	0.002	-0.002	0.047	0.001	-0.001	0.031	-0.000	0.000	0.024
Φ_{22}	0.781	0.019	0.063	0.789	0.011	0.040	0.794	0.006	0.028
κ_{11}	0.926	0.024	0.113	0.946	0.003	0.089	0.949	0.001	0.065
κ_{12}	0.059	-0.009	0.083	0.051	-0.001	0.065	0.049	0.001	0.050
κ_{21}	0.042	0.007	0.091	0.048	0.002	0.064	0.049	0.000	0.050
κ_{22}	0.877	0.023	0.123	0.896	0.004	0.083	0.893	0.007	0.061

Monte-Carlo Simulation results based on 1000 replications. The columns "*Bias*" and "*RMSE*" represent the empirical bias and root mean square error.

TABLE 2.2: Monte-Carlo Simulation results for $\nu_0 = 5$.

	$T = 250$			$T = 500$			$T = 1000$		
	Estimate	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*
ν	5.089	-0.090	1.084	5.075	-0.075	0.693	5.012	-0.012	0.573
Ω_{11}	0.978	0.220	0.121	0.993	0.007	0.086	0.997	0.003	0.068
Ω_{12}	0.000	-0.001	0.075	-0.002	0.002	0.050	-0.001	0.001	0.046
Ω_{22}	0.974	0.025	0.123	0.988	0.012	0.084	0.992	0.008	0.038
ω_1	-2.973	-0.027	0.326	-2.994	-0.006	0.219	-3.002	0.002	0.127
ω_2	5.011	-0.011	0.268	4.995	0.005	0.156	4.997	0.003	0.133
Φ_{11}	0.831	0.019	0.055	0.831	0.019	0.055	0.844	0.006	0.055
Φ_{12}	-0.000	0.001	0.068	-0.000	0.001	0.068	0.000	0.000	0.044
Φ_{21}	-0.000	0.001	0.056	-0.001	0.001	0.056	-0.001	0.001	0.024
Φ_{22}	0.776	0.023	0.069	0.777	0.023	0.069	0.788	0.012	0.039
κ_{11}	0.974	0.002	0.154	0.949	0.001	0.103	0.950	-0.001	0.083
κ_{12}	0.047	0.003	0.115	0.050	0.000	0.075	0.050	0.000	0.027
κ_{21}	0.055	-0.005	0.112	0.055	-0.005	0.073	0.052	-0.002	0.055
κ_{22}	0.896	0.004	0.138	0.900	-0.001	0.099	0.900	-0.000	0.049

TABLE 2.3: Monte-Carlo Simulation results for $\nu_0 = 3$.

	$T = 250$			$T = 500$			$T = 1000$		
	Estimate	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*
ν	2.987	0.013	0.457	2.979	0.021	0.473	3.016	-0.016	0.321
Ω_{11}	0.972	0.028	0.132	0.972	0.028	0.130	0.988	0.011	0.093
Ω_{12}	-0.000	0.000	0.073	-0.004	0.004	0.074	0.000	0.000	0.053
Ω_{22}	0.971	0.029	0.135	0.972	0.028	0.136	0.991	0.008	0.090
ω_1	-2.996	-0.004	0.288	-2.990	-0.009	0.257	-3.008	0.009	0.183
ω_2	5.002	-0.003	0.215	5.004	-0.004	0.207	5.002	-0.002	0.145
Φ_{11}	0.831	0.019	0.063	0.836	0.014	0.062	0.840	0.010	0.040
Φ_{12}	0.001	-0.001	0.082	0.000	0.000	0.083	-0.000	0.001	0.047
Φ_{21}	0.000	0.000	0.069	-0.001	0.001	0.066	-0.000	0.000	0.041
Φ_{22}	0.768	0.032	0.091	0.771	0.029	0.084	0.789	0.011	0.048
κ_{11}	0.955	-0.005	0.184	0.941	0.009	0.117	0.954	-0.004	0.086
κ_{12}	0.052	-0.002	0.142	0.054	-0.004	0.145	0.049	0.000	0.097
κ_{21}	0.054	-0.004	0.149	0.057	-0.007	0.152	0.051	-0.001	0.098
κ_{22}	0.898	0.002	0.182	0.894	0.006	0.186	0.899	0.001	0.121

2.6 Empirical Analysis of Homescan Data Consumer Prices

In order to demonstrate a potential use of the score-driven method, we show an innovative application to the estimation of consumer prices from homescan data. This field of application is gaining interest, due to the growing availability of high frequency and high detail purchase data collected through scanner technologies at the retail point (retail scan) or household level (homescan). The latter of type of data allows one to obtain cost-of-living measures for vulnerable sub-groups of the population, and to explore the distributional effects of fiscal measures. While being a valuable source for detailed price information, post-purchase homescan price data are affected by a measurement noise that can be potentially large in small samples, and the application of filtering techniques may help to mitigate such noise and control for outliers.

Scanner data are collected either at the retail level, e.g. supermarket data, or from households in consumer panels, i.e. homescan data. Retail scanner data are widely used to estimate prices, both for continuity with the traditional price survey methodology, and because they are expected to suffer less from the substitution (unit value) bias (Silver and Heravi, 2001). This bias is due to the fact that scanner data are based on actual transactions, i.e. prices are only observed after the consumer purchases the good. This implies that the observed price embodies a quality choice component, as consumers confronted with a price increase may opt for a cheaper option (or a cheaper retailer) and information on non-purchased items is missing. The bias can be particularly important for aggregated goods, such as those goods commonly represented by category-level prices like food and drinks. Thus, a wide body of research has been devoted to improve sampling strategies and the choice of weights in aggregation. A well-documented problem is the change in the composition of the consumption basket over time, an issue that can be exacerbated by high-frequency data (Feenstra and Shapiro, 2003a). For example, stockpiling of goods during promotion periods generate bias in price indices, as the purchased quantities are not independent over subsequent time periods (Ivancic, Diewert, and Fox, 2011; Melser, 2018).

Although supermarket-level scanner data allow to mitigate the problem, as one expects a wide range of products to be purchased across the population of customers within a given time period, the use of homescan data to estimate prices and price indices has potentially major advantages. These advantages lie in the possibility to exploit household-level heterogeneity. Most importantly, it becomes feasible to estimate prices faced by particular population sub-groups whose consumption basket differs from the average one, as elderly households or low-income groups (Kaplan and Schulhofer-Wohl, 2017; Broda, Leibtag, and Weinstein, 2009). However, the unit value issue is heavier with homescan data, as individual households buy a small range of products. Thus, variable shopping frequencies and zero purchases make it necessary to rely on very large samples of households to control the bias. The problem becomes even more conspicuous for prices at the regional level, for products that are not frequently purchased and for products whose demand is highly seasonal.

Robust filtering techniques may constitute a powerful solution to the above mentioned problems, and may perform well even with relatively small samples of household as the one used in our application.

To illustrate the potential contribution of the proposed score-driven method, we exploit a data-set that has been recently used to evaluate the effects of a tax on sugar-sweetened beveraged introduced in France in 2012 (Capacci et al., 2019). Our data consists of weekly scanner price data for food and non-alcoholic drinks. The data were collected in a single region, within the Italian GfK homescan consumer panel, based on purchase information on 318 households surveyed in the Piedmont region, over the period between January 2011 and December 2012. The regional scope and the relatively small sample provide an ideal setting to test the applicability and effectiveness of the multivariate filtering approach.

2.6.1 Data

The data for our application consist of three time series of weekly unit values for food items, non-alcoholic drinks and Coca-Cola purchased by a sample of 318 households residing in the Piedmont region, Italy, over the period 2011-2012, and collected within the GfK Europanel homescan survey. The data-set provides information on weekly expenditures and purchased quantities for each of the three aggregated items, and unit values are obtained as expenditure-quantity ratios.

Average unit values are shown in Table 2.4. Food and non-alcoholic drinks are composite aggregates, hence they are potentially subject to fluctuations in response to changes in the consumer basket even when prices are stable. Instead, Coca-Cola is a relatively homogeneous good, with little variability due to different packaging sizes.

	2011		2012	
Food	4.343	(0.234)	4.226	(0.255)
Non-alcoholic drinks	0.434	(0.047)	0.426	(0.052)
Coca-Cola	1.000	(0.096)	1.100	(0.172)

TABLE 2.4: Average unit values, € per kilogram, Piedmont homescan data (standard deviations in brackets)

2.6.2 Results

We fit the multivariate score driven model developed in the paper to the considered vector of time series. ML estimation produces the following multivariate dynamic system of time varying locations for Drinks (D), Food (F) and Coca-Cola (C),

$$\begin{bmatrix} \underset{(0.000)}{\mu^D_{1,t+1|t} - 0.443} \\ \underset{(0.000)}{\mu^F_{2,t+1|t} - 4.394} \\ \underset{(0.000)}{\mu^C_{3,t+1|t} - 1.070} \end{bmatrix} = \begin{bmatrix} \underset{(0.011)}{0.839} & \underset{(0.002)}{0.015} & \underset{(0.005)}{0.007} \\ \underset{(0.059)}{-0.528} & \underset{(0.009)}{0.912} & \underset{(0.025)}{0.342} \\ \underset{(0.020)}{0.222} & \underset{(0.003)}{0.023} & \underset{(0.009)}{0.847} \end{bmatrix} \begin{bmatrix} \underset{(0.000)}{\mu^D_{1,t|t-1} - 0.443} \\ \underset{(0.000)}{\mu^F_{2,t|t-1} - 4.394} \\ \underset{(0.000)}{\mu^C_{3,t|t-1} - 1.070} \end{bmatrix} + \begin{bmatrix} \underset{(0.017)}{0.442} & \underset{(0.003)}{-0.023} & \underset{(0.007)}{0.007} \\ \underset{(0.079)}{0.334} & \underset{(0.014)}{0.216} & \underset{(0.038)}{-0.631} \\ \underset{(0.030)}{-0.290} & \underset{(0.005)}{-0.098} & \underset{(0.014)}{-0.014} \end{bmatrix} \begin{bmatrix} u^D_{1t} \\ u^F_{2t} \\ u^C_{3t} \end{bmatrix}$$

where the values in parenthesis are the standard errors and with

$$\hat{\nu} = 6.921 \ \ (0.229), \qquad \hat{\boldsymbol{\Omega}} = \begin{bmatrix} \underset{(0.138)}{0.162} & \cdot & \cdot \\ \underset{(0.913)}{0.348} & \underset{(0.327)}{53.258} & \cdot \\ \underset{(0.057)}{-0.134} & \underset{(0.327)}{-0.579} & \underset{(0.155)}{9.086} \end{bmatrix} \times 10^{-3}.$$

The estimated degrees of freedom are approximately 7. We remark here that the assumption of a (conditional) multivariate Student's t distribution for the data generating process implies that all the univariate marginal distributions are tail equivalent, see Resnick, 2004. This requires the implicit underlying assumption that the level of heavy-tailedness across the observed time series vector is fairly homogeneous. To investigate this issue, and for the sake of comparisons, we have carried out a univariate analysis, as in Harvey and Luati, 2014, from which it resulted that the estimated degrees of freedom were very low for Coca Cola (about 4) and medium size (smaller than 30) for the other two series, as expected. Hence, the multivariate score driven model developed in the paper reveals to be a good compromise between a multivariate non-robust filter, based on a linear Gaussian model, and a robust univariate estimator. Indeed, a multivariate Portmanteau test on the residuals obtained from the three univariate models is carried out to test the null hypothesis $H_0 : \boldsymbol{R}_1 = \cdots = \boldsymbol{R}_m = \boldsymbol{0}$, where $\boldsymbol{R}_i$ is the sample cross-correlation matrix for some $i \in \{1, \ldots, m\}$ against the alternative $H_1 : \boldsymbol{R}_i \neq \boldsymbol{0}$.

m	$Q(m)$	*df*	*p-value*
1	13.7	9	0.000
2	40.8	18	0.000
3	58.6	27	0.000
4	89.6	36	0.000
5	105.9	45	0.000

TABLE 2.5: *Multivariate Portmanteau test.*

The results of Table 2.5 indicate rejection of the null hypothesis of absence of of serial dependence in the trivariate series at the 5% significance level.

The matrix of the estimated autoregressive coefficients $\hat{\boldsymbol{\Phi}}$ measures the dependence across the dynamic locations $\boldsymbol{\mu}_{t|t-1}$, while the estimated scale matrix $\hat{\boldsymbol{\Omega}}$ measures the concurrent relationship between the three series under investigation, i.e. drink, food and Coca Cola prices. For these matrices, we report the estimates of the coefficients and, in parenthesis, the relative standard errors. The diagonal elements of $\hat{\boldsymbol{\Phi}}$ show that each variable of interest is highly persistent. In order to explore the relation among the series, we implement an impulse response analysis. Figure 2.1 shows the estimated impulse response functions.

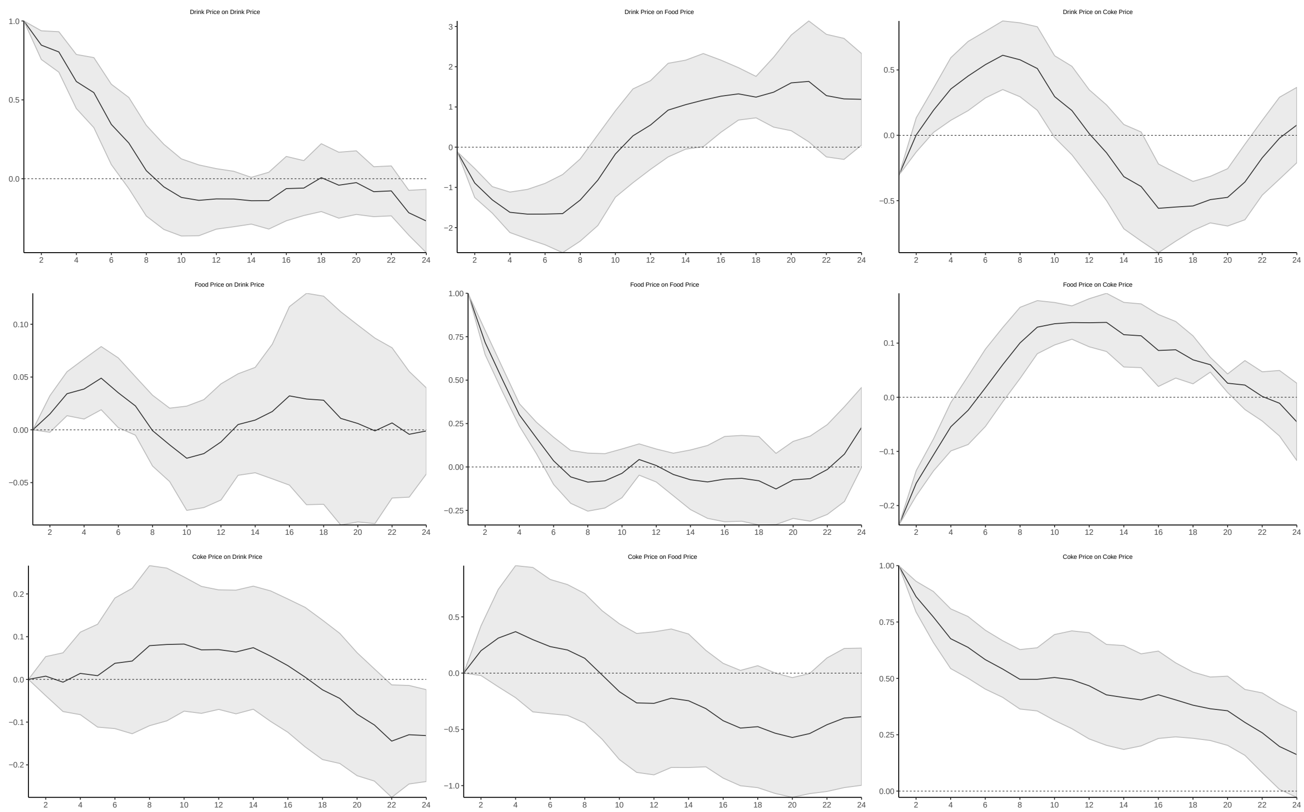


FIGURE 2.1: Estimated impulse response functions of the filtered $\boldsymbol{\mu}_{t|t-1}$ for a unit shock.

What emerges is a negative relation between drink and food prices: a unit shock in drink prices will produce a negative shock in food prices. This may adjustments in purchasing decisions by the households aimed at mitigating the rising cost of their shopping basket. This would be evidence that univariate signals are likely to suffer from the unit value bias. Similarly, a non trivial negative relation exists between food and Coca Cola prices. A unit shock on food prices yields a concurrent negative impact on Coca Cola prices, which is also noted from the analysis of the cross-correlations. As one might expect, a positive correlation exists between Coca Cola prices and drink prices, as the former product belongs to the latter category. Instead, unit shocks on food prices seem to have negligible correlation (if any) on drink prices.

2.6.3 Interpretation

Figure 2.2 shows the original unit value time series and the corresponding signals extracted through the multivariate score driven filter. Noise and outliers, as well as some irregular periodic pattern, are clearly visible in the drinks and food series. On the other hand, the Coca-Cola series is relatively regular, with the exception of few peaks, including a couple of large outliers in the second year. Given the homogeneous nature of the good, it is reasonable to believe that those extreme values are the results of measurement error.

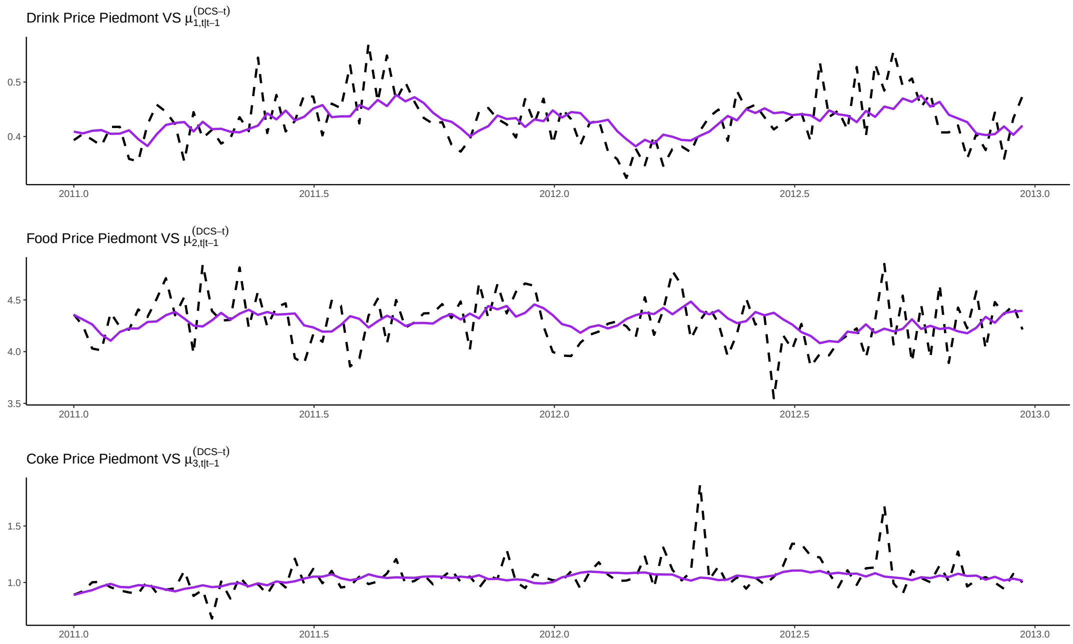


FIGURE 2.2: Original series (dotted line) and estimated signals

The estimates illustrate an effective noise reduction and return patterns that are smoother and more consistent with a regular price time series. As one would expect, the Coca-Cola DCS-t series is very flat, and suggests a relatively stable price over the two-years time window, with no outliers.

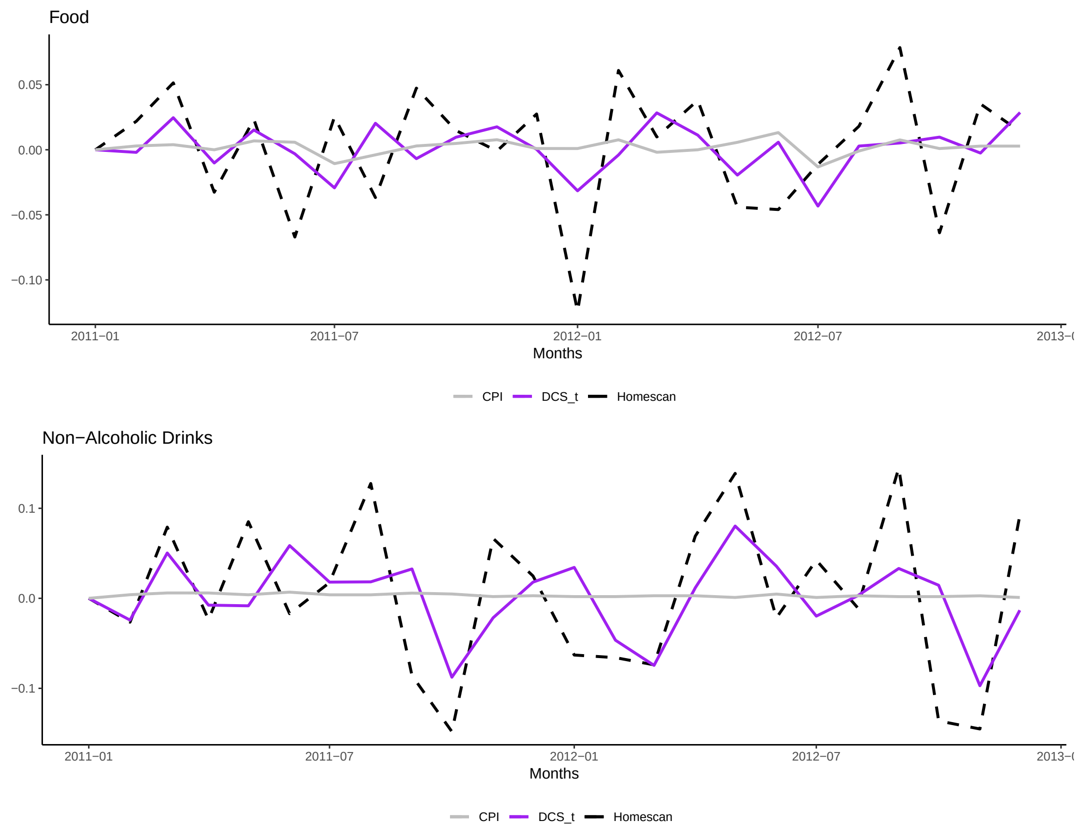


FIGURE 2.3: Raw unit value series (dotted line), estimated signal and Regional CPIs (log differences, grey line)

Figure 2.3 shows the monthly natural logarithm differences of the raw homescan prices (HSP) and the estimated signals, together with changes in the official Regional CPIs (R-CPI) for food and non-alcoholic drinks, whereas no CPI to the brand detail is produced. The R-CPIs are provided by the National Statistical Institute (ISTAT). They have a monthly frequency and are built with a traditional survey-based approach on retailers. The comparison between the score driven filtered values and the R-CPIs is purely indicative, as the unit values from the homescan data are weekly, whereas the official CPIs are monthly. This frequency difference may lead to biased comparisons (Diewert, Fox, and Haan, 2016). Nevertheless, the graphs confirm that the score driven signals are effective in reducing the noise in the data. This is especially true for the food series, whose CPIs are more volatile compared to drinks. The correlation between the raw homescan log-differenced unit value and the log-differenced food CPI is 0.05, against 0.44 when the filtered time series is considered. For the non-alcoholic drinks price series the gain is less conspicuous, as prices evolve very regularly over the time window. Still, an inexistent correlation between the HSP and the R-CPI (-0.02) turns into a positive one (+0.11) when considering the score driven estimates and the R-CPI.

In essence, the empirical evidence suggests that a robust multivariate approach to model-based signal extraction produce meaningful price series from homescan data, especially when noise and outliers in the original data are relevant. We find the approach to perform reasonably well even with a low number of sampled households (318) and price time series (3), and with a relatively short time window (104 weeks). Future research might shed further light on the implications of dealing with a larger number of price series and longer time series.

2.7 Concluding Remarks

We presented a nonlinear and multivariate dynamic location model which enables the extraction of reliable signals from vector processes affected by outliers, structural problems and non-Gaussian errors. The model has two innovative features: (a) it extends the univariate first-order dynamic conditional location score by **Harvey_Luati2014** to the multivariate setting; and (b) it extends the dynamic model for time varying volatilities and correlations by Creal, Koopman, and Lucas, 2011 to the location case.

Its peculiarity lies in the specification of a score-robust updating equation for the time-varying conditional location vector. We derived the stochastic properties of the model: bounded moments, stationarity, ergodicity, and filter invertibility. Parameters are estimated by maximum likelihood and we provided closed formulae for the score vector and the Hessian matrix, which can be directly used for a scoring procedure. Consistency and asymptotic normality have been proved and a Monte Carlo study showed good and reliable finite sample properties.

Our empirical application showed that filtering may lead to satisfactory estimates of price signals from homescan data. We contribute to research in this area with two promising results. First, we show that robust modelling allowing for heavy tails provide is more effective in dealing with noisy series affected by outliers. Second, the multivariate extension of the DCS-t model has shown more appropriate in the case of scanner price data, as price time series are expected to have a good degree of correlation. This proves to be valuable information to reduce the noise across the modelled price time series.

Chapter 3

Dynamic Factor Models with Robust Idiosyncratic Components

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name
Department of Statistical Sciences “Paolo Fortunati”

Doctor of Philosophy in Statistical Sciences

Essays in Robust and Nonlinear Multivariate Time Series Models

by Enzo D’INNOCENZO

Factor volatility models which decompose the conditional volatility of a N-variate time-series into a common factor and N idiosyncratic components can offer a useful and parsimonious way of handling multivariate time-series and panels of data. While the common factor is naturally robust to fat-tailed innovations due to averaging of information over the N cross-sectional elements, the idiosyncratic components are typically sensitive to outliers. We propose a nonlinear dynamic factor model for conditional volatilities with score-robust updating equation for the idiosyncratic components with substantially improved fit on data sets with fat tailed innovations. We derive stochastic properties for the model, including bounded moments, stationarity, ergodicity, and filter invertibility. We also establish consistency and asymptotic normality of the maximum likelihood estimator in large samples. Additionally, we study the small sample properties of the estimator by means of a Monte Carlo study. Finally, we provide an empirical illustration using a panel of ten stocks from the S&P500 which highlights the advantages of the proposed dynamic factor structure, as well as the need for robust filtering techniques for the idiosyncratic component.

3.1 Introduction

Multivariate models play an important role in the financial econometrics literature. Univariate time series models are often too restrictive given the presence of cross-sectional dependences and common features in the data. The seminal work of Engle and Kroner (1995) provided a multivariate specification of the already widely used ARCH model introduced by Engle, 1982 and further generalized by Bollerslev, 1986. Since then, several multivariate model specifications have been introduced which enjoy raising popularity due to their ability to capture covariance spillovers and feedbacks, which are of fundamental importance in many economic or financial applications; see e.g. Bauwens, Laurent, and Rombouts, 2006 for a review.

Unfortunately, despite the advance of computing power and efficient optimization methods, most multivariate GARCH model specification still suffer from the curse of dimensionality, i.e. the number of parameters to be estimated becomes unfeasible and the dimension of the dynamic covariance matrix becomes intractable, as the number of series to be analyzed increases. The factor GARCH model, proposed by Engle, Ng, and Rothschild, 1990, is a popular method to reduce the dimension of the parameter; see also Alexander, 2001 and Weide, 2002, with the orthogonal GARCH (O-GARCH) and its generalized version (GO-GARCH) respectively.

Factor volatility models which decompose the conditional volatility of a N-variate time-series into a common factor and N idiosyncratic components can offer a useful and parsimonious way of handling multivariate time-series and panels of data. While the common factor is naturally robust to fat-tailed innovations due to averaging of information over the N cross-sectional elements, the idiosyncratic components are typically sensitive to outliers. In this paper, we propose a dynamic non-linear conditional volatility model with a robust updating equation for the idiosyncratic conditional volatilities.

Our model is naturally related to the class of single-factor multivariate time series models. The idea dates back to the early works of Engle and Watson, 1981a and Gonzalo and Granger, 1995. Adopting the terminology of **cox**, in both the aforementioned works, the authors propose parameter-driven models which retain a linear and Gaussian specification. It is well-known that in this case the Kalman recursions provides the optimal solutions and the likelihood function can be retrieved from the prediction-error decomposition. However, in financial and economic applications, the normality assumptions may be too restrictive. Once the assumptions are dropped, the Kalman filter will lose its optimality, therefore, closed-formulae are not available and simulation-based techniques are required. As a consequence, statistical properties of the model may remain obscure to the researcher. In direct contrast, we adopt an observation-driven specification. One important advantage of this class of models, is that the predictive likelihood is usually available in closed-form facilitating parameter estimations via standard likelihood-based methods. Even if we drop Gaussianity assumptions and introduce nonlinearities in the model. In particular, relying in a simple system of non-linear recursions, our model decomposes the conditional variance of each series of daily returns into the product of a common and idiosyncratic component.

Our model can be also seen as an extension of the volatility component models introduced by Engle and Lee, 1999 and further explored by Engle and Rangel, 2008, Amado and Teräsvirta, 2013 and the multivariate extension of Hafner and Linton, 2010 combined with the aforementioned one-factor multivariate time series models literature. Even thought they haven proven to be a powerful tool capable to capture complex volatility dynamics, a comprehensive statistical and probabilistic analysis of the multiplicative component models was not available until the recent work of Wang and Ghysels, 2015.

In a different framework, the paper of Barigozzi et al., 2014 propose a similar model. Relying on a semi-nonparametric setup, the authors formulate and estimate a vector multiplicative error model where the common trend component is estimated by a Nadaraya-Watson type estimator, while the idiosyncratic terms are individually modeled as scalar parametric and asymmetric GARCH processes. In contrast, we have decided to engage a fully parametric framework by specifying a simple univariate GARCH-type filter for the common factor. Additionally, we achieve robustness for the idiosyncratic components, by drawing on the score-driven class of models proposed by Creal, Koopman, and Lucas, 2013 and Harvey, 2013, introduced in the next section.

In the remainder of this paper, Section 2 introduces the score factor model. Section 3 studies its stochastic properties. Section 4 establishes the asymptotic properties of the maximum likelihood estimator. Section 5 analyzes small sample properties through a Monte Carlo study. Section 6 provide an empirical illustration using a panel of 10 stocks from the S&P500, and Section 7 concludes.

3.2 The Model

Let $\boldsymbol{x}_t = (x_{1t}, \dots, x_{Nt})^\top \in \mathbb{R}^N$ denote an N-dimensional vector of stock returns at time t and let $\mathscr{F}_{t-1} = \sigma\{\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \dots\}$ be the σ-field generated by the past values of $\boldsymbol{x}_t$. For each asset $i = 1, \dots, N$, the return at day t is modeled as

$$x_{it} = \sqrt{h_t}\epsilon_{it}, \tag{3.1}$$

where $\{\epsilon_{it}\}$ is an independent and identically distributed (*IID*) sequence. We model the time-varying conditional volatility $\{h_t\}$ as the product of a dynamic common factor $\{f_t\}$, which captures the commonalities between assets, and the idiosyncratic volatility processes of the considered asset $\{\sigma_{it}\}$,

$$x_{it} = f_t \sigma_{it} \epsilon_{it}. \tag{3.2}$$

Our score-driven model captures the dynamics of time-varying parameters, $\boldsymbol{\sigma}_t = (\sigma_{1t}, \dots, \sigma_{Nt})^\top$, through an autoregressive term and the scaled score of the conditional density of the sequence of observations. Hence, for $t = 1, \dots, T$ and for each i-th element of this vector, with $i = 1, \dots, N$, we have the updating recursion

$$\sigma^2_{i,t+1} = \delta_i + \phi_i \sigma^2_{it} + \kappa_i s_{it},$$

where δ_i, ϕ_i and κ_i are unknown parameters to be estimated and s_{it} is the conditional scaled score, the driving-force of the process, defined by

$$s_{it} = S_{it} \nabla_{it}, \qquad \nabla_{it} = \frac{\partial \log p(\epsilon_{it} | \sigma^2_{it}, \delta_i, \phi_i, \kappa_i)}{\partial \sigma^2_{it}},$$

with ∇_{it} being the conditional score and S_{it} is the scaling factor. Popular choices for this component are

- unit scaling, that is $S_{it} = 1$,
- the square root of inverse of the Fisher information scaling, that is $S_{it} = \mathcal{I}(\sigma_{it})^{-1/2}$ or
- the inverse of the Fisher information scaling, $S_{it} = \mathcal{I}(\sigma_{it})^{-1}$,

where

$$\mathcal{I}(\sigma^2_{it}) = -\mathbb{E}\left[\frac{\partial^2 \log p(\epsilon_{it} | \sigma^2_{it}, \delta_i, \phi_i, \kappa_i)}{(\partial \sigma^2_{it})^2}\right].$$

For a more detailed discussion we refer to Creal, Koopman, and Lucas, 2013 and Harvey, 2013.

If we further assume that the *IID* sequence follows a standard Student's t distribution with ν_i degrees of freedom, zero mean and unit scale, that is

$$\epsilon_{it}|\mathscr{F}_{t-1} \sim t_{\nu_i}(0,1),$$

we can complete the specifications of the *i-th* idiosyncratic sequence $\{\sigma_{it}\}$, which is a positive conditionally predictable that evolves as a first-order *Beta-t-GARCH*, model of Harvey, 2013. The basic recursion is given by

$$\sigma^2_{i,t+1} = \delta_i + \phi_i\sigma^2_{it} + \kappa_i\sigma^2_{it}\left(\frac{(\nu_i+1)(x^2_{it}/f^2_t)}{(\nu_i-2)\sigma^2_{it} + (x^2_{it}/f^2_t)} - 1\right). \tag{3.3}$$

Assumption 2 ensures positivity of the conditional volatility Additionally, we fix the intercepts $\delta_i = (1-\phi_i)$ to ensure that the unconditional mean of the idiosyncratic conditional volatility in (3.3) is one. Clearly, the requirement $\min_{i=1,\dots,N}\nu_i > 2$ is a necessary condition that comes from the distributional assumptions.

Assumption 2. *For* $i = 1,\dots,N$, $\delta_i = (1-\phi_i) > 0$, $\kappa_i \geq 0$, $\phi_i \geq 0$ *and* $\phi_i - \kappa_i \geq 0$.

The updating rule for the common factor is given by a *GARCH(1,1)-type* process

$$f^2_{t+1} = \omega + \alpha\left(\frac{1}{N}\sum_{i=1}^{N} x^2_{it} - f^2_t\right) + \beta f^2_t, \tag{3.4}$$

Note that the updating equation in (3.4) is driven by the prediction error between the estimated common conditional volatility $\frac{1}{N}\sum_{i=1}^{N} x^2_{it}$ and the filtered common factor f^2_t. Assumption 2 imposes positivity constraints which ensure the positivity of the common conditional volatility f^2_{t+1}.

Assumption 3. $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\beta - \alpha \geq 0$.

The defined GARCH form of equation (3.4) needs some justification. We keep the common factor equation of the GARCH form because in our empirical evidences, it can be seen that averaging over the cross-section will make it robust. The common factor can have a linear update because it averages over the cross-section. The idiosyncratic component needs the nonlinear robust update because all the outliers end up in the idiosyncratic term. Therefore, a score-driven model like the adopted *Beta-t-GARCH*, turns out to be a reasonable choice. In support of this claim for the dynamic common factor in (3.4), we provide some empirical evidence in our empirical application, see 3.6 below.

We can thus summarize our model using the following set of equations,

$$\begin{aligned}
x_{it} &= f_t\sigma_{it}\epsilon_{it},\\
f^2_{t+1} &= \omega + \alpha\frac{1}{N}\sum_{i=1}^{N} x^2_{it} + (\beta-\alpha)f^2_t,\\
\sigma^2_{i,t+1} &= \delta_i + \left[(\phi_i-\kappa_i) + \kappa_i\frac{(\nu_i+1)(x^2_{it}/f^2_t)}{(\nu_i-2)\sigma^2_{it} + (x^2_{it}/f^2_t)}\right]\sigma^2_{it}.
\end{aligned} \tag{3.5}$$

As a conclusion of this section we introduce some useful notation. We define the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\lambda}^\top, \boldsymbol{\psi}_1^\top, \dots, \boldsymbol{\psi}_N^\top)^\top$, where $\boldsymbol{\lambda} = (\omega,\alpha,\beta)^\top \in \boldsymbol{\Lambda} \subset \mathbb{R}^3$ collects the parameters of the common factor process, while each $\boldsymbol{\psi}_i = (\delta_i,\phi_i,\kappa_i,\nu_i)^\top \subset \mathbb{R}^4$ for $i = 1,\dots,N$ and such that $\boldsymbol{\psi}_i \cap \boldsymbol{\psi}_i = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{N}\boldsymbol{\psi}_i = \boldsymbol{\Psi} \subset \mathbb{R}^{4N}$, those who drive the dynamics of the idiosyncratic processes. For any scalar random variable we define the norm $\|x\|_n = (\mathbb{E}[|x|^n])^{1/n}$.

3.2.1 Comparison with the Multivariate GAS of Creal, Koopman, and Lucas, 2011

In this subsection, we show the main differences between the proposed dynamic single-factor score-driven model with the model of (Creal, Koopman, and Lucas, 2011). Furthermore, we compare empirically both models in section 3.6 below.

The time-varying features of a panel of stocks returns may be captured by following the procedure described in Creal, Koopman, and Lucas, 2011, Section 4.3, where, by imposing a priori structure to the common dynamic factor, it is possible to model the full dynamic variance-covariance matrix with a properly selected number of score-driven filters. The objective of this procedure is to reduce the dimension of the panel, still maintaining a reliable approximation of the time-varying multivariate volatility. For the sake of comparison, let us consider a single dynamic factor structure, which dramatically reduces the number of unknown parameters of the volatility model. In practice, by assuming that the conditional correlation matrix $\boldsymbol{R}$ of the vector $\boldsymbol{x}_t$ is constant $\forall t \in \mathbb{N}$, we have

$$\boldsymbol{x}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\epsilon}_t,$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t,$$

where $\boldsymbol{D}_t$ is the dynamic diagonal standard deviation matrix. Now, the a priori common factor structure for $\boldsymbol{D}_t$ is assumed to be of the form

$$\text{diag}(\boldsymbol{D}_t^2) = \boldsymbol{a} + \boldsymbol{B} f_t^2,$$

where f_t^2 is the dynamic common factor driven by the score of the predictive log-likelihood and the vector $\boldsymbol{a}$ and the matrix $\boldsymbol{B}$ could be fixed or treated as unknown and hence to be estimated.

It is clear that the drawback of this specification is that the number of static unknown parameters of the factor loadings in $\boldsymbol{a}$ and $\boldsymbol{B}$ increases as the dimension of the vector of time series $\boldsymbol{x}_t$ increases. Therefore, the curse of dimensionality which affects most of multivariate volatility models, it is still an issue.

In contrast to the common dynamic factor model of Creal, Koopman, and Lucas, 2011, the proposed single dynamic factor model assumes that the volatility of each of the time series in $\boldsymbol{x}_t$ follows an idiosyncratic volatility by its own, and they are linked together with a strong common factor, which inherits the common fluctuations in firm-level, thus capturing the cross-section dependence.

In conclusion, for each time series x_{it} we capture its idiosyncratic volatility with the filtered sequence $\{\sigma_{i,t}^2\}$. It is important to note that with this specification, the common dynamic factor structure does not suffers the curse of dimensionality, since the common factor is extracted with a suitable and parsimonious recursion.

3.3 Statistical Properties of the Model

3.3.1 Stationarity, Ergodicity & Moments

In this section we explore the stationarity of the data generated by model (3.2). To this end, we consider the system of equations for asset i at time t,

$$
\begin{aligned}
x_{it} &= f_t \sigma_{it} \epsilon_{it}, \\
f_{t+1}^2 &= \omega + \left[\alpha \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_{it}^2 \epsilon_{it}^2 - 1 \right) + \beta \right] f_t^2, \\
\sigma_{i,t+1}^2 &= \delta_i + \left[\kappa_i \left(\frac{(\nu_i + 1)\epsilon_{it}^2}{(\nu_i - 2) + \epsilon_{it}^2} - 1 \right) + \phi_i \right] \sigma_{it}^2.
\end{aligned} \tag{3.6}
$$

Proposition 1 establishes the strict stationarity and ergodicity of the data generated by the by our robust factor model. This result is obtained by showing that both the factor volatility $\{f_t^2\}_{t\in\mathbb{Z}}$ and the the idiosyncratic volatilities are themselves stationary and ergodic $\{\sigma_{it}^2\}_{t\in\mathbb{Z}}$ under the following parameter restrictions.

Assumption 4. *For $i = 1, \dots, N$, $|\phi_i| < 1$, and $|\beta| < 1$.*

Proposition 1. *Let assumptions 2–4 be satisfied. Then, the common factor factor process $\{f_t^2\}$ and the idiosyncratic volatility processes $\{\sigma_{it}^2\}$ admit unique stationary solutions $\{f_t^2\}_{t\in\mathbb{Z}}$ and $\{\sigma_{it}^2\}_{t\in\mathbb{Z}}$ respectively. Moreover, the multiplicative components sequence $\{(f_t\sigma_{it})^2\}_{t\in\mathbb{Z}}$ is the unique stationary and ergodic solution of the volatility process. As a corollary, it follows that data $\{x_t\}_{t\in\mathbb{Z}}$ generated by this model is also stationary and ergodic.*

Proof. See Appendix B.1. □

Now we can concentrate on the number of bounded moments of model (3.2). Proposition 2 shows that the data simulated from this model has m bounded moments, where m is a function of the moments n_f of the common factor, the moments n_σ of the idiosyncratic component, and the moments n_ϵ of the innovations.

Proposition 2. *Under assumptions 2–4, the stationary ergodic series $\{x_{it}\}_{t\in\mathbb{Z}}$ has m bounded moments, that is $\mathbb{E}[|x_{it}|^m] < \infty$ where*

$$
m = \frac{n_f n_\sigma n_\epsilon}{n_\sigma n_\epsilon + n_f n_\epsilon + n_f n_\sigma}, \tag{3.7}
$$

and n_f is the number of bounded moments of the common factor process, $n_\sigma = \min_{i=1,\dots,N} n_{\sigma_i}$ the minimum number of bounded moments between the N-idiosyncratic processes and n_ϵ of the IID *sequence. Clearly, $m > 2$.*

Proof. See Appendix B.1. □

To conclude this section, we illustrate the simple moment structure of the data generating process. It is straightforward to see that for $i = 1, \dots, N$ we have $\mathbb{E}[x_{it}|\mathscr{F}_{t-1}] = 0$ and $\mathbb{V}[x_{it}|\mathscr{F}_{t-1}] = (f_t\sigma_{it})^2$. In addition, $\mathbb{V}[x_{it}] = f_t^2$ and this means that we can retrieve asymptotic results for both large T and N without any further assumption on the covariance structure of the model.

3.3.2 Invertibility

We now turn to the invertibility property of the updating equations in model (3.5), and look at the parameter updating recursions as functions of the observed data $\{x_t\}_{t\in\mathbb{N}}$.

It should be noted how the filtering procedure of the model works as follows. First, we filter the common factor $\{\hat{f}_t\}_{t\in\mathbb{N}}$, which does not require knowledge of the idiosyncratic component. Second, we re-scale the data $x_{it}/\hat{f}_t$ and filter for each series the idiosyncratic terms.

In practice, the recursion for the common factor and the idiosyncratic components must be started at some fixed values $\hat{f}_1^2$ and $\hat{\sigma}_{i1}^2$ for $i = 1\ldots,N$. From these starting points we retrieve the estimated paths $\{\hat{f}_t^2\}$ and $\{\hat{\sigma}_{it}^2\}$ for $t \in \mathbb{N}$ and $i = 1\ldots,N$. It is clear then these initialization will crucially affect the whole filtering process, since the actual re-scaled series are $\{x_{it}/\hat{f}_t\}_{t\in\mathbb{N}}$ from which we retrieve the $\{\hat{\sigma}_{it}^2\}_{t\in\mathbb{N}}$. Thus, each $\hat{\sigma}_{it}^2$ is a function of $x_{it}/\hat{f}_t$.

We are now ready to state conditions which ensure the invertibility of the updating equations in (3.6). Assumption 5 imposes parameter restrictions that are sufficient to obtain filter invertibility. Additionally, it imposes the compactness of the parameter space, which will be useful for establishing the consistency of the MLE.

Assumption 5. *The parameter space* $\boldsymbol{\Theta} = \{\boldsymbol{\Lambda}\times\boldsymbol{\Psi}\}$ *is compact and satisfies*

1. $|\beta-\alpha| < 1$ *and*

2. $\max_{i=1,\ldots,N}\{|\phi_i-\kappa_i|,|\phi_i+\kappa_i\nu_i|\} < 1.$

Unlike other recent articles which also deal with robust nonlinear filtering methods for time-varying locations (e.g. Harvey and Luati, 2014, Blasques et al., 2018) we cannot rely on standard contraction theorems, such as Bougerol's Theorem 3.1 in Bougerol, 1993 to obtain invertibility. Instead, given the multivariate factor structure, we will apply a sequential method for proving invertibility, and work instead with the contraction theorem for perturbed stochastic recursions stated in Straumann, 2005, Therem 2.6.4 ; see also Straumann and Mikosch, 2006. The proof of Proposition 3 below relies on first starting with the common factor filter and handling the idiosyncratic filter on a second stage.

Proposition 3 establishes the invertibility of the filtered common factor and the idiosyncratic volatilities under the parameter space restrictions imposed by Assumption 5.

Proposition 3. *Consider the* N*-dimensional vector process* $\{\boldsymbol{x}_t\}_{t\in\mathbb{Z}}$ *be generated from model* (3.2) *under assumptions 2–4, such that is stationary and ergodic. In addition, impose Assumption 5, then the filtered common factor* $\{\hat{f}_t\}_{t\in\mathbb{N}}$*, started at some fixed point* $\hat{f}_1^2 \in \mathbb{R}^+$*, and the perturbed filters* $\{\hat{\sigma}_{it}^2\}_{t\in\mathbb{N}}$ *for* $i = 1\ldots,N$ *converge exponentially almost surely and uniformly to their respective stationary and ergodic solutions* $\{f_t\}_{t\in\mathbb{Z}}$ *and* $\{\sigma_{it}^2\}_{t\in\mathbb{Z}}$ *for* $i = 1\ldots,N$*, that is*

$$\sup_{\theta\in\Theta}\|\hat{f}_t^2 - f_t^2\| \xrightarrow{e.a.s.} 0 \qquad \text{and} \qquad \sup_{\theta\in\Theta}\|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \xrightarrow{e.a.s.} 0 \qquad \text{as} \qquad t\to\infty, \tag{3.8}$$

for $i = 1\ldots,N$*. Moreover, the filtered* $(N+1)$*-dimensional sequence* $\{(\hat{f}_t^2,(\hat{\sigma}_t^2)^\top)^\top\}_{t\in\mathbb{N}}$ *converges to a unique stationary and ergodic solution, for any initialization* $\{(\hat{f}_1^2,(\hat{\sigma}_1^2)^\top)^\top\} \in \mathbb{R}_+^{N+1}$,

$$\sup_{\theta\in\Theta}\|(\hat{f}_t\hat{\sigma}_t)^2 - (f_t\sigma_t)^2\| \xrightarrow{e.a.s.} 0 \qquad \text{as} \qquad t\to\infty.$$

Proof. See Appendix B.1. □

3.4 Maximum Likelihood Estimation

Having established the stochastic properties of our factor model both as a data generating process (Section 3.1) and as a filter (Section 3.2), we can now turn to the sample properties of the MLE. In this section, we introduce the maximum likelihood procedure for the estimation of model (3.2), we write the time-varying common factor and the idiosyncratic terms as explicit function of the unknown parameter vectors, that is

$$f_t = f_t(\boldsymbol{\lambda}) \qquad \textit{and} \qquad \sigma_{it} = \sigma_{it}(\boldsymbol{\psi}_i) \qquad i = 1, \ldots, N.$$

The conditional density is given by

$$p(x_{it}|\mathscr{F}_{t-1}, \boldsymbol{\lambda}, \boldsymbol{\psi}_i) = \frac{\Gamma[(\nu_i+1)/2]}{\Gamma[\nu_i/2]\sqrt{\pi(\nu_i-2)}(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))}\left[1+\frac{x_{it}^2}{(\nu_i-2)(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}\right]^{-[(\nu_i+1)/2]}. \tag{3.9}$$

and then the conditional log-likelihood for a single observation of the i-th time series has the following form

$$\begin{aligned}\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) &= \log\Gamma\left[\frac{\nu_i+1}{2}\right] - \log\Gamma\left[\frac{\nu_i}{2}\right] - \frac{1}{2}\log(\nu_i-2) - \frac{1}{2}\log\pi \\ &\quad - \frac{1}{2}\log\left[(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\right] - \frac{(\nu_i+1)}{2}\log\left[1+\frac{x_{it}^2}{(\nu_i-2)(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}\right].\end{aligned} \tag{3.10}$$

Therefore, the maximum likelihood estimator boils down as the solution of

$$\hat{\boldsymbol{\theta}}_{NT} = (\hat{\boldsymbol{\lambda}}_T^\top, \hat{\boldsymbol{\psi}}_{1T}^\top, \ldots, \hat{\boldsymbol{\psi}}_{NT}^\top)^\top = \arg\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}, \boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_N\in\boldsymbol{\Upsilon}} \sum_{i=1}^N\sum_{t=1}^T \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i).$$

The asymptotic analysis of the model require the definition of the empirical expectations of the observed log-likelihood functions of the filtered processes initialized at some fixed f_1 and σ_{i1} for $i = 1, \ldots, N$ and the unfeasible log-likelihood function where its starting values are drawn from the stationary distribution respectively. However, every marginal likelihood will depends on the common factor when $N > 1$, hence we have

$$\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\ell}_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \qquad \textit{and} \qquad \mathcal{L}_{NT}(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i).$$

Then, we define the limits

$$\mathcal{L}_N(\boldsymbol{\theta}) = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)\right] \qquad \textit{and} \qquad \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)].$$

3.4.1 Consistency

This section establishes the consistency of the MLE as $T \to \infty$. Naturally, the results can also be obtained for both large T and large N. This is relevant for application where N is very large. As a corollary, we describe also consistency as $T \to \infty$ and $N \to \infty$ sequentially.

Assumption 6. *The true parameter* $\boldsymbol{\theta}_0 = (\boldsymbol{\lambda}_0^\top, \boldsymbol{\Upsilon}_0^\top)^\top$ *is an element of the parameter space* $\boldsymbol{\Theta}$.

To prove strong consistency for both T and N large, we make use of two preliminary Lemmas which facilitate the discussion in proving the main consistency theorem. Lemma 3.4.1 builds on the stationarity, moments and invertibility results of Section 3, and establishes the uniform convergence of the log-likelihood $\mathcal{L}_{NT}$ over the parameter space Θ. Lemma 3.4.2 obtains the identification of the true parameter $\boldsymbol{\theta}_0 \in \Theta$.

Lemma 3.4.1. *Let Assumptions 2–6 hold. Then, the limits $\mathcal{L}_N(\boldsymbol{\theta})$ and $\mathcal{L}(\boldsymbol{\theta})$ are both well defined and $\mathbb{E}[|\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)|] < \infty$ for $i = 1,\dots,N$ and every $t = 1,\dots,T$. Thus,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0 \quad as \quad T\to\infty \qquad and \qquad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_N(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\| \to 0 \quad as \quad N\to\infty. \tag{3.11}$$

Proof. See Appendix B.1. □

Lemma 3.4.2. *Under the Assumptions of Lemma 3.4.1, $\mathcal{L}_N(\boldsymbol{\theta})$ and $\mathcal{L}(\boldsymbol{\theta})$ are uniquely maximized at the true parameter $\boldsymbol{\theta}_0$ i.e. $\mathcal{L}_N(\boldsymbol{\theta}_0) > \mathcal{L}_N(\boldsymbol{\theta})$ and $\mathcal{L}(\boldsymbol{\theta}_0) > \mathcal{L}_N(\boldsymbol{\theta})$ for every $\boldsymbol{\theta} = (\boldsymbol{\lambda}^\top,\boldsymbol{\Psi}^\top)^\top \in \boldsymbol{\Theta}$ and $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$.*

Proof. See Appendix B.1. □

We are now in position to prove the main result of this sub-section. Theorem 3.4.3 establishes the consistency of the MLE for the well-specified dynamic factor model, as $T\to\infty$. As a corollary, it also follows that the MLE is consistency when both $T\to\infty$ and $N\to\infty$ sequentially.

Theorem 3.4.3 (Strong Consistency). *Consider model (3.2), satisfying Assumptions 2–6. Then, one obtains*

$$\hat{\boldsymbol{\theta}}_{NT} \to \boldsymbol{\theta}_0 \;\; almost\ surely \;\; as \;\; T\to\infty. \tag{3.12}$$

Proof. See Appendix B.1. □

Corollary 3.4.3.1 (Strong Consistency). *Under the conditions of Theorem 3.4.3 we also have,*

$$\hat{\boldsymbol{\theta}}_{NT} \to \boldsymbol{\theta}_0 \;\; almost\ surely \;\; as \;\; T\to\infty \;\; and \;\; N\to\infty \;\; sequentially. \tag{3.13}$$

Proof. See Appendix B.1. □

3.4.2 Asymptotic Normality

In proving asymptotic normality of the ML estimator several steps are needed. We may begin by exploring the limit behaviour of the differential processes $\{d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{N}}$. To this end, we provide the following Proposition

Proposition 4. *Under Assumptions 2, 3, 4 and 5, the differential $\{d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{N}}$ achieve a unique stationary ergodic solution $\{d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{Z}}$. Moreover, for any initializations of the original filter $\{\hat{f}_1^2,(\hat{\sigma}_1^2)^\top\}$, the perturbed differential $\{d(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2\}$ will eventually converge to the same stationary ergodic solution for $i = 1,\dots,N$.*

Proof. See Appendix B.1. □

The number of bounded moments it will be also of interest later on and so, in addition to the previous proof of the existence and almost sure exponentially fast convergence of the differential processes to a unique stationary ergodic solution, we provide the following result, which deal with the number of bounded moments of this solution.

Proposition 5. *Under Assumptions 2, 3, 4 and 5, the stationary ergodic differential $\{d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{Z}}$ has m' bounded moments uniformly over the parameter space, that is $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2|^{m'}]$ where*

$$m' = \frac{n'_f n_\sigma}{n_\sigma + n'_f} + \frac{n_f n'_\sigma}{n'_\sigma + n_f}, \tag{3.14}$$

where, in addition to the notation of Proposition 2, n'_f are the number of bounded moments of the differential of the common factor $\{d(f_t(\boldsymbol{\lambda})\}$ and $n'_\sigma = \min_{i=1,\dots,N} n'_{\sigma_i}$ the minimum between the number of bounded moments of the differential of the idiosyncratic components $\{d(\sigma_{it}(\boldsymbol{\psi}_i))^2\}$. Clearly, $m' > 2$.

Proof. See Appendix B.1. □

Next, we present the following Lemma where we show that the first differential of the conditional likelihood forms a martingale difference sequence with zero mean and finite variance.

Lemma 3.4.4. *Under Assumptions 2, 3, 4 and 5, the differential of the conditional likelihood $\{d\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)\}$ is a martingale difference sequence, i.e. $\mathbb{E}[d\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)|\mathscr{F}_{t-1}] = 0$ and moreover $\mathbb{E}[|d\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)|^2]$ exists and is finite.*

Proof. See Appendix B.1. □

Now we establish the asymptotic distribution of the score function.

Lemma 3.4.5. *Under Assumptions 2, 3, 4 and 5, we obtain*

$$\sqrt{NT}\mathcal{L}'_{NT}(\boldsymbol{\theta}) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{V}), \qquad as \qquad N, T \to \infty, \qquad where$$

$$\boldsymbol{V} = \mathbb{E}\left[\frac{d\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{d(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)} \frac{d\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{d(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)^\top}\right].$$

Proof. See Appendix B.1. □

We are ready now to enter in the real of the second differentials and present a new proposition.

Proposition 6. *Under Assumptions 2, 3, 4 and 5, the second derivative $\{d^2(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{N}}$ achieve a unique stationary ergodic solution $\{d^2(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{Z}}$. Moreover, for any initialization of the original processes $\{\hat{f}_1^2, (\hat{\sigma}_1^2)^\top\}$ and $\{d\hat{f}_1^2, (d\hat{\sigma}_1^2)^\top\}$, the perturbed second differential $\{d^2(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2\}$ will eventually converge to the same stationary ergodic solution for $i = 1, \dots, N$.*

Proof. See Appendix B.1. □

Almost at the end of our excursion, we turn into the analysis of the likelihood's second differential. Specifically, we need to prove that the empirical second differential processes, converge to their limits, which exist and are both well defined, yielding non singular matrices when taking the derivatives with respect to the vector of parameters.

Lemma 3.4.6. *Let Assumptions 2, 3, 4 and 5 hold. Then, the limits $\mathcal{L}''_N(\boldsymbol{\theta})$ and $\mathcal{L}''(\boldsymbol{\theta})$ are both well defined and moreover the latter is nonsignular for $i = 1, \dots, N$ and every $t = 1, \dots, T$. Thus,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}''_{NT}(\boldsymbol{\theta}) - \mathcal{L}''_N(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0 \quad as \quad T \to \infty \qquad and \qquad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}''_N(\boldsymbol{\theta}) - \mathcal{L}''(\boldsymbol{\theta})\| \to 0 \quad as \quad N \to \infty. \tag{3.15}$$

Proof. See Appendix B.1. □

In conclusion, we present the last theorem which shows the asymptotic gaussianity of the maximum likelihood estimator.

Theorem 3.4.7 (Asymptotic Normality of the MLE). *Consider model (3.2), satisfying Assumptions 2, 3, 4 and 5. Then,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{NT} - \boldsymbol{\theta}_0) \overset{\mathscr{D}}{\Rightarrow} \mathscr{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)^{-1}),$$

where, $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ *is the Fisher Information matrix evaluated at the true parameter vector* $\boldsymbol{\theta}_0$.

Proof. See Appendix B.1. □

3.5 Monte Carlo Experiment

With a simulation study we aim to investigate the finite sample properties of estimators. In particular, we specify a data generating process with model (3.6), and directly generate vector of time series. The idiosyncratic components are simulated from univariate first-order *Beta-t-GARCH* and then plugged into the common factor. With this scheme we ensure that the simulated series of innovations $\{\epsilon_{it}\}$ have standard univariate Student's t distribution for each i. Therefore, the co-movements and cross-sectional dependence betweeen the series $\{x_{it}\}$ is directly induced by the fixed equation for the common factor f_t.

The experiment consider a panel of $N = 5$ time series with $T = 2000$, and we perform the simulation scheme for $M = 1000$ times. The results of the ML estimates for the parameters of the common factors are displayed in Table 3.1, while the results for the idiosyncratic components are in Table 3.2.

To show the capability of our model to extract a dynamic common factor from a panels of returns, we report in Figure 3.1 the distribution of the estimated dynamic common factor, and the real common factor used for the simulation scheme.

From the results we deduce that the ML estimators have good finite sample properties for both the components of our model. The empirical mean and the root mean square error confirm the fact that the bias of the estimator is very low and are rather precise. Also, from the 95% of the simulated distribution of the estimated dynamic common factor it is apparent that the the model tracks the real common trend, yielding a satisfactory e reliable path.

TABLE 3.1: *Monte Carlo Experiment - Parameters of the Common Factor*

	ω	α	β
Actual	0.100	0.050	0.900
Mean	0.101	0.051	0.897
RMSE	0.003	0.015	0.034

TABLE 3.2: *Monte Carlo Experiment - Parameters of the Idiosyncratic Volatilities*

	ϕ_1	κ_1	ν_1
Actual	0.800	0.100	6.000
Mean	0.812	0.102	6.122
RMSE	0.095	0.017	0.564
	ϕ_2	κ_2	ν_2
Actual	0.800	0.100	6.000
Mean	0.798	0.101	6.136
RMSE	0.069	0.016	0.513
	ϕ_3	κ_3	ν_3
Actual	0.800	0.100	6.000
Mean	0.797	0.101	6.167
RMSE	0.076	0.017	0.482
	ϕ_4	κ_4	ν_4
Actual	0.800	0.100	6.000
Mean	0.808	0.102	6.106
RMSE	0.089	0.018	0.826
	ϕ_5	κ_5	ν_5
Actual	0.800	0.100	6.000
Mean	0.810	0.102	6.120
RMSE	0.089	0.017	0.815

3.6 Empirical Application

In this section we present some results about the application of our dynamic factor model and other competing specifications to a panel of ten stocks from S& P500. The financial time series extend from the 1st of January 1999 to the 31 December 2018 and they are plotted in Figure 3.2.

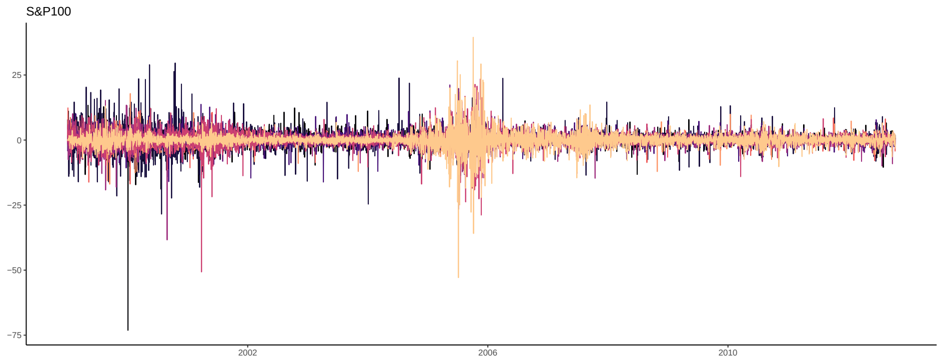


FIGURE 3.2: This panel shows the ten stock returns from S & P500.

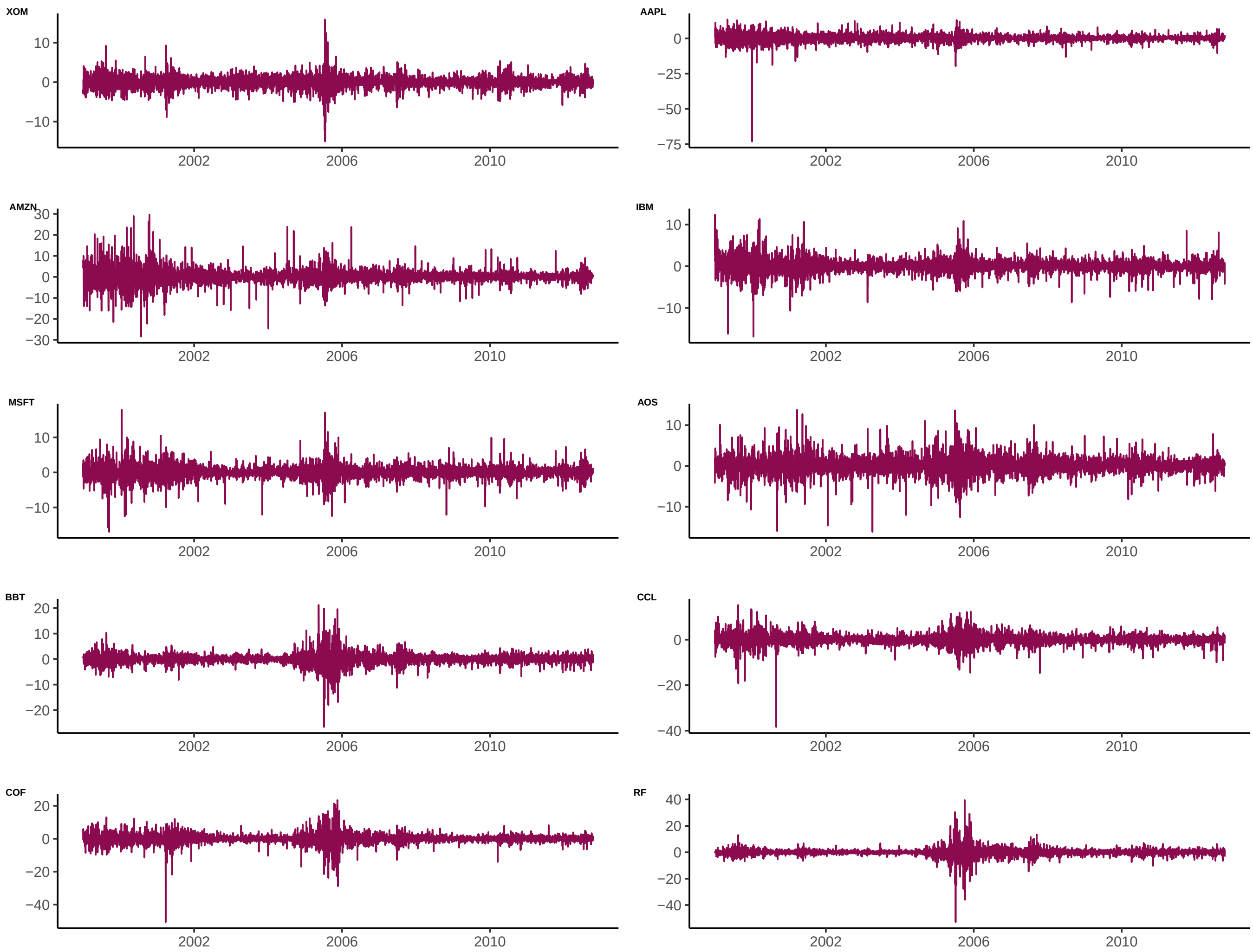


FIGURE 3.3: This panel shows the ten stock returns from S & P500 separately.

To illustrate the importance of robust approach to study the underlying behaviour and properties of portfolio returns, a preliminary analysis might offer some insights about the observed data. The panel of time series displayed in Figure 3.3 highlights the common properties shared by financial time series, the so-called stylized facts of asset returns, such as presence of outliers, volatility clusters ecc.

Moreover, the concurrent correlation matrix in Figure 3.4 shows that the series have significant amount of concurrent correlations, indeed, the multivariate Portmanteau test confirm the existence of this dependence at the 5% significance level However, while the sample correlation matrix is a consistent estimator, it is biased in finite sample. Thus, we follow the suggestion of Tsay, 2005 and with a bootstrap resampling method we overcome that issue. Moreover, from Figures 3.2 and 3.3 it could be seen that there are 5 main stocks and 5 small stocks and just by checking the data in these graphs one can easily note that there is some common pattern in the volatilities, with the presence of several jumps and outliers and in particular, there is much idiosyncratic behaviour.

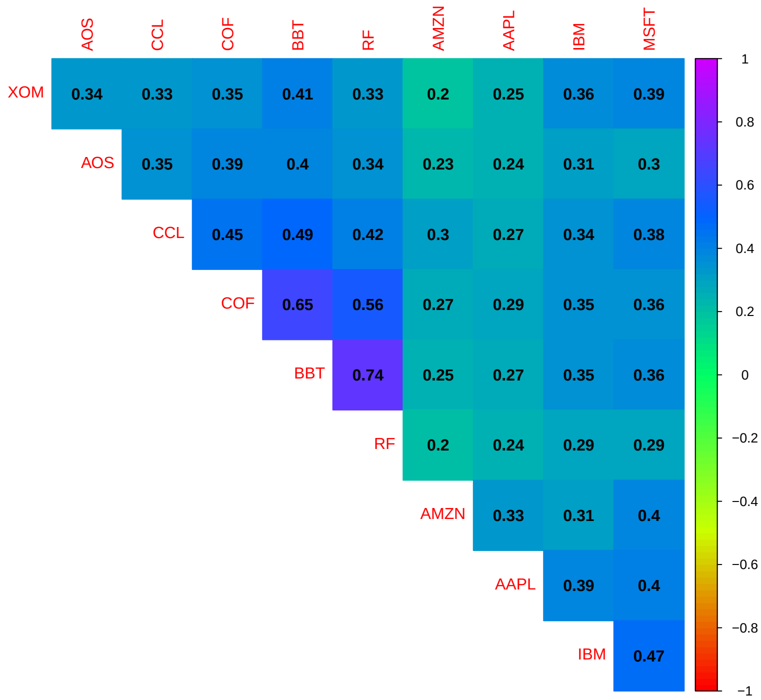


FIGURE 3.4: Sample Cross-Correlation matrix of the daily returns of the considered ten stocks from S& P500. A multivariate Portmanteau test (see Tsay, 2005) were carried out in order to verify the statistical significance at level of 5%.

In order to explore the dependence structure of the observed processes, it might be also useful to have a look at the scatter plot matrix, which is reported in Figure 3.5. This graph is often of interest in empirical works, since it is straightforward method to detect clustering of observations in the joint quadrants. For instance, with this simple descriptive tool one might be able to detect tail dependence between pair of stock returns, which measures the probability of the considered couple of variables, lying below their q-quantile, see for example Oh and Patton, 2017. An inspection of Figure 3.5 for each pair of our panel of time series, reveals several clustering of observations, both in the joint positive and negative quadrants.

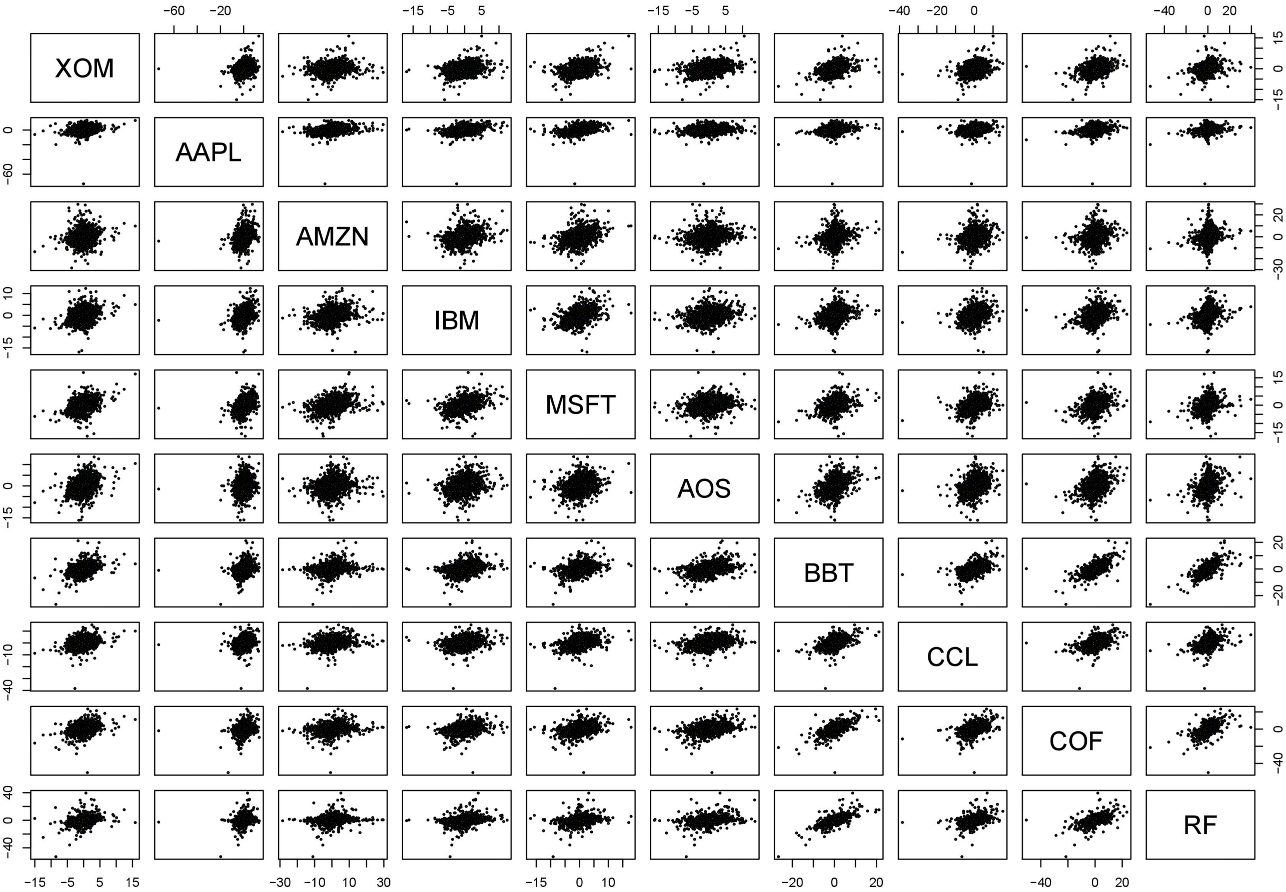


FIGURE 3.5: Scatter plot matrix of pairs of time series.

We begin our empirical analysis with our proposed model, with specifications given in (3.5) and in order to show the differences between the proposed dynamic single-factor score-driven model with the model of Creal, Koopman, and Lucas, 2011, we perform several comparison to the same panel of ten stocks from S&P500. To illustrate the importance of a robust approach when study the underlying behaviour and properties of portfolio returns, we further compare the in-sample performances of both the models specified with a Gaussian distribution and a Student's t distribution.

The maximum likelihood estimates of the single factor model proposed in Creal, Koopman, and Lucas, 2011, Section 4.3 are displayed in Table 3.3, while the maximum likelihood estimates of the proposed single factor model are displayed in Table 3.4.

By comparing these results, it is possible to note that the proposed dynamic factor model outperforms the multivariate dynamic model of Creal, Koopman, and Lucas, 2011 by means of *AIC*, *BIC* and *HIQ* criteria, and hence it may be a better choice when modeling cross-correlated financial returns. However, both the models confirm that the introduction of a dynamic common factor could improve the in-sample fitting performance, and moreover, the Student's t specifications are preferred. This is directly implied by the score-driven filtering procedure, since the change of the conditional distribution assumption for the financial returns yields a different filter, which may be more suitable according to the initial conditions to handle the stylized facts of the considered time series. Several advantages of the method can be listed.

First, when a Student's t distribution is considered, the dynamic multivariate model of Creal, Koopman, and Lucas, 2011 restricts the degrees of freedom parameter ν to be the same across the marginals, so that, the same degree of fatness is imposed to the whole panel of time series. In contrast, the new proposed filtering procedure allows for more flexibility in accommodating the relevant empirical features of the idiosyncratic terms, because we treat each of the term as independent components of the multivariate volatility model. In particular, the idiosyncratic volatilities are allowed to have different degrees of freedom, which can be crucial when the panel of time series shows heterogeneity across the marginal behavior of the return changes. In fact, it could be seen from the ML estimates, reported in Table 3.4 that the estimated degrees of freedom ν_i for the stock returns are quite different, ranging from a minimum of 4 and a maximum of 8 approximately. Therefore, the presence of excess kurtosis in the observed data is largely confirmed. Each of the estimated ϕ's, which are the autoregressive coefficients of the filters, are really high, almost hitting the stationary bound, which may suggest a possible violation of the stationary conditions for the idiosyncratic components. On the other hand, the highly persistence of these terms is well handled by the driving-force, namely the conditional score. In fact, from the estimated κ's we may observe the differences between the less and more influential player of the market, which might be appreciated even more from the filtered paths of conditional volatilities displayed in Figure 3.7. The estimated paths are remarkably different and hence we may conclude that our robust score-driven dynamic factor model is able to capture this source of heterogeneity.

Second, the multivariate model of Creal, Koopman, and Lucas, 2011 involves matrix recursions which need to be evaluated for every available observation, thus its computational burden is challenging. On the other hand, with the proposed approach univariate models con be considered separately and maximize univariate likelihoods. Thus concluding to an increase in the computational speed considerably.

Altogether, the proposed approach to model both the common dynamic factor and the idiosyncratic terms appears to be a good alternative with respect to the multivariate approach of (Creal, Koopman, and Lucas, 2011).

TABLE 3.3: *Estimated Parameters of the Idiosyncratic Volatilities from the Factor Score-Driven Volatility Model of Creal, Koopman, and Lucas, 2011.*

	XOM	APPL	AMZN	IBM	MSFT	AOS	BBT	CCL	COF	RF
					Non-Robust - (Gaussian)					
ϕ	0.960	0.932	0.920	0.913	0.913	0.963	0.980	0.930	0.930	0.963
	(0.001)	(0.002)	(0.010)	(0.012)	(0.023)	(0.022)	(0.012)	(0.022)	(0.013)	(0.013)
κ	0.050	0.052	0.034	0.034	0.034	0.017	0.034	0.034	0.034	0.052
	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.003)
		AIC= 152839.6			BIC= 155984.1			HIQ= 154751.3		
					Non-Robust with Common Factor - (Gaussian)					
ϕ	0.786	0.898	0.986	0.990	0.997	0.976	0.923	0.907	0.889	0.989
	(0.003)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.008)	(0.014)	(0.000)
κ	0.213	0.100	0.014	0.007	0.017	0.016	0.071	0.012	0.098	0.009
	(0.013)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.007)	(0.013)	(0.000)
		AIC= 150649.6			BIC= 155446.8			HIQ= 154395.4		
					Robust - (Student's t)					
ϕ	0.999	0.989	0.998	0.999	0.988	0.988	0.987	0.998	0.999	0.989
	(0.003)	(0.001)	(0.000)	(0.005)	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)
κ	0.010	0.054	0.034	0.010	0.008	0.005	0.010	0.003	0.008	0.005
	(0.000)	(0.001)	(0.020)	(0.050)	(0.001)	(0.000)	(0.000)	(0.012)	(0.000)	(0.000)
ν	3.895									
	(0.845)									
		AIC= 152446.4			BIC= 155102.5			HIQ= 154133.6		
					Robust with Common Factor - (Student's t)					
ϕ	0.970	0.968	0.880	0.975	0.097	0.978	0.959	0.959	0.971	0.914
	(0.002)	(0.001)	(0.012)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.010)
κ	0.027	0.026	0.115	0.021	0.024	0.019	0.033	0.036	0.025	0.075
	(0.001)	(0.002)	(0.012)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.011)
ν	4.325									
	(0.416)									
		AIC= 150232.5			BIC= 155029.6			HIQ= 153529.6		

TABLE 3.4: *Estimated Parameters of the Idiosyncratic Volatilities from the Factor Score-Driven Volatility Model of my thesis.*

	XOM	APPL	AMZN	IBM	MSFT	AOS	BBT	CCL	COF	RF
					Non-Robust - (*Gaussian*)					
ϕ	0.988	0.999	0.989	0.997	0.994	0.999	0.998	0.989	0.999	0.998
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
κ	0.063	0.037	0.019	0.035	0.030	0.020	0.054	0.032	0.038	0.063
	(0.041)	(0.027)	(0.014)	(0.025)	(0.020)	(0.017)	(0.035)	(0.020)	(0.024)	(0.047)
	AIC= 145521.6			BIC= 145652.1			HIQ= 145567.3			
					Non-Robust with Common Factor - (*Gaussian*)					
ϕ	0.997	0.998	0.999	0.999	0.999	0.588	0.999	0.824	0.996	0.999
	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.010)	(0.002)	(0.007)	(0.002)	(0.001)
κ	0.040	0.013	0.008	0.009	0.007	0.094	0.022	0.050	0.010	0.027
	(0.031)	(0.016)	(0.021)	(0.033)	(0.015)	(0.019)	(0.047)	(0.026)	(0.032)	(0.028)
	$\omega = 0.053$			$\alpha = 0.101$			$\beta = 0.974$			
	(0.166)			(0.023)			(0.008)			
	AIC= 137220.7			BIC= 137370.7			HIQ= 137273.2			
					Robust - (*Student's t*)					
ϕ	0.991	0.999	0.999	0.989	0.996	0.999	0.995	0.999	0.999	0.999
	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.002)
κ	0.085	0.075	0.092	0.097	0.092	0.072	0.081	0.072	0.085	0.095
	(0.011)	(0.010)	(0.014)	(0.015)	(0.010)	(0.017)	(0.025)	(0.020)	(0.024)	(0.017)
ν	7.851	4.843	4.059	4.811	4.624	4.405	6.389	4.890	5.098	6.470
	(0.485)	(0.235)	(0.244)	(0.267)	(0.224)	(0.279)	(0.460)	(0.320)	(0.312)	(0.489)
	AIC= 134352.9			BIC= 135548.7			HIQ= 134421.5			
					Robust with Common Factor - (*Student's t*)					
ϕ	0.996	0.996	0.998	0.996	0.990	0.902	0.998	0.932	0.997	0.998
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
κ	0.063	0.038	0.037	0.058	0.066	0.091	0.040	0.059	0.026	0.049
	(0.008)	(0.008)	(0.010)	(0.007)	(0.010)	(0.015)	(0.014)	(0.010)	(0.014)	(0.015)
ν	5.464	4.959	4.614	4.820	4.856	4.788	5.244	5.002	5.100	5.284
	(0.422)	(0.134)	(0.143)	(0.206)	(0.223)	(0.259)	(0.358)	(0.289)	(0.254)	(0.401)
	$\omega = 0.054$			$\alpha = 0.102$			$\beta = 0.974$			
	(0.154)			(0.012)			(0.005)			
	AIC= 134121.4			BIC= 135336.6			HIQ= 134196.8			

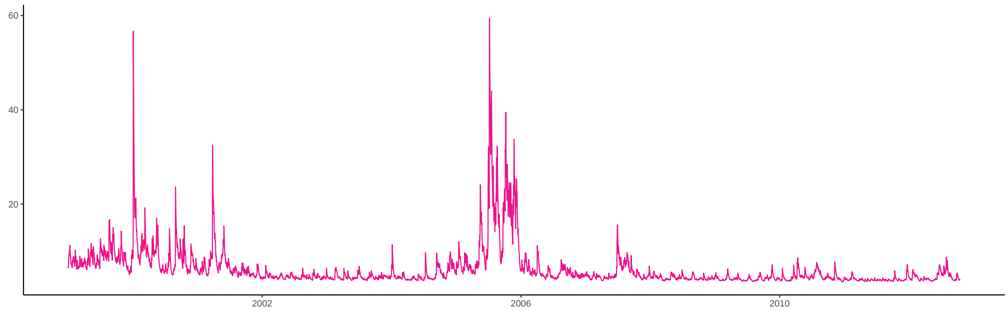


FIGURE 3.6: Common variance extracted from the ten stock returns from S&P 500 by the Factor Score-Driven Volatility Model

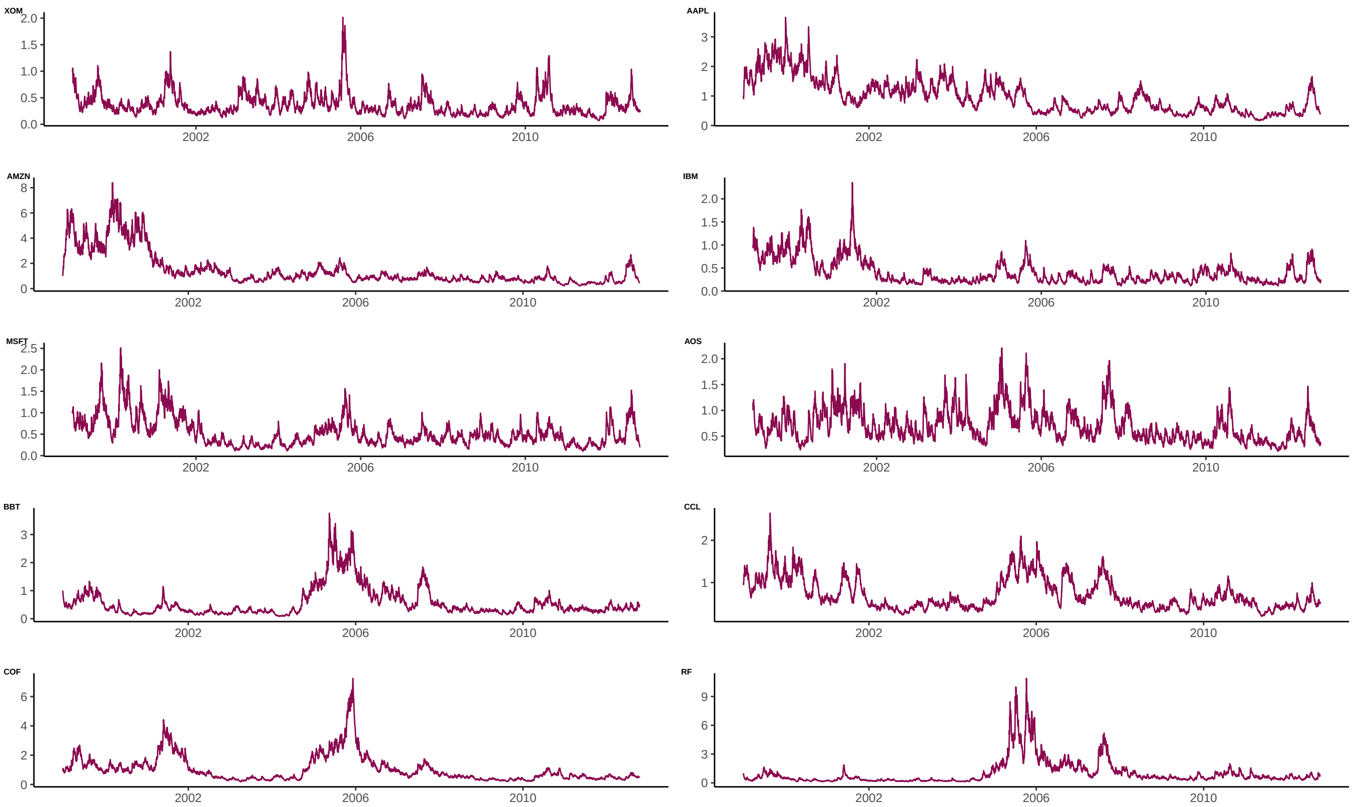


FIGURE 3.7: This panel shows the idiosyncratic volatilities for the ten stock returns from S&P 500 sperately.

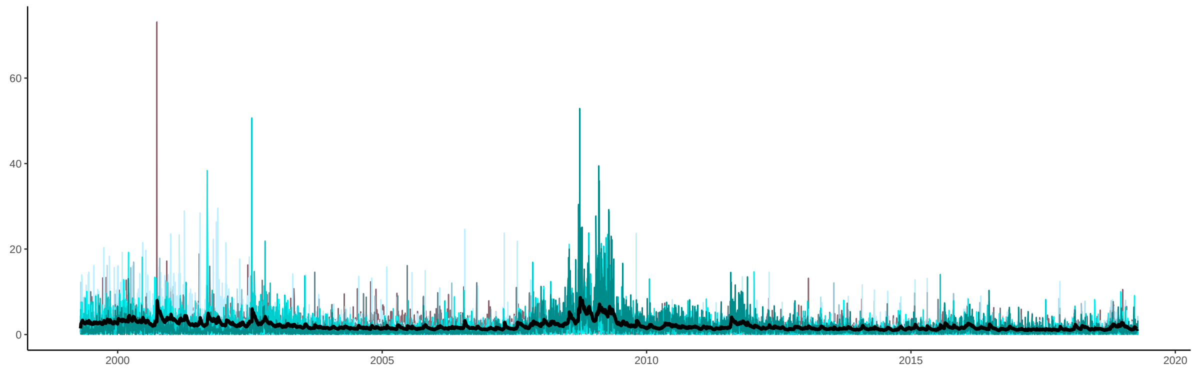


FIGURE 3.8: Absolute values of the observations from the S&P500 time series panel of stock returns. In black the filtered dynamic common factor.

In Figure 3.8 it is displayed the absolute value of the panel of stock returns from S&P500. The square root of the filtered dynamic common dynamic factor is drawn with the thick black line. The presence of the outliers in the panel of time series is evident. However, large values are not embedded in the conditional variance and we still get a robust approximation of the cross-sectional volatility.

3.7 Conclusion

This paper introduced a new nonlinear dynamic factor model for conditional volatilities which features a score-robust updating equation for the idiosyncratic components. We derived stochastic properties for the model, including bounded moments, stationarity, ergodicity, and filter invertibility. Additionally, we established the consistency and asymptotic normality of the maximum likelihood estimator in large samples. A Monte Carlo study showed that the MLE has good finite sample properties. Finally, an empirical illustration using a panel of ten stocks from the S&P500 was used to highlight the advantages of the proposed dynamic factor structure over other competing models as well as the need for robust filtering techniques for the idiosyncratic component in financial data sets.

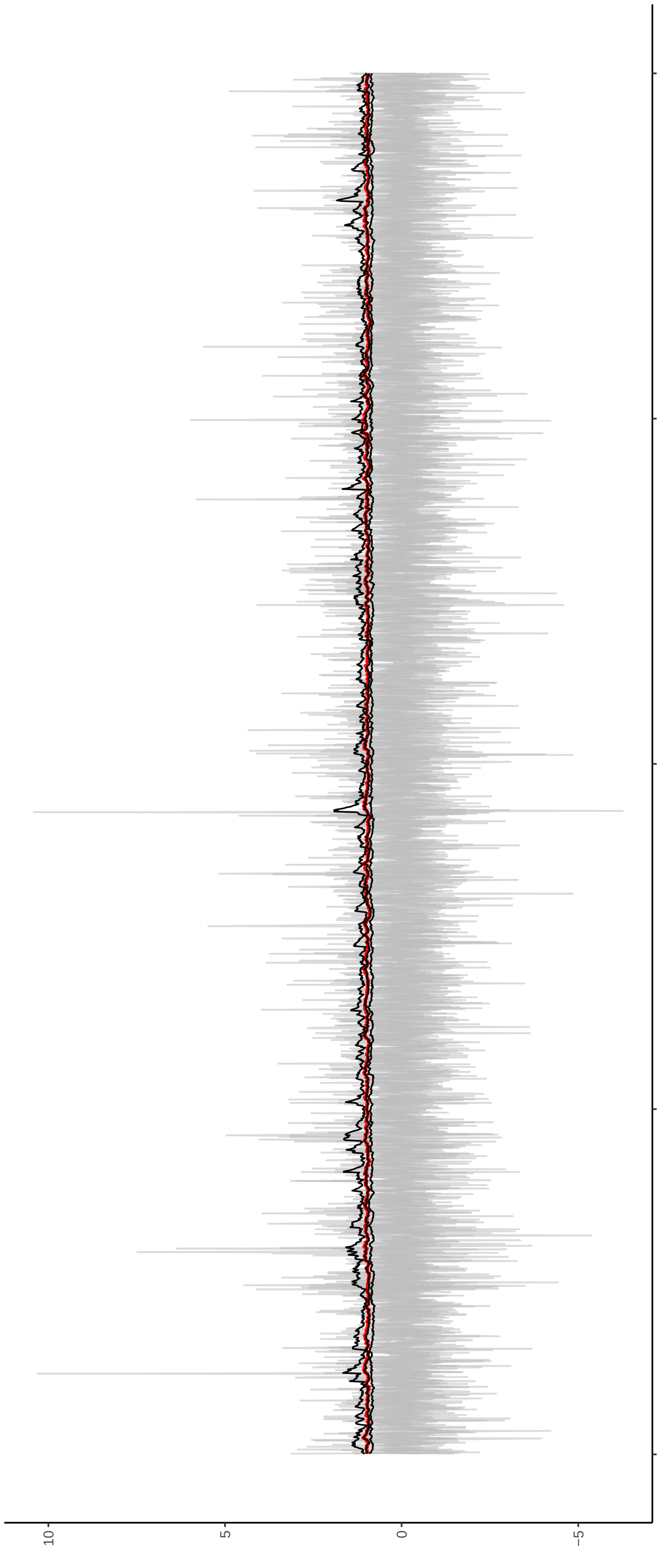


FIGURE 3.1: The graph shows the median of the real dynamic common factor of the DGP (thick red line) together with the 95% quantile range and the median of the distribution of the estimated dynamic common factor. The shaded series are some of the generated return paths.

Chapter 4

Bivariate EGAS Time Series Models

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name
Department of Statistical Sciences “Paolo Fortunati”

Doctor of Philosophy in Statistical Sciences

Essays in Robust and Nonlinear Multivariate Time Series Models

by Enzo D’INNOCENZO

It is well known that taking nonlinear dynamic filters from univariate to multivariate settings is challenging due to the likely failure of the sufficient contraction conditions used to ensure filter invertibility. Our main contribution is to show that only invertible filters may provide reliable and stable miximum likelihood estimators. We first review two newly introduced exponential multivariate score driven models for conditional volatilities and we also incorporate a time-varying conditional correlation coefficient, as a possible extension of the model. After that, we propose an empirical method to verify that the invertibility condition holds. Our method is simple and can be used in practice to find an empirical invertible domain. Therefore, constrained optimization routine can be performed in order to obtain reliable maximum likelihood estimators. This is crucial to ensure that the estimation and filtering results are not spurious. We show the practical relevance of our results by means of an empirical application to stock returns.

4.1 Introduction

In financial econometrics, considerable amount of research has focused on the analysis of multiple time series. More specifically, it is often of interest the study of dynamic interactions and spillovers between two stocks, and the work of Baillie and Myers, 1991 or Herwartz and Lütkepohl, 2000 are two examples on how insightful can be such an analysis.

The early paper of Bollerslev, Engle, and Wooldridge, 1988 and later Bollerslev, 1990 introduced a convenient way to model the covariance and measuring the systematic risk associated between assets. This analyses were performed by means of classes of multivariate *GARCH* models, namely the *VEC* and constant conditional correlations (*CCC*) GARCH models, allowing for the first time the possibility to define general structures able to deal with multivariate financial time series. As a natural extension, Jeantheau, 1998 proposed the class of extended *CCC-GARCH*, or *ECCC-GARCH*. These models has proved to be particularly useful when measuring financial spillovers is in our concerns, since the volatility of one stock is modeled as a linear combination of lagged squared innovations and volatility of its own plus the same components of the other equations in the system. Since the work of Conrad, Gultekin, and Kaul, 1991, there has been considerable attention on the investigation of volatility spillovers with multivariate *GARCH*. In this latter work, the author provides profound evidences of the fact that impacts between firms may be relevant to understand the future dynamics of their own returns, the so-called causality in variance. An important example where it is shown the usefulness of multivariate GARCH, and in particular the *ECCC-GARCH*, in measuring such a causality properties, can be found in Conrad and Weber, 2013. Other authors, such as Hafner and Herwartz, 2008 and Nakatani and Teräsvirta, 2009, propose several testing procedure for volatility contagion.

Due to its intuitive and simple structure, the class of extended constant conditional correlations *GARCH* models is nowadays considered as a benchmark in order to measure how fluctuations in the price of an asset influence changes in the prices of other assets. Indeed, before turning on more complex specifications, it is often useful to test phenomena such as volatility interactions among markets. For instance, Nakatani and Teräsvirta, 2009 propose a Wald-type test which can helps researchers and model-builders in substantial computational efficiency gains, by avoiding unnecessary parameter estimations, and merely consider to estimate a *CCC-GARCH*. More recently, Pedersen, 2017 concentrate his work on the properties of the quasi-maximum likelihood estimator (QMLE) in the case where some of the elements of the coefficient matrices are on the boundary of the parameter space, under the null hypothesis of no spillovers.

Another fundamental aspect during multivariate analyses of financial returns lies in the so called negative spillovers. In particular, empirical evidences demonstrate the importance of the *ECCC-GARCH* in allowing for negative *ARCH* and *GARCH* propulsion in the model. For this reason, the work of Conrad and Karanasos, 2010 focuses on the sufficient conditions which ensures the positive definiteness of the conditional covariance matrix even when some of the parameters are negative. A further possible reason for their success relies in the fact that the study of the stationarity and regularity conditions is remarkably simple, see Francq and Zakoian, 2019, so that a comprehensive asymptotic theory is readily available. Considering the above discussion, it is worth mentioning that Francq and Zakoïan, 2012 provide the asymptotic theory for the general class of asymmetric *CCC-GARCH*, since it has recently received distinguished attention in empirical applications, especially from the model-builder who wants to consider such specifications for testing economic theories.

Despite their empirical usefulness and efficient likelihood-based estimation techniques, *ECCC-GARCH* still suffer of common problems shared by several multivariate specifications of autoregressive conditionally heteroskedastic models. Among the others, we first mention the curse of dimensionality, since the number of parameters that need to be estimated dramatically increases as the number of the assets grows. Some simplified version have been proposed in the literature, but sometimes leading to over-simplification and then landing into the realm of *CCC-GARCH* where no spillover is allowed. In addition, the linearity assumptions imposed by the recursions of this class of models may result rather restrictive in several applications. To overcome these drawbacks, our aim is to propose a flexible yet parsimonious multivariate *GARCH* model which does not suffer of such limitations. Thus, in our framework, we will consider dynamic equations for the conditional variances which include nonlinear features, such as residuals and conditional variances interactions, and rely in a feasible parametric structure, in order to maintain the total computational effort reasonable.

Indeed, motivated by the difficulties which affect a general multivariate conditional heterosckedastic model we propose a novel class of observation-driven bivariate models, namely the bivariate score-driven *EGAS* models, where the last acronym stands for exponential generalized autoregressive score. Our work may be also considered as an attempt to try to break the linearity constraints imposed by *CCC-GARCH*, while still relying on verifiable and appropriate regularity conditions. Perhaps, the biggest challenge during the development of a proper asymptotic theory of dynamic, and possibly multivariate, nonlinear models, is the definition of feasible and verifiable stability conditions, because as it is well-known, these models cannot be written with the endogenous variables equal to a vector valued function plus an additive error. For this reason, another focus of the present paper is in the analysis of the form of the full region that ensure the invertibility of the underlying stochastic processes. More precisely, relying on the well-known Theorem 3.1 of Bougerol, 1993, we obtain necessary conditions under which the considered bivariate score-driven model is invertible. However, this conditions are often very restrictive or not possible to being verified in practice. Thus, inspired by the empirical approach proposed by Blasques et al., 2018 for univariate observation-driven models, we study the shape of the invertibility region as the values of the static parameters vary.

The knowledge of the underlying stochastic properties of nonlinear models is of crucial importance for several reasons. For instance, when deriving asymptotic properties of the corresponding parameter estimates. Indeed, when considered as a filter for retrieving the dynamics of the volatilities, the invertibility of the model give us the opportunity to approximate with arbitrary low degrees of uncertainty the driving noise of the process. This has also another interpretation, namely, none of the initial conditions of the filtering process are relevant and, more precisely, they are asymptotically negligible and one can easily acknowledges the fundamental importance of that property for purposes of reliable empirical analysis. Score-driven models have proven to be a very powerful and flexible tool able to capture complex and possibly nonlinear dynamics in different scientific fields. Thanks to their generality and observation-driven set up, this class of models allows the researchers in specifying highly-nonlinear robust models able to deal with different features of nonlinear time series. As an example, one can easily accommodate the heavy-tailedness nature or the non-Gaussianity of the observed processes by specifying a fat-tailed score-driven model.

Nevertheless, statistical properties of this models may be very challenging to be established and often, they are not verifiable in practice. Things become harder when moving to multivariate frameworks, where the dependence structure of the filters makes the regularity conditions analytically intractable. Among these issues, we concentrate our work on the invertibility property of our bivariate score-driven volatility model and in particular we will consider two different specifications. For the sake of its analytical tractability, the first is a Gaussian based volatility score-driven model, while the second is a bivariate extension of the *Beta-t-EGARCH*, model of Harvey, 2013, which as already prove to be able to deal with extreme observations.

4.2 Bivariate Score-Driven *EGAS* Models

In this section we introduce the class of Bivariate Score-Driven *EGAS* Models and present the explicit analytical form of the score which is the driving-force of the whole system. To be specific, the two class of model that we are going to define are labelled as the *EGAS-g* and the *EGAS-t*. The first class rely on an error term which is normally distributed, while the second class on the Student's t distribution. However, it is necessary to first introduce the general framework which we collocate our class of score-driven models. Thus, following the specification introduced by Bollerslev, 1990 or Jeantheau, 1998, we consider a zero-mean bivariate constant conditional correlation *GARCH* (*CCC-GARCH*) volatility model. More precisely, we consider a bivariate vector-valued sequence of random variables $\{\boldsymbol{\epsilon}_t\}$ for $t = 1, \ldots, T$, where $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})^\top$, is generated by the *CCC-GARCH* framework,

$$\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\eta}_t, \qquad \mathbb{V}[\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1}] = \boldsymbol{\Sigma}_t = \boldsymbol{D}_t\boldsymbol{R}\boldsymbol{D}_t, \qquad \boldsymbol{\eta}_t \sim IID(\mathbf{0}_2, \boldsymbol{I}_2), \tag{4.1}$$

where $\mathscr{F}_{t-1}$ collects the past information up to $t-1$ and

$$\boldsymbol{D}_t = \begin{bmatrix} \exp f_{1t} & 0 \\ 0 & \exp f_{2t} \end{bmatrix}, \qquad \boldsymbol{f}_t = \log[\mathrm{diag}(\boldsymbol{D}_t)], \qquad \boldsymbol{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{4.2}$$

$\boldsymbol{\Sigma}_t$ is the dynamic conditional variance-covariance matrix, which is modeled by using the the well-known decomposition specified with the matrix $\boldsymbol{D}_t$, with the volatilities on the main diagonal, where the exponential link function is adopted in order to ensure positivity. $\boldsymbol{R}$ is the static correlation matrix and we note that since

$$\rho = \frac{\mathbb{E}_{t-1}[\epsilon_{1t}\epsilon_{2t}]}{\mathbb{E}_{t-1}[\epsilon_{1t}^2]\mathbb{E}_{t-1}[\epsilon_{2t}^2]}$$

by construction, the conditional correlation coefficient between the pair of returns is time-invariant. However, it is still possible to retrieve indirectly a time-varying covariance process by using $\mathbb{E}_{t-1}[\epsilon_{1t}\epsilon_{2t}] = \rho \exp\{f_{1t} + f_{2t}\}$

For the dynamics of the time-varying volatilities $\exp \boldsymbol{f}_t \in \mathbb{R}^2$, a two-dimensional score-driven updating scheme is adopted, such that

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \alpha\boldsymbol{s}_t + \beta\boldsymbol{f}_t, \tag{4.3}$$

where $\boldsymbol{f}_t = (f_{1t}, f_{2t})^\top$ are updated thanks to the driving force $\boldsymbol{s}_t = (s_{1t}, s_{2t})^\top$, that is the score vector of the predictive log-likelihood $\partial \ln p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1})/\partial f_{it}$ for $i = 1, 2$. $\boldsymbol{\omega} \in \mathbb{R}^2$ is a bivariate vector of intercepts and $\alpha, \beta \in \mathbb{R}$, with $|\beta| < 1$, are scalar parameters. We collect all the involved unknown parameters in a vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta} \in \mathbb{R}^5$ is the compact parameter space. Notably, no positivity constraints are required for the parameters since the score functions in $\boldsymbol{s}_t$ and the adopted exponential link function will accomodate the volatilities f_{it} in the proper space for every t and $i = 1, 2$, which is enough to ensure that they still strictly positive.

We are now ready to present the distributional assumptions needed in order to properly define our class of score-driven volatility models. We begin with the details for the bivariate Gaussian *EGAS* (*EGAS-g*) and then we turn to the bivariate Student's *t EGAS* (*EGAS-t*).

4.2.1 Bivariate Gaussian

The bivariate Gaussian conditional density with zero mean and an exponential link-function for the variance parameters assumes the form

$$\begin{aligned} p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1}) =& \frac{1}{2\pi \exp\{f_{1t} + f_{2t}\}\sqrt{1-\rho^2}} \\ &\times \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right], \end{aligned} \tag{4.4}$$

where $0 \leq \rho \leq 1$ is the constant conditional correlation coefficient.

Thus, by taking logs of (4.4), we get

$$\ln p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1}) = -\ln 2\pi - f_{1t} - f_{2t} - \frac{1}{2}\ln(1-\rho^2) - \frac{1}{2(1-\rho^2)} \times \Big[\epsilon_{1t}^2 \exp\{-2f_{1t}\} - 2\rho\epsilon_{1t}\epsilon_{2t}\exp\{-(f_{1t}+f_{2t})\} + \epsilon_{2t}^2\exp\{-2f_{2t}\}\Big], \tag{4.5}$$

and so, in order to complete the dynamic specifications of the two-dimensional updating scheme, for $i = 1,2$ we need

$$s_{it} = \frac{\partial \ln p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1})}{\partial f_{i,t}} = \frac{1}{(1-\rho^2)}\Big(\epsilon_{it}^2\exp\{-2f_{it}\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(f_{1t}+f_{2t})\}\Big) - 1. \tag{4.6}$$

4.2.2 Bivariate Student's *t*

The bivariate Student's *t* conditional density with zero mean and an exponential link-function for the variance parameters assumes the form

$$p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1}) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)(\nu-2)\exp\{f_{1t}+f_{2t}\}\sqrt{1-\rho^2}} \times \left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right]^{(\nu+2)/2}, \tag{4.7}$$

where $0 \leq \rho \leq 1$ is the constant conditional correlation coefficient and $\nu > 2$ is the kurtosis parameter, or in other words the degrees of freedom of the distribution. Note that we have parametrized the variance-covariance matrix by imposing $nu > 2$, such that variances exist. Otherwise, one always work in terms of the scale matrix, which allows to drop this parameter constraint. However, in what follows, we still impose $\nu > 2$ and work with the convariance specification.

As before, by taking logs of (4.7), we get

$$\ln p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1}) = \ln\Gamma\left(\frac{\nu+2}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \ln[2\pi(\nu-2)] - f_{1t} - f_{2t} - \frac{1}{2}\ln(1-\rho^2) - \frac{\nu+2}{2}\ln\left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right] \tag{4.8}$$

and so, analogously, in order to complete the dynamic specifications of the two-dimensional updating scheme, for $i = 1,2$ we need

$$s_{it} = \frac{\partial \ln p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1})}{\partial f_{i,t}} = \frac{(\nu+2)}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{it}^2\exp\{-2f_{it}\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(f_{1t}+f_{2t})\}}{1+\frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)}\right) - 1. \tag{4.9}$$

4.2.3 Spillovers and Impact of the Scores

The aim of this subsection is to highlights the capability of the bivariate *EGAS* models in producing volatility spillovers. In particular, we will show that our proposed model is able to transmit information between the two equations of the system, while retaining in a parsimonious and restriction-free parameter structure. The linkages also allow for negative spillovers, which is known to be relevant for empirical applications. Moreover, we will further illustrate that one may attain totally different filters, which may or may not be suitable for a particular application.

To start the discussion, consider the Gaussian conditional score in equation (4.6). For fixed values of the volatility processes, one can investigate the properties of that curve as a function of the observations. Figure 4.1 below displays the susceptibility of the Gaussian score function s_{it} to the observations ϵ_{1t} and ϵ_{2t} for fixed t.

Although its smoothness and well behaviour, it is trasparent that score-driven volatility filter with Gaussian specification is not suitable for heavy-tailed or contaminated time series. In fact, as $|\epsilon_{it}| \rightarrow \infty$ for $i = 1, 2$, one sees that the score also diverges to infinity. Stated differently, the function is unbounded and monotonically increasing in $|\epsilon_{it}|$ for $i = 1, 2$. It follows that extreme observations may have arbitrarily large impact on the score-driven recursion (4.3).

In contrast, specifying a score-driven filter with the Student's t specification will yelds a different scenario. It is well-known that the Student's t family of distributions posses certain robustness properties, which are particularly useful for financial applications.

From Figure 4.2 one sees that the conditional score is bounded both from below and above and hence the driving-force s_t is robust. As a consequence, the filters in f_t with this specifications will inherits that property, yielding a robust filter which mitigates the impact of aberrant observations or outliers. It is worthwhile to note that as the responses are paired, outliers can occur in either or both coordinates and therefore, the boundedness of volatility spillovers may be crucial.

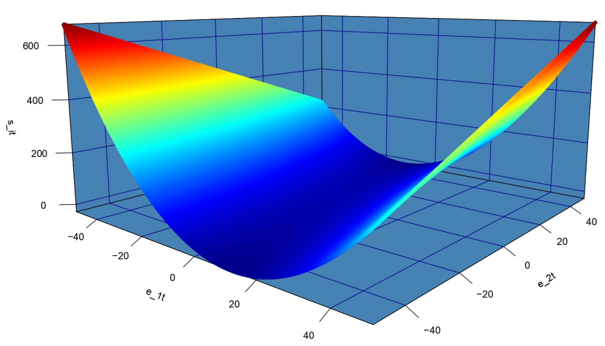


FIGURE 4.1: *Sensitivity of the bivariate Gaussian score as a function of* ϵ_t

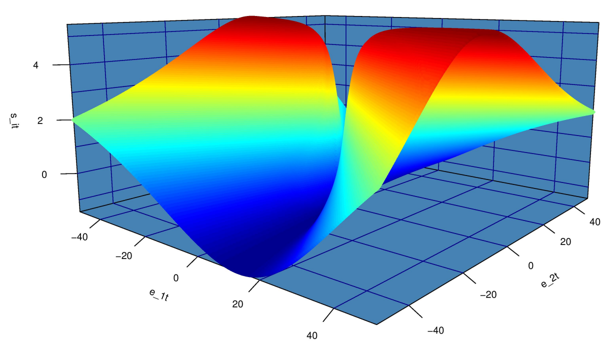


FIGURE 4.2: *Sensitivity of the bivariate Student's* t *score with* $\nu = 4$ *as a function of* ϵ_t.

To summarize, when an extreme observation hits the system, the Gaussian score will reacts fiercely in the updating equation for the time-varying volatilities, while the Student's t it will smooth out. However, the difference between the Student's t score in (4.9) and the Gaussian in (4.6), is only substantial for finite and low values of ν, the tail parameter, since as $\nu \rightarrow \infty$ the *EGAS-t* collapses to the *EGAS-g*.

4.3 Stochastic Properties

In this section, we explore the stochastic properties of the model as a Data Generating Process (DGP) and as a filter. We shall affront and give primitive conditions for the asymptotic properties of maximum likelihood estimator for both the Gasussian and Student's t case.

4.3.1 Stationarity, Ergodicity and Moments of the Model

The discussion about the properties of the updating equations as a DGP can be made by first re-writing the dynamics of the volatility processes as a function of the innovations, which might be useful to highlight its stochastic properties as Markov chain. More precisely, we could rewrite the transition process as an homogeneous chain

$$\boldsymbol{f}_t = \boldsymbol{\Phi}_{\boldsymbol{\theta}_0}(\boldsymbol{\Sigma}_{t-1}^{1/2}\boldsymbol{\eta}_{t-1}, \boldsymbol{f}_{t-1}), \qquad t \in \mathbb{Z}, \tag{4.10}$$

where the nonlinear map $\boldsymbol{\Phi}_{\boldsymbol{\theta}_0}(\,\cdot\,)$ should be seen as a random transformation evaluated with the true parameter $\boldsymbol{\theta}_0$.

Since $|\beta| < 1$, one can note that the sequences $\{f_{it}\}$ for $i = 1, 2$, constitute causal $AR(1)$ processes. Therefore, from the standard theory of linear processes, it is straightforward to see that $|\beta| < 1$ is the necessary and sufficient condition for the existence of the unique strictly stationary and ergodic solutions $\{f_{it}\}_{t\in\mathbb{Z}}$ for $i = 1, 2$, which are given by the following representations

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \sum_{k=0}^{\infty} \alpha\beta^k \boldsymbol{s}_{t-k}, \qquad t \in \mathbb{Z}. \tag{4.11}$$

We can summarize our findings with the following theorem, which establishes the existence of a unique strictly stationary and ergodic solution with n bounded moments for both of the random sequences $\{f_{1,t+1}\}$ and $\{f_{2,t+1}\}$.

Theorem 4.3.1. *Let η_{1t} and η_{2t} be* IID *with mean* 0 *and such that $\mathbb{E}[\|\boldsymbol{\eta}_t \otimes \boldsymbol{\eta}_t\|^n] < \infty$ for $n > 0$. They have time-invariant positive definite covariance matrix $\boldsymbol{R} = [\rho_{ij}]$, for $i, j = 1, 2$ and $i \neq j$, with ones on the main diagonal. Suppose further that $|\beta| < 1$. Then,* (4.10) *admits a unique strictly stationary and ergodic solutions $\{f_{it}\}_{t\in\mathbb{Z}}$ for $i = 1, 2$ with n bounded moments.*

The stationary and ergodic solutions of the EGAS-g *and* EGAS-t *have for $i = 1, 2$ the following representations*

$$f_{i,t+1} = \frac{\omega}{1-\beta} + \sum_{k=0}^{\infty} \alpha\beta^k \left[\frac{1}{(1-\rho^2)} \eta_{i,t-k}^2 - \frac{\rho}{(1-\rho^2)} \eta_{1,t-k}\eta_{2,t-k} - 1 \right], \qquad t \in \mathbb{Z},$$

for the former, while for the latter

$$f_{i,t+1} = \frac{\omega}{1-\beta} + \sum_{k=0}^{\infty} \alpha\beta^k \left\{ \frac{(\nu+2)}{(1-\rho^2)(\nu-2)} \left[\frac{\eta_{i,t-k}^2 - \rho\eta_{1,t-k}\eta_{2,t-k}}{1 + \frac{1}{(\nu-2)(1-\rho^2)} \left(\eta_{1,t-k}^2 - 2\rho\eta_{1,t-k}\eta_{2,t-k} + \eta_{2,t-k}^2 \right)} \right] - 1 \right\} \quad t \in \mathbb{Z}.$$

Proof. See Appendix C.1. □

4.3.2 Invertibility

Having solved the problem of stationarity, it is possible to explore the properties of the dynamic equations in (4.3) as a bivariate filter. As it is common in empirical applications, given the observations $\epsilon_{i1}, \ldots, \epsilon_{iT}$ for $i = 1, 2$, we wish to approximate the bivariate volatility process $\{\boldsymbol{f}_t\}$. This may be done by choosing fixed points f_{i1} for $i = 1, 2$, and then start some sensible recursions which are able to extract the desired unobserved processes. These recursions are known as the filtering equations. We stress the facts that the process $\{\hat{f}_{it}(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ for $i = 1, 2$, is now a filter and hence should be considered as a random elements in a space of continuous functions, which take values from the product space $\boldsymbol{\mathcal{E}} \times \boldsymbol{\theta}$, where $\boldsymbol{\mathcal{E}} = \mathbb{R}^2$, or more compactly, $\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) : \boldsymbol{\mathcal{E}} \times \boldsymbol{\theta} \to \mathbb{R}^2$. Of course, the true parameter vector, $\boldsymbol{\theta}_0$ is unknown and we try to obtain its best approximation with $\boldsymbol{\theta}$, that in turn must be estimated.

It is important to note that $\{\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\}$ is not stationary ergodic and hence, in order to obtain a reliable approximation of the underlying volatility bivariate process, it is necessary that our filter satisfy some stability conditions, which will ensure its convergence to a stationary ergodic solution. The notion of invertibility is extremely important since it ensures that the approximative stochastic recursion $\{\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\}$ started at $\boldsymbol{f}_1$ will eventually coincide with the true $\{\boldsymbol{f}_t(\boldsymbol{\theta})\}$. For univariate observation-driven models, the works of Wintenberger, 2013, Blasques et al., 2018 and Martinet and McAleer, 2018 provide an extensive discussion about the contraction conditions of this type and the importance of its implications. For multivariate nonlinear frameworks as ours, however, a little is disposable in the literature. As regards multivariate *GARCH* models, the relevant conditions required for the stability of the filters are quite complicated or unknown, as it is for *DCC-GARCH* model of Engle, 2002. Recently, the discussion of invertibility for a class of multivariate *GARCH* model has been tackled by Darolles, Francq, and Laurent, 2018.

As we will see, under the appropriate Bougerol's contraction condition, one can obtain the exponentially fast almost sure convergence (see Straumann, 2005 for a formal definition) of the filtered parameters $\{\hat{f}_{i,t+1}(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ to the stationary and ergodic solutions $\{f_{it}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, for any fixed starting points f_{i1} with $i=1,2$. In particular, the random maps of the recursions in $\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})$ have to be contractive on average. As it is a common practice in non linear models, to study the aforementioned stability conditions, one may rely again on Theorem 3.1 of Bougerol, 1993, or analogously Theorem 6.12 of Pötscher and Prucha, 1997.

Here a sufficient condition for the invertibility of the bivariate *EGAS* is

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) = \mathbb{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \left\| \boldsymbol{\Lambda}_t(\boldsymbol{\theta}) \right\| \right] = \mathbb{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \left\| \frac{\partial}{\partial \boldsymbol{f}_t^\top} \left\{ \boldsymbol{\omega} + \alpha \boldsymbol{s}_t + \beta \boldsymbol{f}_t \right\} \right\| \right] < 0. \tag{4.12}$$

Other weaker conditions may be available. This latter coefficient is known as the Bougerol's contraction, (Bougerol, 1993). It differs from the well-known Lyapunov exponent from the fact that it takes the sup over the parameter $\boldsymbol{f}$, while the Lyapunov exponent, it just takes the expectation, see Tong, 1990 and McCaffrey et al., 1992 for a comprehensive theoretical and empirical treatment of the subject.

The Lyapunov exponent is usually employed to quantify the level of perturbation of a nonlinear system. In other words, one can study how a little perturbation on the initial state of a dynamic system will affect the subsequent states. Apparently, Bougerol's contraction will get arbitrarily close to the Lyapunov coefficient as we unfold the model backwards, but it is always more restrictive. At this point it is useful to show what are the implications of the stated new contraction condition.

With some abuse of notation, we let $\tilde{\boldsymbol{f}}_t(\boldsymbol{\theta})$ denotes a set of points that are between $\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})$ and $\boldsymbol{f}_t(\boldsymbol{\theta})$, element-by-element. Thus, an application of the multivariate mean value theorem yields

$$\|\hat{\boldsymbol{f}}_{t+1}(\boldsymbol{\theta}) - \boldsymbol{f}_{t+1}(\boldsymbol{\theta})\| \leq \left\| \frac{\partial \hat{\boldsymbol{f}}_{t+1}(\boldsymbol{\theta})}{\partial \tilde{\boldsymbol{f}}_t(\boldsymbol{\theta})^\top} \right\| \|\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\|$$

for $i=1,2$. In the terminology of McCaffrey et al., 1992, the right-hand side of the latter inequality is also known as the tangent map system.

Clearly, the contraction condition require the calculation of

$$\frac{\partial \hat{\boldsymbol{f}}_{t+1}(\boldsymbol{\theta})}{\partial \hat{\boldsymbol{f}}_t(\boldsymbol{\theta})^\top} = \begin{bmatrix} \frac{\partial \hat{f}_{1,t+1}(\boldsymbol{\theta})}{\partial \hat{f}_{1t}(\boldsymbol{\theta})} & \frac{\partial \hat{f}_{1,t+1}(\boldsymbol{\theta})}{\partial \hat{f}_{2t}(\boldsymbol{\theta})} \\ \frac{\partial \hat{f}_{2,t+1}(\boldsymbol{\theta})}{\partial \hat{f}_{1t}(\boldsymbol{\theta})} & \frac{\partial \hat{f}_{2,t+1}(\boldsymbol{\theta})}{\partial \hat{f}_{2t}(\boldsymbol{\theta})} \end{bmatrix} = \beta \boldsymbol{I}_2 + \alpha \frac{\partial \hat{\boldsymbol{s}}_t(\boldsymbol{\theta})}{\partial \hat{\boldsymbol{f}}_t(\boldsymbol{\theta})^\top} = \begin{bmatrix} \beta & 0 \\ 0 & \beta \end{bmatrix} + \alpha \begin{bmatrix} \frac{\partial \hat{s}_{1t}(\boldsymbol{\theta})}{\partial \hat{f}_{1t}(\boldsymbol{\theta})} & \frac{\partial \hat{s}_{1t}(\boldsymbol{\theta})}{\partial \hat{f}_{1t}(\boldsymbol{\theta})} \\ \frac{\partial \hat{s}_{2t}(\boldsymbol{\theta})}{\partial \hat{f}_{1t}(\boldsymbol{\theta})} & \frac{\partial \hat{s}_{2t}(\boldsymbol{\theta})}{\partial \hat{f}_{2t}(\boldsymbol{\theta})} \end{bmatrix}. \tag{4.13}$$

It is evident that if the norm of the obtained stochastic matrix is contractive on average, together with some other regularity condition explained below, we are able to ensure that our sequence of random maps converges almost surely and exponentially fast to the strictly stationary and ergodic solution, as a corollary of Theorem 2.10 of Straumann and Mikosch, 2006 and existence of the log-moment condition of the stationary solution given by Theorem 4.3.1 above.

To summarize, by properly restricting the parameter space, one can obtain the desired exponentially fast convergence of the approximated process towards the stationary ergodic solution. In this respect, it is important to note that the norm has to be carefully chosen as it changes substantially the value of (4.12) which may yields a more or less restrictive parameter spaces.

Thus, the above discussion leads us to the next general theorem, which will ensure the invertibility of our class of bivariate models, at the cost of restricting the parameter space, in particular α and β.

Theorem 4.3.2. *Let* $\boldsymbol{\Theta}$ *be the compact parameter space, let* $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ *generated by* (4.1)-(4.2) *be strictly stationary and ergodic with* $\mathbb{E}[\|\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t\|^n] < \infty$ *for* $n > 0$. *Suppose that there exist a nonrandom initial condition* $\boldsymbol{f}_1 = (f_{11}, f_{21})^\top$, *where* $f_{i1} \in [(\omega_i - \alpha)/(1-\beta), \infty)$ *for* $i = 1, 2$, *such that*

$$\mathbb{E}\left[\log^+ \left\|\frac{\partial}{\partial \boldsymbol{f}_1^\top}\left\{\boldsymbol{\omega} + \alpha \boldsymbol{s}_t + \beta \boldsymbol{f}_1\right\}\right\|\right] < \infty, \tag{4.14}$$

and

$$\mathbb{E}\left[\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}} \left\|\frac{\partial}{\partial \boldsymbol{f}_t^\top}\left\{\boldsymbol{\omega} + \alpha \boldsymbol{s}_t + \beta \boldsymbol{f}_t\right\}\right\|\right] < 0. \tag{4.15}$$

Then the filtered parameters $\{\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ *are invertible and irrespective of the choice of* $\boldsymbol{f}_1$ *one has that*

$$\|\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\|_1 \xrightarrow{e.a.s.} 0 \qquad \text{as} \qquad t \to \infty.$$

Proof. See Appendix C.1. □

Unfortunately, it seems to be impossible to derive explicit conditions for Theorem 4.3.2. However, it is shown in Harvey, 2013, Pag. 217 that if the score vector is pre-multiplied by the inverse of the information matrix, it the situation may be much better. In fact, for the Gaussian case, the re-scaled volatility filters are as for a univariate model and there is no need to assume that the α and β parameters are the same in each equation. It is therefore important to see if a similar result holds for the bivariate Student's t case. If it does, then invertibility results for the univariate *Beta-t-EGARCH* model can be applied; see Harvey and Lange, 2017, Pag. 182 where a sufficient condition for invertibility in terms of the model parameters is given.

We then derive the information matrix for the considered bivariate Student's t volatility model and further inspect the properties of the resulting filters, in order to verify if the aforementioned sufficient condition of Harvey and Lange, 2017 applies.

In matrix form, we can write the log-likelihood bivariate Student's t density as

$$\ell_t(\boldsymbol{\theta}) = \ln\Gamma\left(\frac{\nu+2}{\nu}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \ln[\pi(\nu-2)] - \frac{1}{2}\ln|\boldsymbol{\Sigma}_t| - \frac{\nu+2}{2}\ln\left(1 + \frac{\boldsymbol{\epsilon}_t^\top \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t}{\nu-2}\right). \tag{4.16}$$

Differentiating (4.16) with respect to the vec-covariance matrix yields the following score vector

$$\boldsymbol{\nabla}_t^{\Sigma} = \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \operatorname{vec}\boldsymbol{\Sigma}_t} = \frac{1}{2}(\boldsymbol{\Sigma}_t^{-1} \otimes \boldsymbol{\Sigma}_t^{-1})\left[\frac{\nu+2}{\nu - 2 + \boldsymbol{\epsilon}_t^\top \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\epsilon}_t}(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t) - \operatorname{vec}\boldsymbol{\Sigma}_t^{-1}\right]. \tag{4.17}$$

where It can be shown that by differentiating (4.17) with respect to the vec-covariance matrix and taking conditional expectations, we get the conditional information matrix

$$\begin{aligned}\mathcal{I}^{\boldsymbol{\Sigma}}_{t|t-1} &= -\,\mathbb{E}\left[\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \operatorname{vec} \boldsymbol{\Sigma}_t \partial \operatorname{vec} \boldsymbol{\Sigma}_t^\top}\right] \\ &= (\boldsymbol{\Sigma}_t^{-1/2} \otimes \boldsymbol{\Sigma}_t^{-1/2}) \left[\frac{\nu+2}{\nu+4}\boldsymbol{I}_4 - \frac{1}{2(\nu+4)}(\operatorname{vec} \boldsymbol{I}_2)(\operatorname{vec} \boldsymbol{I}_2)^\top\right](\boldsymbol{\Sigma}_t^{-1/2} \otimes \boldsymbol{\Sigma}_t^{-1/2}),\end{aligned} \tag{4.18}$$

however, since $\boldsymbol{\Sigma}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t$ and $\boldsymbol{f}_t = \log[\operatorname{diag}(\boldsymbol{D}_t)]$, we need the Jacobian matrix

$$\boldsymbol{\Psi} = \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}_t}{\partial \boldsymbol{f}_t^\top} = [(\boldsymbol{D}_t \boldsymbol{R} \otimes \boldsymbol{I}_2) + (\boldsymbol{I}_2 \otimes \boldsymbol{D}_t \boldsymbol{R})](\boldsymbol{W}_{\boldsymbol{D}_t}),$$

where $\boldsymbol{W}_{\boldsymbol{D}_t}$ is a 4×4 matrix with diagonal elements $\operatorname{vec}(\boldsymbol{D}_t^{-1})$ and then dropping the zero-columns, see Creal, Koopman, and Lucas, 2011.

Therefore, the driving force of the vector recursion $\boldsymbol{f}_t$ is defined as

$$\boldsymbol{s}_t = (\boldsymbol{\Psi}^\top \mathcal{I}^{\boldsymbol{\Sigma}}_{t|t-1} \boldsymbol{\Psi})^{-1} (\boldsymbol{\Psi}^\top \boldsymbol{\nabla}^{\boldsymbol{\Sigma}}_t). \tag{4.19}$$

The resulting volatility filters obtained by scaling the conditional score as in equation (4.19), have a very complex structures, which are not as for the univariate Student's t case. Therefore, there the results shown in Harvey, 2013, Pag. 217 for the bivariate Gaussian case, do not apply for the bivariate Student's t case and in conclusion, the sufficient condition stated in Harvey and Lange, 2017 does not applies.

4.4 Empirical Estimation of the Invertibility Region

Unfortunately, a big drawback implied by the flexibility of our bivariate nonlinear class of models lies in the introduction of the product between the couple ϵ_{1t} and ϵ_{2t}. The presence of that term relates the recursions of the dynamic volatilities with past observations in a highly nonlinear fashion, making impossible to derive explicit invertibility conditions. Moreover, we should stress that the expectation operator in the conditions stated in Theorem 4.3.2 are taken with respect to the actual distribution, which is of course unknown. Moreover, even if we where told which is the real data generating process, we still remain with the issue that the real parameter vector $\boldsymbol{\theta}_0$ is indeed unknown.

Thus, a possible solution is to assume the correct specification of the model and restrict our parameter space such that the condition (4.12) is verified. Nevertheless, as well documented in the work of Blasques et al., 2018 this straightforward approach may give us a parameter space which is too small for empirical applications or even degenerate, making the resulting filter unusable for practical purposes. However they do not actually show if they are (or not) for this bivariate model.

This further motivate us to propose the class of bivariate *EGAS* models as a valid statistical tool to estimate and making forecast of correlated bivariate financial time series. Indeed, in the present section, we propose a new method to estimate the unfeasible contraction condition (4.12) and then to restrict our parameter space $\boldsymbol{\Theta}$ to a compact subset $\widehat{\boldsymbol{\Theta}}_T$ which ensures the invertibility conditions (4.14) and (4.15). With this approach, we are able to deliver feasible and easy to check sufficient conditions. Therefore, once maximum likelihood estimations are carried out under the obtained restrictions, one could rely on filters which are asymptotically stable, i.e. they do not depend on the chosen initialization of the recursions and will eventually converge to the unique stationary and ergodic solution, as $T \to \infty$.

We propose to restrict the optimization procedures of the predictive likelihood under the empirical invertibility constraint

$$\widehat{\boldsymbol{\Lambda}(\boldsymbol{\theta})} = T^{-1}\sum_{t=1}^{T}\log\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{f}\left\|\widehat{\boldsymbol{\Lambda}_t(\boldsymbol{\theta})}\right\|_1 \leq -\delta, \tag{4.20}$$

where with $\|\cdot\|_1$ we denote the L^1-Norm, that is $\|\boldsymbol{A}\|_1 = \max_j \sum_{i=1}^{n}|a_{ij}|$. In other words, L^1 matrix norm of a matrix is equal to the maximum of L^1 norm of a column of the matrix. $\delta > 0$ can be chosen as small as we want and where

$$\widehat{\boldsymbol{\Lambda}_t(\boldsymbol{\theta})} = \left\|\frac{\partial \hat{\boldsymbol{f}}_{t+1}(\boldsymbol{\theta})}{\partial \hat{\boldsymbol{f}}_t(\boldsymbol{\theta})^\top}\right\|_1.$$

As a consequence, we obtain the following restricted and compact parameter space

$$\widehat{\boldsymbol{\Theta}}_T = \left\{\boldsymbol{\theta}\in\boldsymbol{\Theta} : \widehat{\boldsymbol{\Lambda}(\boldsymbol{\theta})} \leq -\delta,\ \delta > 0\right\}. \tag{4.21}$$

In practice, with a simple simulation scheme, one can easily obtain the desired region. The next subsection will show the obtained regions for our bivariate score-driven volatility models.

4.4.1 Empirical Inveritibility Region of the *EGAS* Filters

We obtain the compact parameter space in (4.21) by simulating data from a fixed DGP and letting several combinations of α and β to vary across of a dense grid constructed by values which produce stationary and ergodic path. The values assumed by the intercepts ω_1 and ω_2 do not matters for the contraction condition. However, for both the Gaussian and the Student's t cases the contraction condition depend on the static correlation coefficient ρ and for the Student's t case, we also need to specify the degrees of freedom ν. Thus, we allow the DGPs used for the simulations to generate fairly correlated time series with $\rho = 0.5$ for both the *EGAS* models and we consider a quite amount of heavy tailedness by fixing $\nu = 5$ in the *EGAS-t*.

It is of interest to have a visual interpretation of the results. The graphs in Figures 4.3 and 4.4 display the obtained compact parameter space (4.21), from which it is possible to gain some useful insights. In fact, these graphs serves to visualize the level of severity that one should impose to the parameters α and β during the optimization procedures in order to ensure the invertibility of the *EGAS* filters implied by the estimated version of the Bougerol's contraction condition in (4.4). The blue area represents low values of $\widehat{\boldsymbol{\Lambda}(\boldsymbol{\theta})}$ which increase until we reach the red area, which means that the filters are very likely to be unstable.

Interestingly, the shape of the regions are quite skewed, showing elliptical and asymmetric area in both cases which further highlights the fact that ensure the invertibility property is not an easy task. It confirms also the well-known fact that as $|\beta| \to 1$ the value of the Bougerol's contraction condition increases very fast. We also note that the resulting $\widehat{\boldsymbol{\Theta}}_T$ for the *EGAS-t* is much larger than the one obtained from the Gaussian counterpart *EGAS-g*. It is evident then that the functional form and therefore, the shape of the conditional score is of crucial importance since condition (4.20) builds upon the L_1-norm of the matrix which contains its partial derivatives, i.e. the Hessian. The overall boundedness of the driving force of the *EGAS-t*, i.e. the function s_{it} for $i = 1, 2$ in (4.9) displayed in Figure 4.2, helps the filter to be contractive on average.

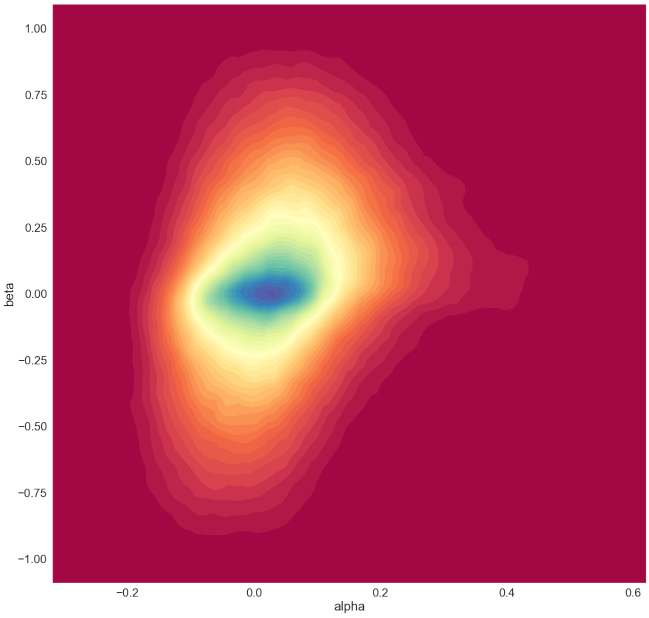


FIGURE 4.3: *Emprical invertibility region of the* EGAS-g

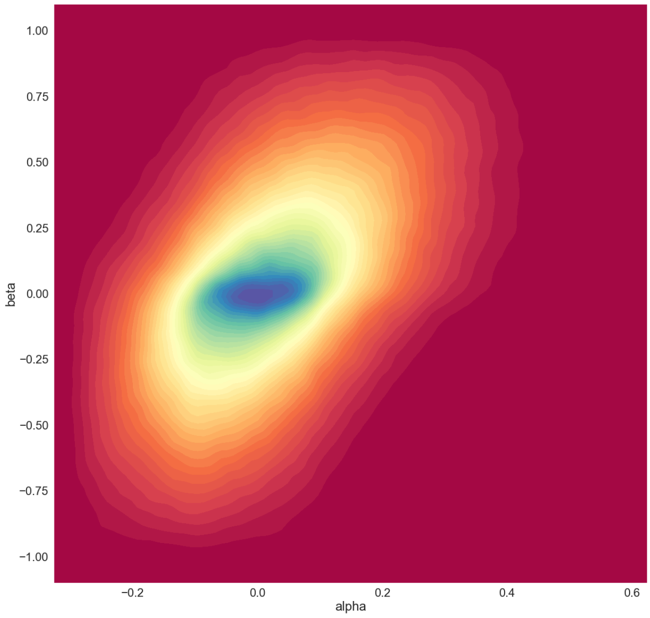


FIGURE 4.4: *Emprical invertibility region of the* EGAS-t

4.5 Issues in Maximum Likelihood Estimation

We now discuss some important issues of the behaviour of the maximum likelihood (ML) estimators. In particular, we show with some simulations the behaviour of the ML estimators in the case where there is clear non-invertibility, e.g. when α increases too much. To illustrate the problem, we assume that we observe a path of observations $\{\boldsymbol{\epsilon}_t\}$ for $t = 1, \dots, T$, generated by the model in (4.1) and (4.2), where the error

$$\boldsymbol{\eta}_t \sim \boldsymbol{\mathcal{N}}(\mathbf{0}_2, \boldsymbol{I}_2) \qquad or \qquad \boldsymbol{\eta}_t \sim \boldsymbol{t}_\nu(\mathbf{0}_2, \boldsymbol{I}_2).$$

Then the log-likelihood function for the bivariate *EGAS* model is given by

$$\ell_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln p(\boldsymbol{\epsilon}_t | \mathscr{F}_{t-1}), \tag{4.22}$$

where $\ln p(\boldsymbol{\epsilon}_t|\mathscr{F}_{t-1})$ can assume the form in (4.5) for the Gaussian case, and (4.8) for the Student's t. By plugging the filtered dynamic parameter $\{\exp \hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\}$ for $t = 1, \dots, T$ into (4.22) we obtain

$$\hat{\ell}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta}, \exp \boldsymbol{f}_1)) = \sum_{t=1}^{T} \ln p(\boldsymbol{\epsilon}_t | \exp \boldsymbol{f}_1, \mathscr{F}_{t-1}), \tag{4.23}$$

Then, the estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell_T(\boldsymbol{\theta}).$$

where we recall that $\boldsymbol{f}_1 = (f_{11}, f_{21})^\top$ is some fixed initial condition. Therefore, the invertibility property of the filters is crucial for the well-behaviour of the log-likelihood function. In what follows, we investigate the behaviour of the maximum likelihood estimator of the coefficients α and β.

4.5.1 Monte Carlo experiments

To study the effect of non-invertibility on the likelihood function, we consider a simple Monte Carlo simulation exercise based on 2000 replications with $T = 3000$, where we repeatedly generate time series from the bivariate *EGAS* model in (4.1) and (4.2), and estimate. Then we show the resulting performances and distributions of the ML estimators.

Firstly, the parameters α and β of the time-varying f_t, will be fixed on the boundary of the estimated invertibility regions, displayed in Figures 4.3 and 4.4. This implies that the filtering equations are likely to become highly unstable, since the perturbation due to the initial condition may not vanishes as $t \to \infty$.

Secondly, we choose empirically relevant values of the parameters α and β that properly satisfy the empirical invertibility condition, i.e. inside the estimated regions, implying that the started recursion are asymptotically stationary.

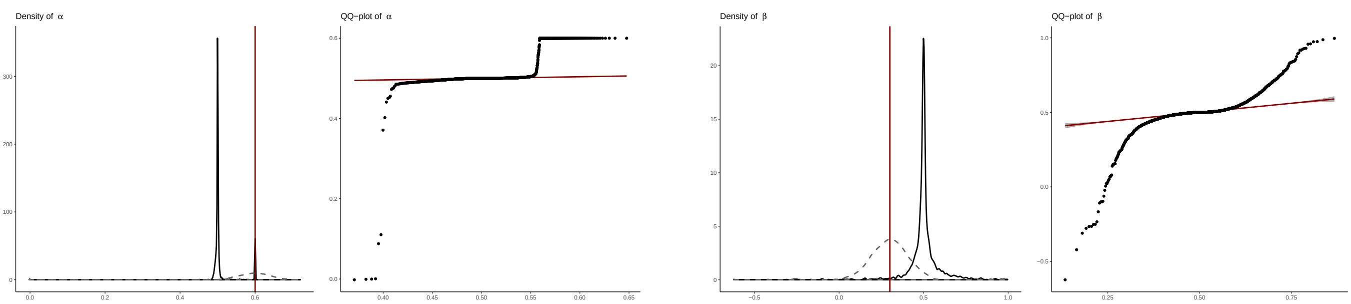
Gaussian Case

It is clear from the first line of result of Table 4.1 and graphs in Figure 4.5 that the behaviour of the ML estimators is totally unreliable, showing a substantial amount of bias, which suggest that the estimators are inconsistent. In addition, we note that α is firmly underestimated, while β is overestimated. These findings are in line with the fact that, without the invertibility property, the likelihood function is not sufficiently well-behaved, and therefore, neither the law of large numbers nor the asymptotic normality are available in that case.

Completely different is the scenario from the second line of of Table 4.1 and Figure 4.5, where the simulations confirm that, the invertibility of the filters positively affects the distribution of the MLEs. However, notice that there is still a little of bias in the estimates of β, since it tends to be always underestimated.

TABLE 4.1: *Bias and RMSE of the α and β in the Gaussian* EGAS *model. The estimators are ML based and we perform* 2000 *replications with* $T = 3000$. *The first line represent the result of the performance of the estimators when the parameters are on the boundary of the region in 4.3. The second line report the same results for an empirically relevant case.*

α			β		
Actual Value	*Bias*	*RMSE*	***Actual Value***	*Bias*	*RMSE*
0.600	0.091	0.099	0.300	-0.200	0.226
0.050	0.000	0.003	0.950	0.002	0.007



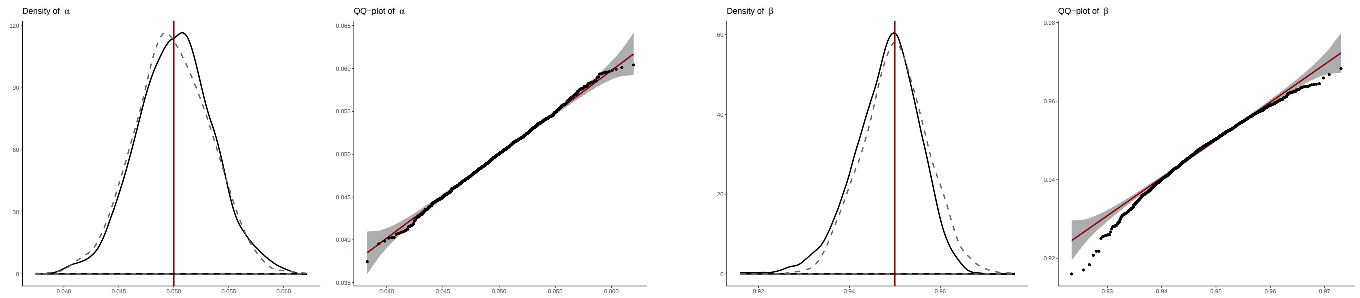


FIGURE 4.5: *Kernel density of the α and β in the Gaussian* EGAS *model with the specifications given in Table 4.1*

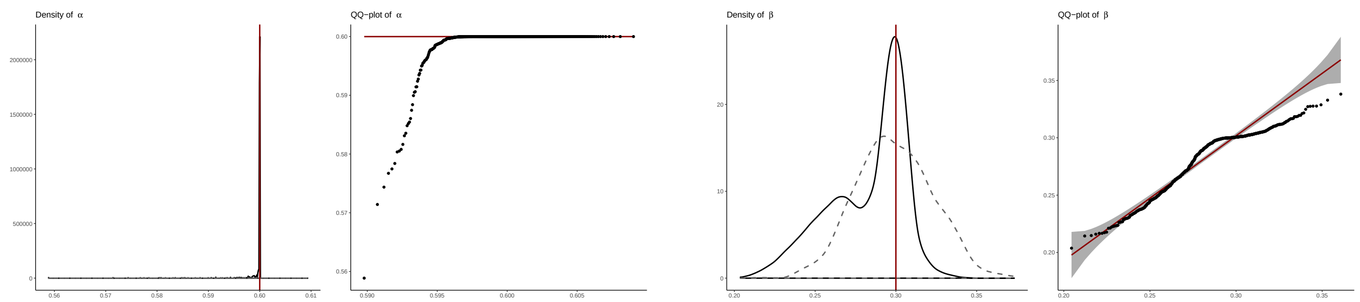
Student's *t* Case

In contrast to the result obtained for the Gaussian case, the results of the Monte Carlo simulation in the first line of Table 4.2 and Figure 4.6 shows a quite strange behaviour for the ML estimator of α. Indeed, when its actual value is on the boundary of the invertibility region, the ML estimator still delivers good results, since the bias is very low. However, during the simulation exercise we have noticed an high probability of the failure of the optimization routine, which means that the actual value may be completely missed. As a result the ML estimator is not reliable. A similar argument holds for the ML estimator of β, where the failure of its convergence is highlighted by the bimodality of the estimated density.

In line with the Gaussian specification, when the parameters are inside the invertibility region, the ML estimator of α for the *EGAS-t* is stable and provide strongly reliable results. Moreover, the second line Table 4.2 and Figure 4.6 confirm the small amount of positive bias, which was already founded in the *EGAS-g* for the estimates of β.

TABLE 4.2: *Bias and RMSE of the α and β in the Student's t* EGAS *model. The estimators are ML based and we perform* 2000 *replications with* $T = 3000$. *The first line represent the result of the performance of the estimators when the parameters are on the boundary of the region in 4.3. The second line report the same results for an empirically relevant case.*

α			β		
Actual Value	*Bias*	*RMSE*	***Actual Value***	*Bias*	*RMSE*
0.600	0.001	0.003	0.300	0.017	0.029
0.050	-0.001	0.005	0.950	0.003	0.009



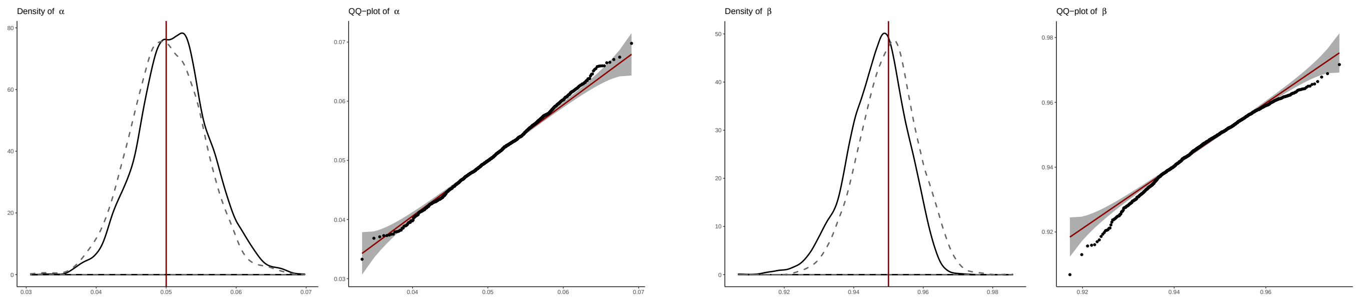


FIGURE 4.6: *Kernel density of the α and β in the Student's t* EGAS *model with the specifications given in Table 4.2*

4.6 Incorporating a Time-Varying Correlation

Consider again the framework given by the specifications in (4.1) and (4.2), but instead of a constant conditional correlation matrix $\boldsymbol{R}$, we allow to the conditional correlation coefficient to be dynamic. This means that we have to substitute the static $\boldsymbol{R}$ with its dynamic counterpart, that is

$$\boldsymbol{R}_t = \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix}.$$

Therefore, we need also to substitute in all the previous formulae, the constant conditional correlation coefficient ρ with the dynamic ρ_t. To retrieve its time evolution, it is necessary to specify a new score-driven filter. With the same approach used for modeling the volatilities f_{1t} and f_{2t}, we work with an unconstrained real valued sequence $\{\gamma_t\}_{t\in\mathbb{N}}$ and then apply a link function in order to ensure that the conditional correlation coefficients lies in the appropriate range, i.e. $\rho_t \in (-1, 1)$. In what follows, we adopt the Fisher transformation $\rho_t = \tanh \gamma_t$.

The new filter assumes the form

$$\gamma_{t+1} = \delta + \phi\gamma_t + \kappa u_t \qquad t \in \mathbb{N},$$

where $\delta \in \mathbb{R}$ is the intercept, $\phi \in \mathbb{R}$ is the autoregressive coefficient with the constrain that $|\phi| < 1$, and finally $\kappa \in \mathbb{R}$ which weight the intensity of the driving-force, u_t, that is the score. In the Appendix C.2 we report the analytical forumlae of the scores for both the models *EGAS-g* and *EGAS-t*, together with the required further derivatives (the Hessian matrices) needed in order to evaluate the empirical invertibility region discussed in Section 4.4.

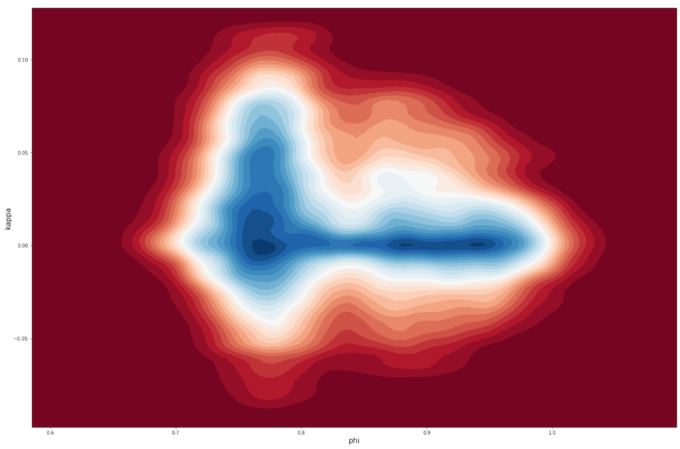


FIGURE 4.7: *Emprical invertibility region of the bivariate* EGAS *models with time-varying conditional correlation.*

4.7 Empirical Applications

In this section we apply the bivariate *EGAS* models to several time series and we shall compare the results between the Gaussian and Student's t specifications. We fit our bivariate class of score-driven models to different pairs of of stock returns, in order to show the most important feature of the considered model, such as volatility spillovers and the volatility interaction between the two stock price series. More importantly, the aim of this section is to show the usefulness of the adopted empirical approach in order to estimate the contraction of the estimated model. As i turns out, each of the estimated model satisfy the (empirical) contraction condition stated in (4.20).

We consider seven pairs of stock returns from different industries (i.e. MSFT/IBM, MSFT/AAPL, MSFT/AMZN, AAPL/IBM, AAPL/AMZN, IBM/AMZN, JPM/BAC). The time series extend from October 2, 2000 to December 31 2018, amounting in 4590 observations for each return series. We present the analysis by first showing the estimation results over the full samples. Then, we plot the filtered conditional volatilities paths for both the Gaussian and the Student's t specifications, in order to stress the crucial importance of the different driving force in the filtering equation. In particular, it will be clear that by changing the conditional distribution one can obtain tremendously different approximations of the underlying processes. For this reason it is important to conduct the analyses by the most suitable filter, which should be able to highlights the relevant properties of the observed time series.

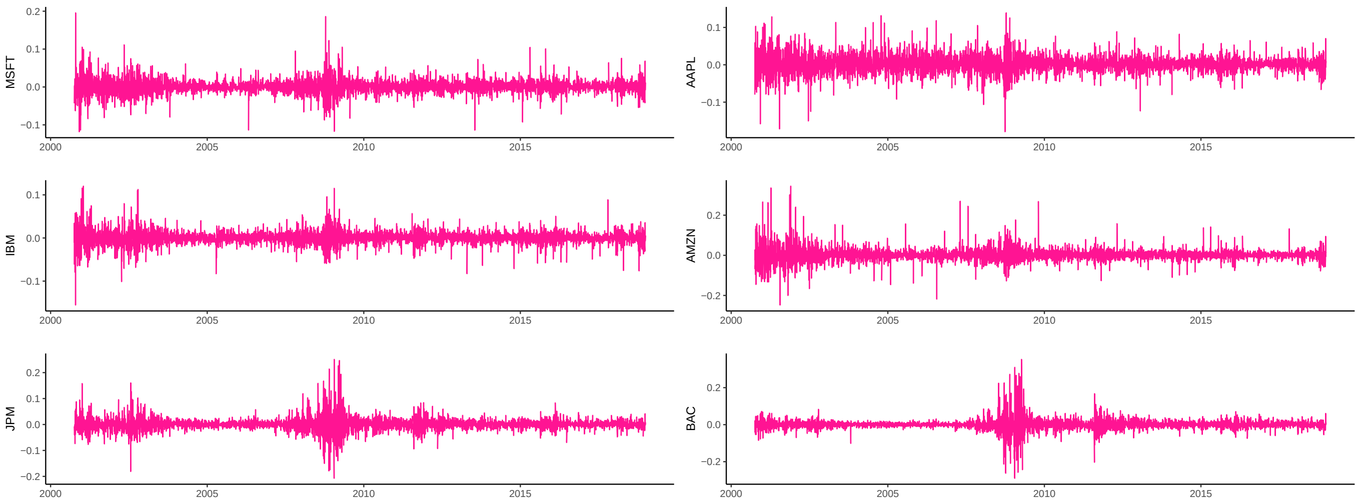


FIGURE 4.8: *Time series graph of the 6 stock returns for the whole sample period.*

In table 4.3, the in-sample maximum likelihood estimation results are displayed. In particular, the first six columns show the estimated parameters and their standard errors for each of the considered models with respect to the selected couple of returns. The seventh and eighth columns report some information about the likelihood function, namely the maximized log-likelihood and the value of the AIC, in order to compare the fitting performances between the models since the different number of estimated parameters. Lastly, in the rightmost column we have the resulting value of the empirical contraction condition, estimated by using the approach discussed in Section 4.4 above. As a diagnostic test, we apply the multivariate Portmanteau statistic, where we choose the order $h = 10$, see Lütkepohl, 2007. For each couple of returns and for both the models we report its estimated statistic together with the relative p-value. The following observations can be made.

First, as expected, the score-driven models with the Student's t specifications always outperform the Gaussian counterpart by means of the maximum log-likelihood obtained and the *AIC*. Hence, this confirm the heavy-tailed nature of the phenomena under study, indeed, the estimated degrees of freedom of the conditional distribution are quite low and almost identical.

Second, the estimates of α fall dramatically when the bivariate Gaussian is adopted, which means that this estimates may be severely biased toward zero, yielding an over estimation of the intercepts ω_1 and ω_2. This effect is slightly less evident for the estimates of β at first glance, but still an issue. However, from the estimates of the static correlation coefficients, it is hard to infer the implications of the corresponding models, since in some cases the MLE are markedly different.

Third, as regards the multivariate Portmanteau test for the estimated residuals we note that, except for the last three couples of returns, the *EGAS-g* always reject correct specification at 5%, thus, the specification test confirms the fact that the Gaussianity assumption might be not adequate for empirical applications, even if we allow for nonlinearities in the filter equations. Indeed, it is not surprising that a fat-tailed model, such as the *EGAS-t* volatility model, tends to deliver better performances when modeling financial returns.

Next, we filter the underlying volatility paths by using the score-driven bivariate system with the Gaussian and Student's t specifications which can be found in (C.3) and (C.6) respectively. We will apply these recursions to all the couple of financial returns. With this visual analysis, further evidences between the two specification could be acknowledged. In particular, it will be transparent the empirical advantages of a robust modeling procedure, indeed, it can be seen from Figure 4.9, the *EGAS-t* is able to mitigate the impact of the outliers and extreme values of the returns, yielding a reliable path of volatilities. On the other hand, the volatility paths filtered with the *EGAS-g* presents several spikes, due to the fact that an observation can have an arbitrary impact on the on f_{t+1}.

In Table 4.3 we report the comparison of the maximum likelihood estimates for the bivariate Gaussian and Student's t EGAS models and in parenthesis the corresponding standard errors. The last column report the values of contraction condition estimated with equation (4.20). The multivariate portmanteau test with order $h = 10$ is performed as a diagnostic test.

MSFT/IBM	ω_1	ω_2	α	β	ρ	ν	$\log L$	*AIC*	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.076	-0.078	0.011	0.982	0.422		-26,035.4	52,080.8	-0.081
	(0.008)	(0.008)	(0.000)	(0.002)	(0.013)				
Q_{10}	est	57.6	p-value	0.04					
EGAS-t	-0.030	-0.031	0.044	0.992	0.259	4.118	-23,657.1	**47,326.2**	-0.019
	(0.008)	(0.008)	(0.000)	(0.003)	(0.010)	(0.009)			
Q_{10}	est	**48.3**	p-value	**0.17**					

MSFT/AAPL	ω_1	ω_2	α	β	ρ	ν	$\log L$	*AIC*	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.041	-0.040	0.015	0.990	0.178		-23,836.5	47,683.0	-0.026
	(0.018)	(0.017)	(0.003)	(0.005)	(0.016)				
Q_{10}	est	89.6	p-value	0.00					
EGAS-t	-0.002	-0.002	0.033	0.999	0.217	4.492	-21,428.2	**42,868.4**	-0.014
	(0.005)	(0.005)	(0.004)	(0.006)	(0.013)	(0.008)			
Q_{10}	est	**85.8**	p-value	**0.05**					

MSFT/AMZN	ω_1	ω_2	α	β	ρ	ν	$\log L$	*AIC*	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.061	-0.051	0.011	0.986	0.195		-22,586.9	45,183.8	-0.022
	(0.007)	(0.006)	(0.001)	(0.002)	(0.015)				
Q_{10}	est	57.3	p-value	0.04					
EGAS-t	-0.002	-0.001	0.034	0.999	0.219	4.114	-20,781.9	**41,575.8**	-0.014
	(0.000)	(0.000)	(0.005)	(0.003)	(0.007)	(0.006)			

Q_{10}	est	**64.5**	p-value	**0.05**					

AAPL/IBM	ω_1	ω_2	α	β	ρ	ν	$\log L$	AIC	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.033	-0.036	0.011	0.992	0.242		-24,630.8	49,271.6	-0.017
	(0.009)	(0.003)	(0.001)	(0.001)	(0.0015)				
Q_{10}	est	56.8	p-value	0.04					
EGAS-t	-0.017	-0.019	0.037	0.995	0.200	4.526	-22,123.5	**44,259.3**	-0.015
	(0.000)	(0.001)	(0.005)	(0.007)	(0.010)	(0.008)			
Q_{10}	est	**57.6**	p-value	**0.06**					

AAPL/AMZN	ω_1	ω_2	α	β	ρ	ν	$\log L$	AIC	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.011	-0.010	0.006	0.997	0.185		-21,422.5	42,855.0	-0.007
	(0.008)	(0.000)	(0.000)	(0.000)	(0.015)				
Q_{10}	est	**51.4**	p-value	**0.11**					
EGAS-t	-0.001	-0.001	0.029	0.999	0.195	4.169	-19,334.2	**38,680.4**	-0.012
	(0.000)	(0.000)	(0.005)	(0.003)	(0.008)	(0.006)			
Q_{10}	est	**52.4**	p-value	**0.09**					

IBM/AMZN	ω_1	ω_2	α	β	ρ	ν	$\log L$	AIC	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.015	-0.012	0.006	0.996	0.306		-23,344.5	46,699.0	-0.008
	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)				
Q_{10}	est	**50.6**	p-value	**0.12**					
EGAS-t	-0.001	-0.001	0.035	0.999	0.186	4.069	-21,383.3	**42,778.6**	-0.015
	(0.000)	(0.001)	(0.002)	(0.003)	(0.007)	(0.006)			
Q_{10}	est	**56.6**	p-value	**0.05**					

JPM/BAC	ω_1	ω_2	α	β	ρ	ν	$\log L$	AIC	$\widehat{\Lambda(\theta)}$
EGAS-g	-0.014	-0.014	0.009	0.997	0.783		-25,951.6	51,913.2	-0.008
	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)				
Q_{10}	est	**45.3**	p-value	**0.27**					
EGAS-t	-0.011	-0.010	0.043	0.997	0.427	4.372	-22,411.6	**44,835.2**	-0.010
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)			
Q_{10}	est	**36.1**	p-value	**0.65**					

Estimation Period from 2000-10-01 to 2018-12-31 Length of each time series $T = 4590$

TABLE 4.3: *Maximum likelihood estimates for the bivariate EGAS models and multivariate Portmanteau test.*

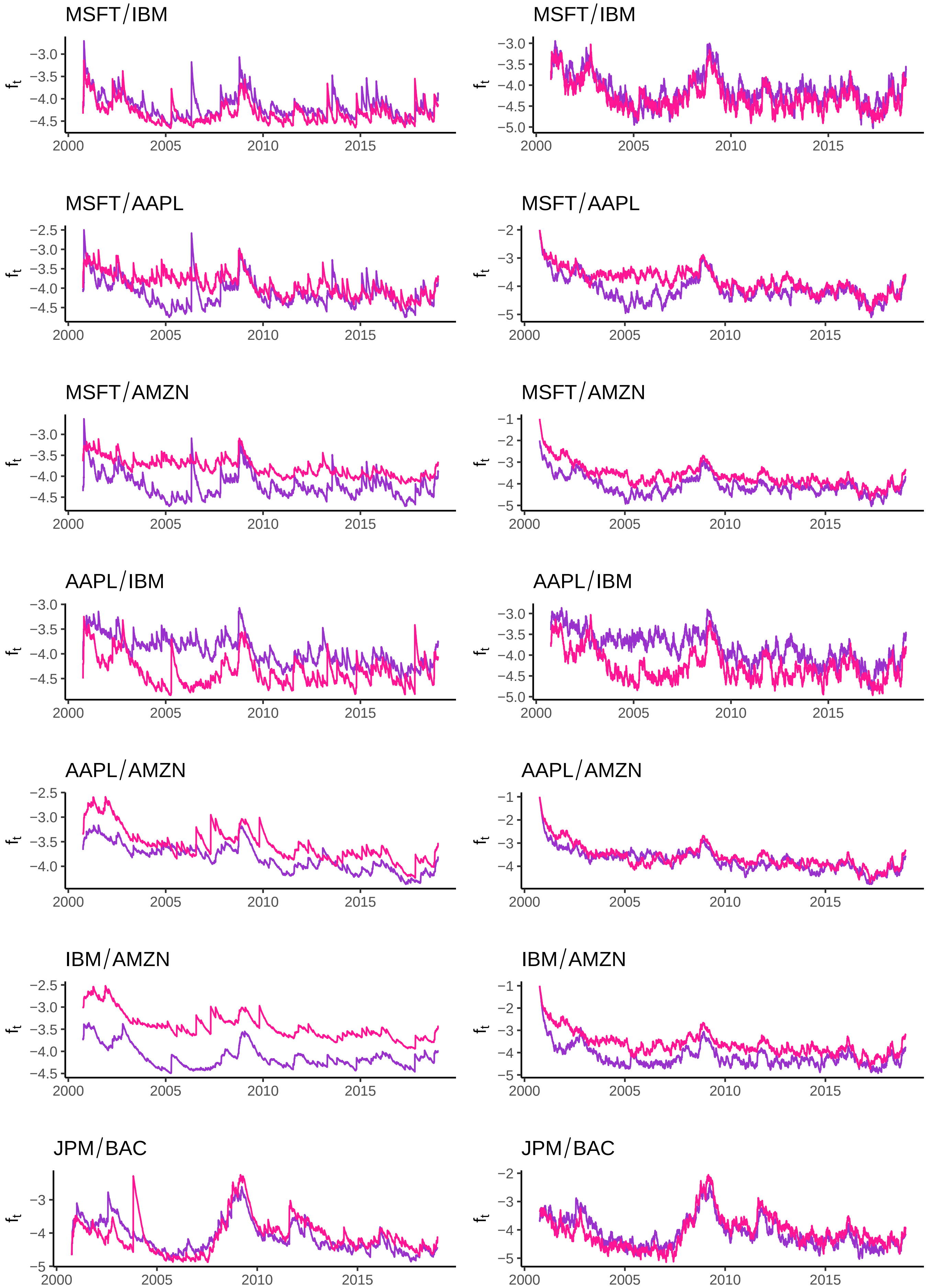


FIGURE 4.9: *Comparison of the filtered volatilities gathered from the Gaussian (LHS PANELS) and the Student's* t *(RHS PANELS) bivariate* EGAS.

The Q-test reported in Table 4.3 was done with squared standardized residuals. However, a score-based test is much preferable, because scores are related to the Lagrange multiplier test and incorporates information about the level of correlation, see Harvey and Thiele, 2016. The alternative test builds upon the Ljung-Box statistics described in section 3 of Harvey and Thiele, 2016. Then, in Table 4.4, we summarize the comparison between the two tests performed on the same panel of coupled time series. Some comments are in order.

TABLE 4.4: *Multivariate Ljung-Box Q Statistics and their p-values. With $Q(10)$ it is denoted the Q-test based on the standardized squared residuals, while with $Q_u(10)$ the score-based Q-test.*

	$Q(10)$		$Q_u(10)$	
	EGAS-g	*EGAS-t*	*EGAS-g*	*EGAS-t*
MSFT/IBM				
est.	57.6	48.3	239.0	211.0
p-value	0.04	0.17	0.00	0.00
MSFT/AAPL				
est.	89.6	85.8	89.0	89.0
p-value	0.00	0.05	0.00	0.04
MSFT/AMZN				
est.	57.3	64.5	61.0	79.0
p-value	0.04	0.05	0.02	0.01
AAPL/IBM				
est.	56.8	57.6	109.2	61.2
p-value	0.04	0.06	0.00	0.02
AAPL/AMZN				
est.	51.4	52.4	165.6	165.0
p-value	0.11	0.09	0.00	0.00
IBM/AMZN				
est.	50.6	56.6	205.2	165.0
p-value	0.12	0.05	0.00	0.00
JPM/BAC				
est.	45.3	36.1	214.3	67.57
p-value	0.27	0.65	0.00	0.00

From the results collected in Table 4.4 it is possible to see that the estimated values of the Ljung-Box statistics are quite different. The multivariate Ljung-Box Q test based on the squared standardized residuals it is often accepted for both the Gaussian and Student's t specification, whereas it is always rejected when the equivalent score-based test is considered. Therefore, this may be seen as a further evidence of the fact that the scores incorporate information about the correlation structure between the time series which are not accounted when using the cross-product of standardized residuals.

4.7.1 Volatility Filtering with the Bivariate *EGAS*

To illustrate in more details the Bivariate *EGAS* models and stress the importance of the invertibility condition for empirical applications, we now focus on the bivariate time series of returns composed by the tickers JPM/BAC, displayed in Figure 4.10.

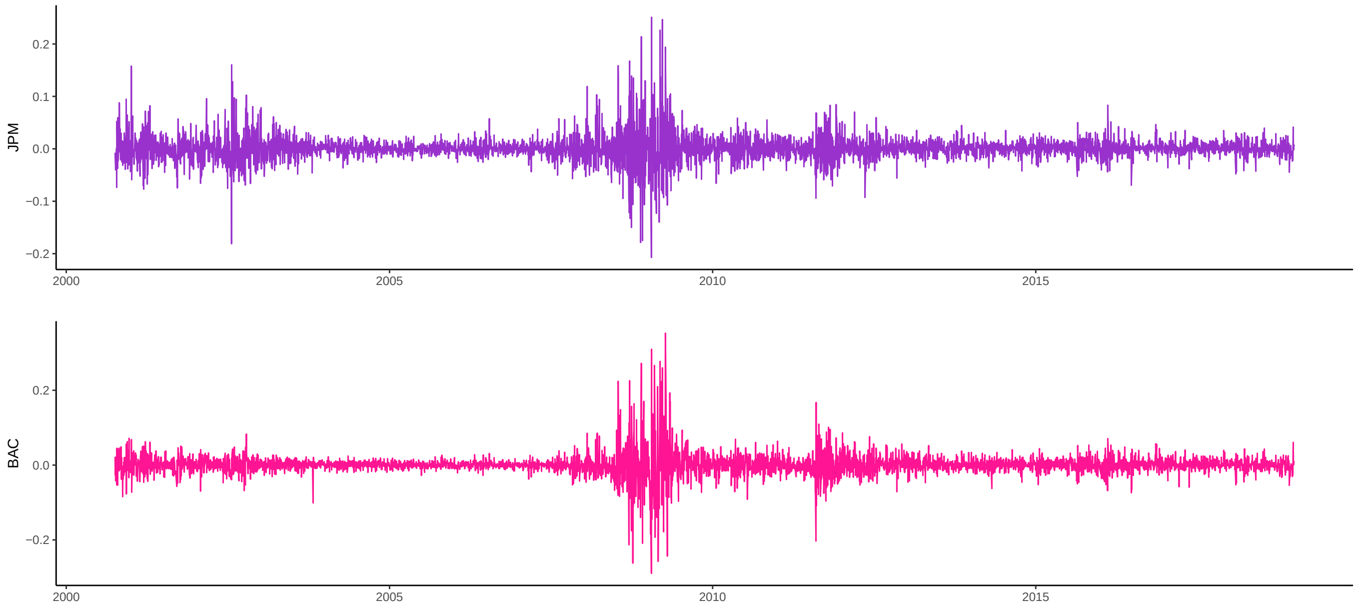


FIGURE 4.10: *Time plots of daily demeaned returns from* 2000-10-01 *to* 2018-12-31.

The reason behind the choice of these tickers lies in the fact that they are highly correlated, showing a very similar pattern of returns. Moreover, they also present the usual stylized facts of financial returns, such as volatility clustering, heavy-tails and tail-dependence. Table 4.5 provides summary statistics of the two time series of returns, which shows that the distribution of both the returns are slightly positively skewed, highly leptokurtic and the p-value of the Jarque Bera Test confirm that the normality assumption is strongly rejected.

TABLE 4.5: *Summary statistics of the returns.*

	Mean	*Std. Dev.*	*Skewness*	*Kurtosis*	*Jarque Bera Test*
JPM	-0.000	0.022	0.914	16.24	0.000
BAC	0.000	0.034	0.927	28.37	0.000

We also document the results of the summary statistics with Figures 4.11 and 4.12 which show the bivariate kernel density estimation of the returns and the respective contour plot.

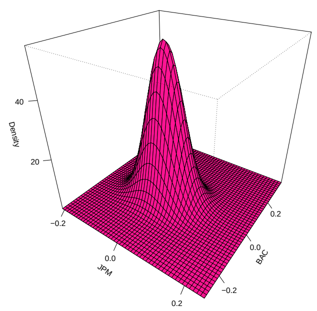


FIGURE 4.11: *Bivariate Kernel Density Plot of the returns.*

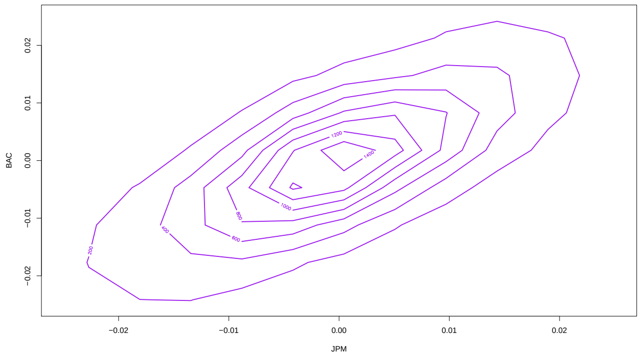


FIGURE 4.12: *Contour Plot of the returns.*

The dependence structure between the tickers can be even more appreciated by looking at Figures 4.13 and 4.14, where two different type of scatter plots are displayed. From these plots it is possible to see the presence of positive tail-dependence.

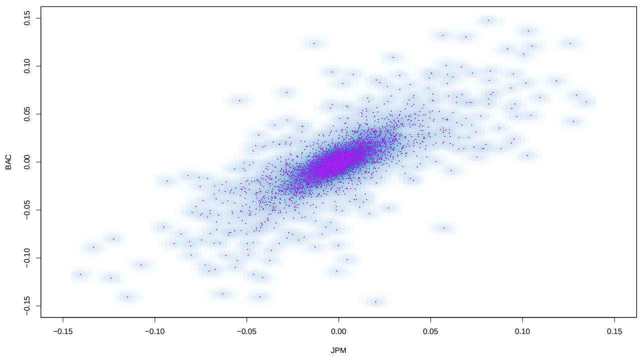


FIGURE 4.13: *Scatter Plot of the returns.*

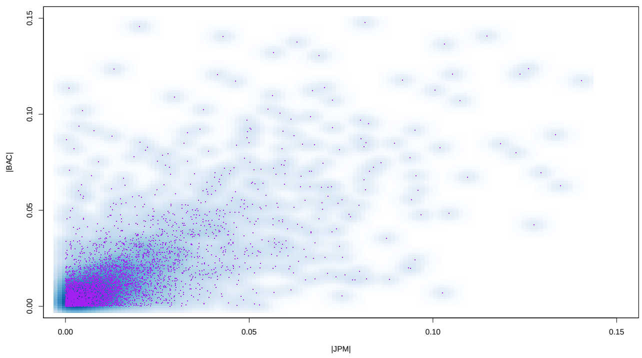


FIGURE 4.14: *Scatter Plot of the absolute returns.*

We now turn to the study of the estimated model for the pair of returns. Specifically, from the results of Table 4.3 one can see that the bivariate *EGAS* with the Student's t specification outperform the Gaussian model, since we can achieve better fitting performances in terms of the likelihood evaluated at the estimates and the AIC. Furthermore, the evidences provided by the descriptive analysis suggest that a fat-tailed model may be a proper choice to model the couple of time series returns.

Thus, to model the daily returns, we consider the estimated version of the *EGAS-t* model, defined with the framework given in (4.1) and (4.2), where the distribution of the *IID* random variables $\boldsymbol{\eta}_t \sim \boldsymbol{t}_{4.372}(\mathbf{0}_2, \boldsymbol{I}_2)$, the constant conditional correlation coefficient $\hat{\rho} = 0.427$ and the estimated score-driven filter is given by the recursions

$$\begin{bmatrix} f_{1,t+1} \\ f_{2,t+1} \end{bmatrix} = \begin{bmatrix} -0.011 \\ -0.010 \end{bmatrix} + 0.043 \begin{bmatrix} s_{1t} \\ s_{2t} \end{bmatrix} + 0.997 \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}, \tag{4.24}$$

where the driving-force $\boldsymbol{s}_t = (s_{1t}, s_{2t})^\top$ is as in the recursions of equation (C.6). We note that the estimated autoregressive parameter $\hat{\beta} = 0.997$, which is very close to unity, and as a consequence the volatility series of the two tickers may exhibit *IGARCH* behaviour.

In Table 4.3 it was also shown that the estimate of the empirical version of the Bougerol's contraction condition stated in (4.20) is $\widehat{\boldsymbol{\Lambda}(\boldsymbol{\theta})} = -0.010 < 0$ satisfying the invertibility condition. Thus, to better understand what are the implication of this result, we filter the underlying volatility from the pair of returns JPM/BAC with the filters in (4.24). Since the starting values are unknown, one may start the recursions from several arbitrary fixed point $\bar{f} = (\bar{f}_1, \bar{f}_2)^\top \in \mathbb{R}^2$.

Figure 4.15 provide the empirical evidence of the fact that our estimated version of the invertibility region of the parameter space for the filters in f_t is reasonably large. Therefore, maximum likelihood estimations carried out directly on the region in (4.21) provides an empirically useful nonlinear and invertible volatility filter. The latter result may also suggest that the likelihood function in equation (4.23) is sufficiently well-behaved such that it allows for appropriate law of large numbers and asymptotic normality of the (restricted) maximum likelihood estimator.

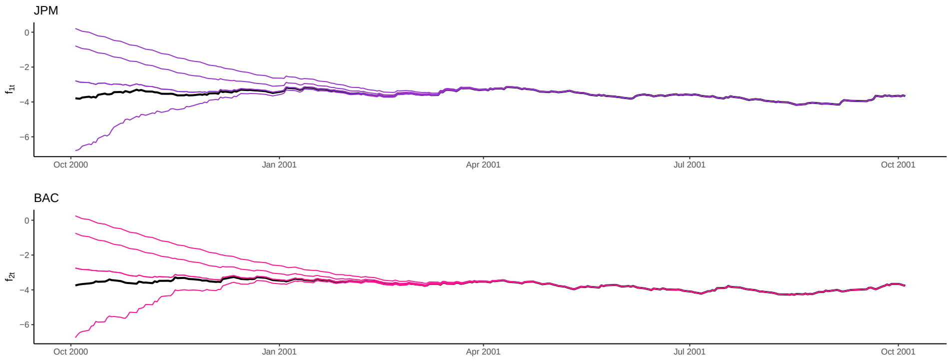


FIGURE 4.15: *In purple and pink, it is shown the paths of the first* 250 *filtered volatilities from the pair of returns JPM/BAC with the invertible* EGAS-t. *The recursion has been started with arbitrary fixed values* $\bar{f} = (\bar{f}_1, \bar{f}_2)^\top \in \mathbb{R}^2$. *In black we have their respective stationary and ergodic solution*

Thus, in Figure 4.16 we also report the estimated conditional volatilities for the whole samples of daily demeaned-return in Figure 4.10, which extends from 2000-10-01 to 2018-12-31. Not surprisingly, the estimated conditional volatilities show a very similar path, with common upward and downward movement.

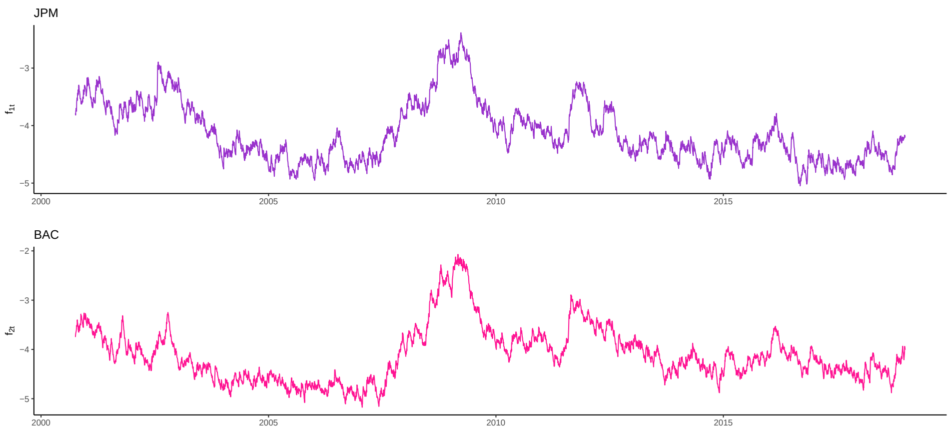


FIGURE 4.16: *In purple and pink, estimated conditional volatilities from the pair of returns JPM/BAC with the* EGAS-t *in equation* (4.24).

Therefore, we analyze this relationship between the series by adopting the methods proposed by Gallant, Rossi, and Tauchen, 1993, see also Herwartz and Lütkepohl, 2000 for an application to multivariate volatility analysis. The estimated impulse response function are displayed in Figure 4.17. The first property that one could note is that the persistence which characterize both the series is remarkably high. Indeed, when a shock hits in the two variables hits the system, there will be a substantial increase in volatility. Interestingly, the covariances reacts in a different way. When a shock in JPM hits BAC, leads to a decreasing in volatility, while when an impulse on BAC hits JPM, yields a small amount of increase in volatility. Hence, JPM innovations have a stronger impact on BAC than BAC on JPM. To conclude this part, we stress the fact that this asymmetric spillover effect may be well handled by our score-driven model, since it allows for *ARCH* as well as *GARCH* spillover effects in the filtering recursions.

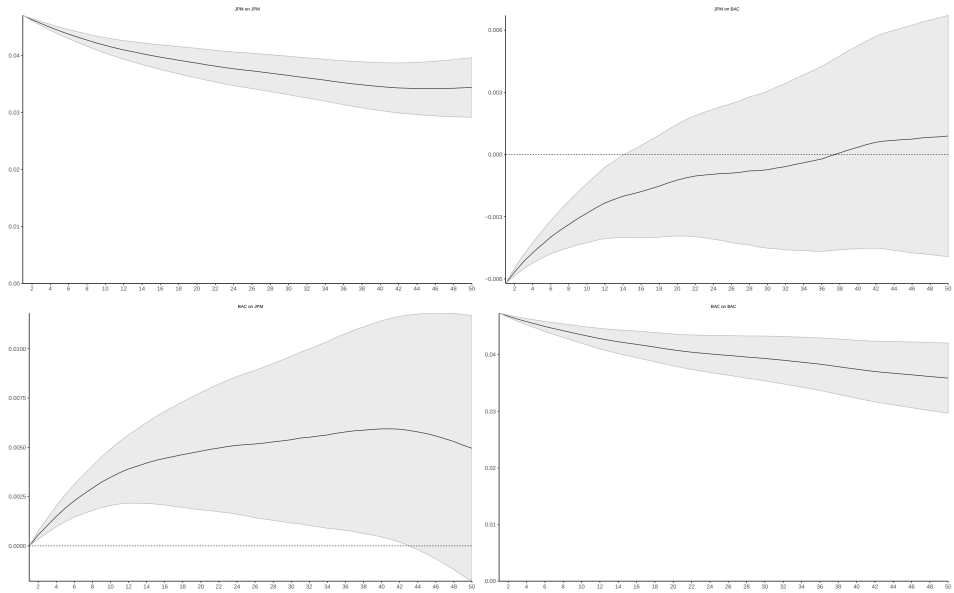


FIGURE 4.17: *Plots of Impulse Responce Functions of standard deviation shock.*

4.8 Concluding Remarks

In this paper we have introduced a new robust multivariate score driven model for conditional volatilities. We further used a multivariate version of the empirical method proposed in Blasques et al., 2018 to verify that the invertibility condition holds and ensure that the estimation and filtering results are not spurious.

With a Monte Carlo exercise we have documented the crucial importance of the invertibility condition for the well-behaviour of the log-likelihood function and the ML estimators.

In our empirical application we considered seven pairs of stock returns from different industries and concluded that filter invertibility holds on all these data set using our empirical method. This allows us to use these bivariate robust filters with confidence that the estimation and filtering results are sound. With a particular focus on the the tickers JPM/BAC we have provided further evidences of the ability of our model in handle heavy-tailed financial returns and capture spillover effects.

Appendix A

Proofs of Chapter 2

A.1 Main Proofs

A.1.1 Proofs of Lemmata for the Dynamic Location

Proof of lemma 2.2.1

Proof. The fact that $b_t \sim \mathcal{B}eta\left(\frac{N}{2}, \frac{\nu}{2}\right)$ follows trivially from the properties of the multivariate Student's t, see for instance Pag. 19 of Kotz and Nadarajah, 2004 or Proposition 39 of Harvey, 2013. Thus, it can be expressed as the ratio random variable $b_t = r_t/(s_t + r_t)$, where $r_t \sim \chi^2_N$ is independent of $s_t \sim \chi^2_\nu/\nu$. Now it easy to see that it is possible to rewrite $\boldsymbol{u}_t$ in (2.4) as $\boldsymbol{u}_t = \sqrt{\nu}\sqrt{b_t(1-b_t)}\boldsymbol{\Omega}^{1/2}\mathbf{u}_t$. The random vector $\mathbf{u}_t$ and the random variables b_t are independent of each other by construction, see Fang, Kotz, and Ng, 1990, and thus for even integers $m \geq 0$, the moments of $\boldsymbol{u}_t$ can also be expressed in terms of a one-dimensional integral, that are

$$
\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{u}_t\|^m\right] =& \nu^{m/2}\|\boldsymbol{\Omega}\|^{m/2}\mathbb{E}\left[b_t^{m/2}(1-b_t)^{m/2}\right]\mathbb{E}\left[\|\mathbf{u}_t\|^m\right] \\
=& \frac{\|\boldsymbol{\Omega}\|^{m/2}}{\mathcal{B}\left(\frac{N}{2}, \frac{\nu}{2}\right)}\left(\frac{\nu}{N}\right)^{m/2} \\
& \times \int b_t^{\frac{N+m}{2}-1}(1-b_t)^{\frac{\nu+m}{2}-1}\mathrm{d}b_t \\
=& \|\boldsymbol{\Omega}\|^{m/2}\left(\frac{\nu}{N}\right)^{m/2}\frac{\mathcal{B}\left(\frac{N+m}{2}, \frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2}, \frac{\nu}{2}\right)}.
\end{aligned}
$$

Notice that, with this Lemma, we also satisfy the Sufficient Condition for Robustness, see Proposition 1 of Calvet, Czellar, and Ronchetti, 2015. □

Proof of lemma 2.3.1

Proof. The stability condition $\varrho(\boldsymbol{\Phi}) < 1$, is a well-known condition in the theory of linear systems, see Hannan and Deistler, 1987, Hannan, 1970 or Lütkepohl, 2007. This condition, however, also extends to our nonlinear model. This can be seen from the fact that if we consider the recursion of the dynamic location as a function of the innovations like in (2.6). This recursion is linear in $\boldsymbol{\mu}_t$ for a given $\boldsymbol{\epsilon}_t$ and hence the condition needed for the process to produce stationary ergodic paths of the former boils down to the familiar stability condition $\varrho(\boldsymbol{\Phi}) < 1$.

As regards the bounded moments, consider (2.6) such that, by combining Lemma 2.2.1, we take the expectation and use the triangle, Hölder and Minkowsky inequalities such that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|\boldsymbol{\mu}_{t+1}-\boldsymbol{\omega}\|^m\right] = \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\sum_{j=0}^{\infty}\boldsymbol{\Phi}^j\boldsymbol{K}\boldsymbol{u}_{t-j}\right\|^m\right] \leq \mathbb{E}\left[\sum_{j=0}^{\infty}\sup_{\theta\in\Theta}\|\boldsymbol{\Phi}^j\boldsymbol{K}\boldsymbol{u}_{t-j}\|^m\right]$$
$$\leq \left\{\bar{c}\sum_{j=0}^{\infty}\bar{\rho}^j\left(\mathbb{E}\left[\|\boldsymbol{u}_{t-j}\|^m\right]\right)^{1/m}\right\}^m < \infty,$$

where here and elsewhere we have that $\bar{c} = \sup_{\theta\in\Theta} N\|\boldsymbol{K}\|$ and $\bar{\rho} = \varrho(\boldsymbol{\Phi}) < 1$. Clearly, the last inequality is a standard result from linear algebra, in fact a simple eigendecomposition can show that $\|\boldsymbol{\Phi}\| = \|\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\| = \text{tr}(\boldsymbol{\Lambda}) = \sum_{i=1}^N \rho_i$ where ρ_i are the eigenvalues of $\boldsymbol{\Phi}$. □

Proof of Lemma 2.3.2

Proof. The desired result follows trivially since

$$\mathbb{E}[\|\boldsymbol{y}_t\|^m] \leq \bar{C}_1\mathbb{E}[\|\boldsymbol{\mu}_t\|^m] + \bar{C}_1\mathbb{E}[\|\boldsymbol{\epsilon}_t\|^m] < \infty,$$

by the c_r-inequality. $\boldsymbol{\mu}_t$ is uniformly bounded from Lemmata 2.2.1 and 2.3.1 and by the properties of the multivariate Student's t. □

Proof of lemma 2.3.3

Proof. The contraction condition 2 follows as a straightforward modification of Theorem 6.12 in Pötscher and Prucha, 1997, where the authors show that the aforementioned sufficient condition may be obtained by applying the mean value theorem to the Lipschitz map of the dynamic location vector. Then we note that, in our model this condition boils down to the following sufficient condition

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\prod_{k=1}^{j}\frac{\partial\boldsymbol{\mu}_{t-k+1|t-k}}{\partial\boldsymbol{\mu}_{t-k|t-k-1}^\top}\right\|\right] = \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\prod_{k=1}^{j}\boldsymbol{X}_{t-j}\right\|\right] < 1.$$

This condition can be derived as follows. Consider the dynamic equation in terms of the observations $\boldsymbol{y}_t \in \mathbb{R}^N$

$$\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega} = \boldsymbol{\Phi}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})}{1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu}.$$

This recursion can be embedded in a first order nonlinear dynamic system

$$\boldsymbol{\mu}_{t+1|t} = \phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}), \qquad t \in \mathbb{N},$$

where we suppse that the dynamic location vector takes his values in a Borel subset $\boldsymbol{\mathcal{M}}$ of $\mathbb{R}^N$. This allow us to define inductively for $k \geq 1$ and any initialization $\boldsymbol{\mu}_{1|0} \in \boldsymbol{\mathcal{M}}$, a sequence of Lipschitz maps $\phi^{(k+1)} : \boldsymbol{\mathcal{M}} \times \mathbb{R}^N \times \boldsymbol{\Theta} \mapsto \boldsymbol{\mathcal{M}}$ for $k \geq 1$ such that

$$\phi^{(k+1)}(\boldsymbol{\mu}_{1|0}, \boldsymbol{y}_1, \dots, \boldsymbol{y}_{k+1}, \boldsymbol{\theta}) = \phi\left(\phi^{(k)}(\boldsymbol{\mu}_{1|0}, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k, \boldsymbol{\theta}), \boldsymbol{y}_{k+1}, \boldsymbol{\theta}\right).$$

Thus, following Pötscher and Prucha, 1997, we apply the *mean value theorem* to the Lipschitz map $\phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})$ and see that

$$\phi(\hat{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) - \phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) = \phi'(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}),$$

from which we can recover

$$\boldsymbol{\mu}_{t+1|t} = \phi'(\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})\boldsymbol{\mu}_{t|t-1} + \phi(\hat{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) - \phi'(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})\hat{\boldsymbol{\mu}}_{t|t-1}, \tag{A.1}$$

since $\boldsymbol{\mu}_{t+1|t} = \phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})$ and where $\boldsymbol{\mu}^{\star}_{t|t-1}$ is on the segment connecting $\hat{\boldsymbol{\mu}}_{t|t-1}$ and $\boldsymbol{\mu}_{t|t-1}$. Furthermore, $\phi'(\cdot)$ denotes the first partial derivatives of $\phi(\cdot)$ with respect to the transpose of $\boldsymbol{\mu}^{\star}_{t|t-1}$, yielding a $N \times N$ random matrix.

A more convenient way to represent (A.1) could be achieved by noting that

$$\boldsymbol{\mu}_{t+1|t} = \widehat{\boldsymbol{X}}^{\star}_t \boldsymbol{\mu}_{t|t-1} + \varphi(\hat{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}), \tag{A.2}$$

with $\widehat{\boldsymbol{X}}^{\star}_t = \phi'(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})$ and $\varphi(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) = \phi(\hat{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) - \phi'(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})\hat{\boldsymbol{\mu}}_{t|t-1}$. Equation (A.2) is a multivariate stochastic recurrence equation (MSRE), that can be viewed as vectorial autoregressive process with random coefficients $\{\widehat{\boldsymbol{X}}^{\star}_t\}$ and $\{\varphi(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})\}$. With this representation, it is possible to backsolve the recursion in order to obtain

$$\boldsymbol{\mu}_{t+1|t} = \left(\prod_{i=0}^{t-1} \widehat{\boldsymbol{X}}^{\star}_{t-i}\right) \boldsymbol{\mu}_{1|0} + \varphi(\boldsymbol{\mu}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) + \sum_{j=1}^{t-1} \left(\prod_{i=0}^{j} \widehat{\boldsymbol{X}}^{\star}_{t-i}\right) \varphi(\boldsymbol{\mu}^{\star}_{t-j|t-j-1}, \boldsymbol{y}_{t-j}, \boldsymbol{\theta}),$$

for $t \in \mathbb{N}$ and where $\boldsymbol{\mu}_{1|0}$ is some fixed random vector used for starting the process.

Now, there is the second condition that remains to be verified from the Pötscher and Prucha, 1997's Theorem 6.12, that is, the uniform integrability at the starting point of the dynamic system, but this is trivially satisfied since

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta \in \Theta} \|\boldsymbol{X}_1\|\right] &= \mathbb{E}\left[\sup_{\theta \in \Theta} \|\boldsymbol{\Phi} + 2(1-b_1)^2/\nu \boldsymbol{K}(\boldsymbol{y}_1 - \boldsymbol{\mu}_{1|0})(\boldsymbol{y}_1 - \boldsymbol{\mu}_{1|0})^{\top}\boldsymbol{\Omega}^{-1} - (1-b_1)\boldsymbol{K}\|\right] \\
&\le \bar{\rho} + c_K\left\{2\mathbb{E}[b_1(1-b_1)]\mathbb{E}\left[\|\mathbf{u}_1\|^2\right] + \mathbb{E}[(1-b_1)]\right\} \\
&\le \bar{\rho} + c_K \frac{\nu(\nu+N+4)}{(\nu+N)(\nu+N+2)} < \infty,
\end{aligned}$$

and moreover, note that $\mathbb{E}\left[\sup_{\theta \in \Theta} \varphi(\boldsymbol{\mu}^{\star}_{1|0}, \boldsymbol{y}_1, \boldsymbol{\theta})\right] = \mathbb{E}\left[\sup_{\theta \in \Theta} \|\boldsymbol{R}_1\|^m\right] < \infty$ from Lemma 2.2.1, 2.3.1 and A.3.5.

Thus, by recursive arguments we obtain

$$\begin{aligned}
\sup_{\theta \in \Theta} \|\hat{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\mu}_{t+1|t}\| &= \sup_{\theta \in \Theta} \left\| \left(\prod_{i=0}^{t-1} \widehat{\boldsymbol{X}}^{\star}_{t-i}\right) \left\{\hat{\boldsymbol{\mu}}_{1|0} - \boldsymbol{\mu}_{1|0}\right\} \right\| \\
&\le \varrho^t c_1,
\end{aligned}$$

where $c_1 > 0$ and $0 < \varrho < 1$ are constants.

Finally, the exponentially fast almost sure convergence of the filtered $\{(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})\}_{t \in \mathbb{N}}$ may be obtained as an application of Theorem 3.1 in Bougerol, 1993 or Theorem 2.7 in Straumann and Mikosch, 2006, that is,

$$\sup_{\theta \in \Theta} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}\| \xrightarrow{\text{e.a.s.}} 0 \qquad as \qquad t \to \infty, \tag{A.3}$$

for any initialization of the filtering recursion, since the contraction condition 2 and the verified integrability at some fixed value $(\mu_{1|0} - \omega)$ are more than enough to satisfy the requirement of Bougerol, 1993's Theorem. Hence, the desired result follows. □

A.1.2 Proof of Consistency and Asymptotic Normality of the MLE

To avoid confusions, we also define the empirical average log-likelihood function based on the chosen initial value $\mu_{1|0}$ and then the recursion $\hat{\mu}_{t|t-1}$

$$\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\ell}_t(\boldsymbol{\theta}), \tag{A.4}$$

and the likelihood based on the stationary solution $\mu_{t|t-1}$

$$\mathcal{L}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}), \tag{A.5}$$

with the following limit

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\ell_t(\boldsymbol{\theta})]. \tag{A.6}$$

Proof of Consistency

Proof. The proof of strong concistency builds up the following steps. We have,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})|.$$

By Lemma A.3.3

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})| \xrightarrow{\text{a.s.}} 0 \qquad \textit{as} \qquad t\to\infty,$$

and by Lemma A.3.4

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| \xrightarrow{\text{a.s.}} 0 \qquad \textit{as} \qquad t\to\infty.$$

Also, by the Ergodic Theorem

$$\lim_{T\to\infty}\widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0) = \lim_{T\to\infty}\mathcal{L}_T(\boldsymbol{\theta}_0) = \mathcal{L}(\boldsymbol{\theta}_0),$$

and in conclusion, by Lemma A.3.2

$$\mathcal{L}(\boldsymbol{\theta}) < \mathcal{L}(\boldsymbol{\theta}_0) \qquad \textit{for any} \qquad \boldsymbol{\theta}\neq\boldsymbol{\theta_0}.$$

Following similar arguments of Theorem 3.4 in White, 1994, we can show that strong consistency holds if $\forall\ \boldsymbol{\theta}\neq\boldsymbol{\theta}_0$, $\exists\ \mathcal{B}_\eta(\boldsymbol{\theta})$, where $\mathcal{B}_\eta(\boldsymbol{\theta}) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| > \eta, \eta > 0\}$ s.t. $\forall$ sequence of maximizers $\{\boldsymbol{\theta}^\star\}\in\boldsymbol{\Theta}$ and $\boldsymbol{\theta}^\star\in\mathcal{B}_\eta(\boldsymbol{\theta})$,

$$\limsup_{T\to\infty}\sup_{\boldsymbol{\theta}^\star\in\mathcal{B}_\eta(\boldsymbol{\theta})}\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) < \lim_{T\to\infty}\widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0) \qquad \textit{almost surely.}$$

Thus, similar argument as before, the reverse Fatou's Lemma and the Ergodic Theorem we get

$$\limsup_{T\to\infty} \sup_{\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})} \widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = \limsup_{T\to\infty} \sup_{\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})} \mathcal{L}_T(\boldsymbol{\theta}) = \limsup_{T\to\infty} \sup_{\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})} \frac{1}{T}\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta})$$
$$\leq \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})} \ell_t(\boldsymbol{\theta}) = \mathbb{E}\left[\sup_{\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})} \ell_t(\boldsymbol{\theta})\right],$$

and therefore, $\forall\, \varepsilon > 0\ \exists\, \eta > 0$ s.t.

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})} \ell_t(\boldsymbol{\theta})\right] < \mathbb{E}\Big[\ell_t(\boldsymbol{\theta})\Big] + \varepsilon = \mathcal{L}(\boldsymbol{\theta}) + \varepsilon.$$

Note that ε can be made arbitrarily small, therefore, strong consistency follows by the compactness of the parameter space $\boldsymbol{\Theta}$ and the identifiability uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, ensured by Lemma A.3.2. □

Proof of Asymptotic Normality

Proof. Standard arguments for the proof of asymptotic normality and the Taylor's theorem lead to the expansion of the conditional likelihood's score function around a neighborhood of $\boldsymbol{\theta}_0$, which yields

$$\begin{aligned}
\mathbf{0} = \sqrt{T}\mathcal{L}_T'(\hat{\boldsymbol{\theta}}_T) =& \sqrt{T}\Big[\widehat{\mathcal{L}}_T'(\boldsymbol{\theta}_0) - \mathcal{L}_T'(\boldsymbol{\theta}_0)\Big] + \sqrt{T}\mathcal{L}_T'(\boldsymbol{\theta}_0) \\
&+ \Big[\Big(\mathcal{L}_T''(\boldsymbol{\theta}_0) - \mathcal{L}''(\boldsymbol{\theta}_0)\Big) + \Big(\widehat{\mathcal{L}}_T''(\boldsymbol{\theta}^\star) - \mathcal{L}_T''(\boldsymbol{\theta}_0)\Big) + \mathcal{L}''(\boldsymbol{\theta}_0)\Big] \\
&\times \Big[\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)\Big],
\end{aligned} \tag{A.7}$$

where $\boldsymbol{\theta}^\star$ lies on the chord between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$, componentwise.

Considering equation (B.31), the convergence of the first difference in square brackets is ensured by Lemma A.3.10 whereas the fact that $\sqrt{T}\mathcal{L}_T'(\boldsymbol{\theta}_0)$ obeys the CLT for martingales is entailed in Lemma A.3.9. As regards the second line we have that the middle term vanishes almost surely and exponentially fast, since Lemma A.3.12 demonstrates that the initial conditions for the likelihood's second derivative are asymptotically irrelevant and the consistency theorem further ensures the convergence in the same point by continuity arguments of the likelihood's second derivatives. In addition, the first term in the brackets of the second line vanishes as well by the uniform law of large numbers discussed in Lemma A.3.13. Finally, with Lemma A.3.11 at hand, we can easily solve equation (B.31), since $\mathcal{L}''(\boldsymbol{\theta}_0)$ is nonsingular.

Thus, the Slusky's Lemma (see Lemma 2.8 *(iii)* of Vaart, 1998) completes the proof. □

A.2 Computational Aspects

This Appendix is devoted to the construction of score vector and the Hessian matrix, which are needed for estimation and inference. Our approach to tackle this problem follows the matrix differential calculus style of Magnus and Neudecker, 2019. As argued by the authors, one of the advantages to represent the conditional log-density in its differential form is that we can straightforwardly retrieve all the partial derivatives, thus avoiding the problem of dealing with the dimensions of the matrices and vectors involved.

A.2.1 The Score Vector

To construct the score vector, we take the first differential of the likelihood function (2.8), and get

$$\begin{aligned}\mathrm{d}\ell_t(\boldsymbol{\theta}) =& \frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{N}{\nu}+\frac{\nu+N}{\nu}b_t-\ln w_t\right](\mathrm{d}\nu)\\ &+\frac{1}{2}(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}))^\top\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\left[\frac{\nu+N}{\nu}\frac{1}{w_t}(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)-\mathrm{vec}\,\boldsymbol{I}_N\right]\\ &+\frac{\nu+N}{\nu}\frac{1}{w_t}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1}),\end{aligned} \tag{A.8}$$

where $\psi(x) = d\ln\Gamma(x)/d(x)$ is the digamma function and $\boldsymbol{\mathcal{D}}_N$ the duplication matrix, which allow us to write $\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega} = \boldsymbol{\mathcal{D}}_N(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}))$, since the scale matrix is symmetric. Now, in order to continue the construction of the score vector we define

$$\boldsymbol{s}_t(\boldsymbol{\theta}) = \frac{d\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}}.$$

However, since the decomposition of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top)^\top$ the score vector can be partitioned into two blocks, thus, two different applications of the chain rule are required.

Specifically, for $\boldsymbol{\xi} = (\boldsymbol{\omega}^\top, (\mathrm{vech}(\boldsymbol{\Omega}))^\top, \nu)^\top$, we have

$$\boldsymbol{s}_t^{(\xi)}(\boldsymbol{\theta}) = \frac{d\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\xi}} = \frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\xi}} + \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right)^\top\frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}}, \tag{A.9}$$

while for $\boldsymbol{\psi} = ((\mathrm{vec}\,\boldsymbol{\Phi})^\top, (\mathrm{vec}\,\boldsymbol{K})^\top)^\top$, we have

$$\boldsymbol{s}_t^{(\psi)}(\boldsymbol{\theta}) = \frac{d\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\psi}} = \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right)^\top\frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}}. \tag{A.10}$$

We start by considering the first differential of the dynamic location

$$\begin{aligned}\mathrm{d}(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega}) =& \boldsymbol{\Phi}\mathrm{d}(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega}) + \left[(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})^\top\otimes\boldsymbol{I}_N\right]\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Phi}\\ &+\left[(\boldsymbol{u}_t)^\top\otimes\boldsymbol{I}_N\right]\mathrm{d}\,\mathrm{vec}\,\boldsymbol{K}+\boldsymbol{K}(\mathrm{d}\boldsymbol{u}_t).\end{aligned} \tag{A.11}$$

Now, because the driving-force $\boldsymbol{u}_t$ is a function of the parameters of the multivariate Student's t, it needs to be differentiated with respect to those parameters, by expanding $\mathrm{d}\boldsymbol{u}_t$ as

$$\begin{aligned} \mathrm{d}\boldsymbol{u}_t =& (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}) b_t (1-b_t)/\nu (\mathrm{d}\nu) \\ &+ (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(1-b_t)^2/\nu (\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top (\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2}) \boldsymbol{\mathcal{D}}_N (\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega})) \\ &+ 2(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(1-b_t)^2/\nu (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1} (\mathrm{d}\boldsymbol{\mu}_{t|t-1}) - (1-b_t)(\mathrm{d}\boldsymbol{\mu}_{t|t-1}). \end{aligned} \tag{A.12}$$

We can achieve a more compact form by embeddig the dynamic differential in a stochastic recurrence equation representation, such that

$$\mathrm{d}(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega}) = \boldsymbol{X}_t \mathrm{d}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{R}_t,$$

where

$$\boldsymbol{X}_t = \boldsymbol{\Phi} + \boldsymbol{K}\boldsymbol{\mathcal{C}}_t, \tag{A.13}$$

and

$$\boldsymbol{R}_t = \boldsymbol{K}\boldsymbol{a}_t \mathrm{d}\nu + \boldsymbol{K}\boldsymbol{B}_t \mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega} + \boldsymbol{D}_t \mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Phi} + \boldsymbol{E}_t \mathrm{d}\,\mathrm{vec}\,\boldsymbol{K}. \tag{A.14}$$

The terms of the latter equation are

$$\begin{aligned} \boldsymbol{a}_t &= \frac{\partial \boldsymbol{u}_t}{\partial \nu} = (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}) b_t (1-b_t)/\nu, \\ \boldsymbol{B}_t &= \frac{\partial \boldsymbol{u}_t}{\partial (\mathrm{vech}(\boldsymbol{\Omega}))^\top} = (1-b_t)^2/\nu \, (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t \otimes \boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t)^\top \boldsymbol{\mathcal{D}}_N, \\ \boldsymbol{\mathcal{C}}_t &= \frac{\partial \boldsymbol{u}_t}{\partial \boldsymbol{\mu}_{t|t-1}^\top} = 2(1-b_t)^2/\nu \, (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1} - (1-b_t)\boldsymbol{I}_N, \end{aligned} \tag{A.15}$$

which we present also in their vectorized form

$$\boldsymbol{a}_t = b_t^{3/2}(1-b_t)^{1/2}/\nu \boldsymbol{\Omega}^{1/2} \mathbf{u}_t, \tag{A.16}$$

$$\mathrm{vec}\,\boldsymbol{B}_t = \nu b_t^{3/2}(1-b_t)^{1/2} (\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{1/2})(\mathbf{u}_t \otimes \mathbf{u}_t \otimes \mathbf{u}_t), \tag{A.17}$$

$$\mathrm{vec}\,\boldsymbol{\mathcal{C}}_t = 2b_t(1-b_t)\,(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{1/2})(\mathbf{u}_t \otimes \mathbf{u}_t) - (1-b_t)\,\mathrm{vec}\,\boldsymbol{I}_N. \tag{A.18}$$

We also need the partial derivatives

$$\boldsymbol{C} = \frac{\partial(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} = (\boldsymbol{I}_N - \boldsymbol{\Phi}),$$

$$\boldsymbol{D}_t = \frac{\partial(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{\partial(\mathrm{vec}\,\boldsymbol{\Phi})^\top} = \Big[(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})^\top \otimes \boldsymbol{I}_N\Big], \tag{A.19}$$

$$\boldsymbol{E}_t = \frac{\partial(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{\partial(\mathrm{vec}\,\boldsymbol{K})^\top} = \Big[(\boldsymbol{u}_t)^\top \otimes \boldsymbol{I}_N\Big], \tag{A.20}$$

in order to obtain the final recursions, needed for the iterative procedure

$$\begin{aligned}
\frac{d(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d\nu} =& \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu} + \boldsymbol{K}\boldsymbol{a}_t, \\
\frac{d(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top} =& \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top} + \boldsymbol{K}\boldsymbol{B}_t, \\
\frac{d(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d\boldsymbol{\omega}^\top} =& \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\omega}^\top} + \boldsymbol{C}, \\
\frac{d(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top} =& \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top} + \boldsymbol{D}_t, \\
\frac{d(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top} =& \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top} + \boldsymbol{E}_t.
\end{aligned} \tag{A.21}$$

Similarly, we affront the discussion for the required partial derivatives of the log-likelihood function.

From (A.8) the calculation are straightforward, we define

$$\begin{aligned}
\alpha_t &= \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \nu} = \frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{N}{\nu} + \frac{\nu+N}{\nu} b_t - \ln w_t\right], \\
\boldsymbol{\beta}_t &= \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial(\mathrm{vech}(\boldsymbol{\Omega}))} = \frac{1}{2}\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2})\left[\frac{\nu+N}{\nu}\frac{1}{w_t}(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t) - \mathrm{vec}\,\boldsymbol{I}_N\right], \\
\boldsymbol{\varsigma}_t &= \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}_{t|t-1}} = \frac{\nu+N}{\nu}\frac{1}{w_t}\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}),
\end{aligned}$$

which completes the construction of the score vector.

A.2.2 The Hessian Matrix

With the same spirit, we obtain the second differential of the conditional log-likelihood by differentiating (A.8), which yields

$$
\begin{aligned}
\mathrm{d}^2\ell_t(\boldsymbol{\theta}) =& \frac{1}{2}\left[\frac{1}{2}\psi'\left(\frac{\nu+N}{2}\right)-\frac{1}{2}\psi'\left(\frac{\nu}{2}\right)+\frac{N}{\nu^2}-\frac{N}{\nu^2}b_t-\frac{\nu+N}{\nu^2}b_t(1-b_t)+\frac{1}{\nu}b_t\right](\mathrm{d}^2\nu)\\
&+\left[\frac{\nu+N}{2\nu^2}(1-b_t)^2(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\otimes\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top)(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\
&-\left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right]-\left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\right]\\
&+\left[\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\
&-\left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\
&+\left[2\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top(\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\
&+\left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right]\\
&+\left[\frac{1}{2}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\
&+\left[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t+\frac{1}{2}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)\right]\\
&\times\left[\frac{\nu+N}{\nu^2}b_t(1-b_t)-\frac{N}{\nu^2}(1-b_t)\right](\mathrm{d}\nu),
\end{aligned}
\tag{A.22}
$$

where $\psi'(x) = d^2\ln\Gamma(x)/d(x)^2$ is the trigamma function.

We thus define the Hessian matrix

$$
\boldsymbol{\mathcal{H}}_t(\boldsymbol{\theta}) = \frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top},
$$

Similar arguments of those for the score vector, lead us to the conclusion that further applications of the chain rule divide the Hessian into four blocks.

Start with the first set $\boldsymbol{\xi} = (\boldsymbol{\omega}^\top, (\mathrm{vech}(\boldsymbol{\Omega}))^\top, \nu)^\top$,

$$
\begin{aligned}
\boldsymbol{\mathcal{H}}_t^{(\xi)}(\boldsymbol{\theta}) &= \frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\xi}d\boldsymbol{\xi}^\top}\\
&= \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^\top}+\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right)^\top\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right)+\frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}^\top}\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}d\boldsymbol{\xi}^\top}.
\end{aligned}
\tag{A.23}
$$

As regards the second vector of parameters $\boldsymbol{\psi} = ((\mathrm{vec}\,\boldsymbol{\Phi})^\top, (\mathrm{vec}\,\boldsymbol{K})^\top)^\top$, we have

$$
\begin{aligned}
\boldsymbol{\mathcal{H}}_t^{(\psi)}(\boldsymbol{\theta}) &= \frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\psi}d\boldsymbol{\psi}^\top}\\
&= \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right)^\top\frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right)+\frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}^\top}\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}d\boldsymbol{\psi}^\top},
\end{aligned}
\tag{A.24}
$$

and finally, by virtue of its symmetry, we conclude the remaining blocks with the cross-derivatives

$$\begin{aligned}\mathcal{H}_t^{(\xi,\psi)}(\boldsymbol{\theta}) &= \frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\xi} d\boldsymbol{\psi}^\top} \\ &= \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right)^\top \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right) + \frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}^\top}\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi} d\boldsymbol{\psi}^\top}.\end{aligned} \tag{A.25}$$

It should be clear now that, according to the above rules, the calculation of the Hessian matrix will also involve the second differentials of the dynamic equation, which we are going to explore.

We have,

$$\begin{aligned}\mathrm{d}^2\boldsymbol{\mu}_{t+1|t} =& \boldsymbol{\Phi}\mathrm{d}^2\boldsymbol{\mu}_{t|t-1} + 2[\mathrm{d}(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})^\top \otimes \boldsymbol{I}_N]\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Phi} \\ &+ 2[\mathrm{d}(\boldsymbol{u}_t)^\top \otimes \boldsymbol{I}_N]\,\mathrm{vec}\,\boldsymbol{K} + \boldsymbol{K}(\mathrm{d}^2\boldsymbol{u}_t),\end{aligned} \tag{A.26}$$

that in turn imply to expand $\mathrm{d}^2\boldsymbol{u}_t$, with respect to the parameters of the multivariate Student's t.

After some algebra we get the second differential of the driving-force

$$\begin{aligned}\mathrm{d}^2\boldsymbol{u}_t =& 2(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})/\nu\Big[b_t^2(1-b_t)/\nu - b_t(1-b_t)\Big](\mathrm{d}^2\nu) \\ &+ 2(1-b_t)^3/\nu^2\Big\{\Big[(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top \otimes (\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top\Big]\,\mathrm{vec}(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\Big\} \\ &\qquad\qquad \times\Big[(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\Big] \\ &- 2(1-b_t)^2/\nu\Big\{\Big[(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\otimes\boldsymbol{\Omega}^{-1})\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}\Big](\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\Big\} \\ &+ 8(1-b_t)^3/\nu^2\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\Big]\,\mathrm{vec}\,\boldsymbol{\Omega}^{-1}\Big\} \\ &\qquad\qquad \times\Big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\Big] \\ &- 2(1-b_t)^2/\nu\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_N\Big]\Big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\otimes\boldsymbol{I}_N+\boldsymbol{I}_N\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\Big](\mathrm{d}\boldsymbol{\mu}_{t|t-1})\Big\} \\ &- 2(1-b_t)^2/\nu\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_N\Big]\Big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\otimes\boldsymbol{I}_N\Big](\mathrm{d}\boldsymbol{\mu}_{t|t-1})\Big\} \\ &+ 2(1-b_t)^2/\nu\Big\{\Big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})\Big]\Big\} - (1-b_t)\Big\{(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})\Big\} \\ &+ 4(1-b_t)^3/\nu^2\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\otimes\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\Big](\mathrm{vec}\,\boldsymbol{\Omega}^{-1})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\Big\} \\ &- (1-b_t)^2/\nu\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\otimes\boldsymbol{I}_N\Big](\mathrm{vec}\,\boldsymbol{I}_N)\Big[(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\Big]\Big\} \\ &- 2(1-b_t)^2/\nu\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\otimes\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\Big](\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\Big\} \\ &+ \Big\{[(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top]\,\mathrm{vec}(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\Big\} \\ &\qquad\qquad \times\Big[2b_t(1-b_t)^2/\nu^2-(1-b_t)^2/\nu^2)\Big](\mathrm{d}\nu) \\ &+ 2\Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\Big]\,\mathrm{vec}\,\boldsymbol{\Omega}^{-1}\Big\}\Big[2b_t(1-b_t)/\nu^2-(1-b_t)^2/\nu^2\Big](\mathrm{d}\nu) \\ &- \Big\{\Big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\otimes\boldsymbol{I}_N\Big](\mathrm{vec}\,\boldsymbol{I}_N)\Big\}\Big[b_t(1-b_t)/\nu\Big](\mathrm{d}\nu).\end{aligned} \tag{A.27}$$

It is useful to consider the dynamic second differential as a stochastic recurrence equation

$$\mathrm{d}^2(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega}) = \boldsymbol{X}_t \mathrm{d}^2(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\mathrm{d}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})^\top \boldsymbol{\mathcal{C}}_t' \mathrm{d}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{Q}_t,$$

where again

$$\boldsymbol{X}_t = \boldsymbol{\Phi} + \boldsymbol{K}\boldsymbol{\mathcal{C}}_t,$$

and

$$\begin{aligned} \boldsymbol{Q}_t =& \boldsymbol{K}\boldsymbol{a}_t'\mathrm{d}^2\nu + \boldsymbol{K}\boldsymbol{B}_t'\mathrm{d}^2 \operatorname{vec}\boldsymbol{\Omega} + \boldsymbol{K}(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})^\top \widehat{\boldsymbol{aB}}_t' \mathrm{d}\nu \\ &+ \boldsymbol{D}_t'\mathrm{d}^2 \operatorname{vec}\boldsymbol{\Phi} + \boldsymbol{E}_t'\mathrm{d}^2 \operatorname{vec}\boldsymbol{K} + (\mathrm{d}\operatorname{vec}\boldsymbol{\Phi})^\top \widehat{\boldsymbol{DE}}_t' (\mathrm{d}\operatorname{vec}\boldsymbol{K}). \end{aligned} \tag{A.28}$$

We now derive the terms of recursion (A.28).

We first need a set of partial derivative

$$\boldsymbol{a}_t' = \frac{\partial^2 \boldsymbol{u}_t}{\partial \nu^2} = 2(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu \left[b_t^2(1-b_t)/\nu - b_t(1-b_t) \right], \tag{A.29}$$

$$\begin{aligned} \boldsymbol{B}_t' = \frac{\partial^2 \boldsymbol{u}_t}{\partial(\operatorname{vech}(\boldsymbol{\Omega}))\partial(\operatorname{vech}(\boldsymbol{\Omega}))^\top} =& 2(1-b_t)^3/\nu^2 \\ &\times \left\{ \left[\boldsymbol{\mathcal{D}}_N^\top \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top \right] \operatorname{vec}(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2}) \right\} \\ &\qquad \times \left[(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top (\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2}) \boldsymbol{\mathcal{D}}_N \right] \\ &- 2(1-b_t)^2/\nu \\ &\times \left\{ \left[\boldsymbol{\mathcal{D}}_N^\top (\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t \otimes \boldsymbol{\Omega}^{-1}) \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1} \right] \boldsymbol{\mathcal{D}}_N \right\}, \end{aligned} \tag{A.30}$$

$$\begin{aligned} \boldsymbol{\mathcal{C}}_t' = \frac{\partial^2 \boldsymbol{u}_t}{\partial \boldsymbol{\mu}_{t|t-1} \partial \boldsymbol{\mu}_{t|t-1}^\top} =& 8(1-b_t)^3/\nu^2 \left\{ \left[\boldsymbol{I}_N \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \right] \operatorname{vec}\boldsymbol{\Omega}^{-1} \right\} \\ &\qquad \times \left[(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1} \right] \\ &- 2(1-b_t)^2/\nu \left\{ \left[\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{I}_N \right] \left[(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}) \otimes \boldsymbol{I}_N + \boldsymbol{I}_N \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}) \right] \right\} \\ &- 2(1-b_t)^2/\nu \left\{ \left[\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{I}_N \right] \left[(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}) \otimes \boldsymbol{I}_N \right] \right\}. \end{aligned} \tag{A.31}$$

Secondly, a set of partial cross-derivatives

$$\begin{aligned}
\widehat{\boldsymbol{aB}}_t' = \frac{\partial^2 \boldsymbol{u}_t}{\partial(\mathrm{vech}(\boldsymbol{\Omega}))\partial\nu} =& [\boldsymbol{I}_N \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top] \,\mathrm{vec}(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2}) \\
& \times \left[2b_t(1-b_t)^2/\nu^2 - (1-b_t)^2/\nu^2)\right], \qquad \text{(A.32)}\\
\widehat{\boldsymbol{a\mathcal{C}}}_t' = \frac{\partial^2 \boldsymbol{u}_t}{\partial\boldsymbol{\mu}_{t|t-1}\partial\nu} =& 2\left\{\left[\boldsymbol{I}_N \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top\right] \mathrm{vec}\,\boldsymbol{\Omega}^{-1}\right\} \\
& \times \left[2b_t(1-b_t)/\nu^2 - (1-b_t)^2/\nu^2\right] \\
& - \left\{\left[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top \otimes \boldsymbol{I}_N\right](\mathrm{vec}\,\boldsymbol{I}_N)\right\} \\
& \times \left[b_t(1-b_t)/\nu\right], \qquad \text{(A.33)}\\
\widehat{\boldsymbol{B\mathcal{C}}}_t' = \frac{\partial^2 \boldsymbol{u}_t}{\partial\boldsymbol{\mu}_{t|t-1}\partial(\mathrm{vech}(\boldsymbol{\Omega}))^\top} =& 4(1-b_t)^3/\nu^2\left\{\left[\boldsymbol{I}_N \otimes \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\right](\mathrm{vec}\,\boldsymbol{\Omega}^{-1})(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\right\} \\
& - (1-b_t)^2/\nu\left\{\left[\boldsymbol{I}_N \otimes \boldsymbol{I}_N\right](\mathrm{vec}\,\boldsymbol{I}_N)\left[(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\right]\right\} \\
& - 2(1-b_t)^2/\nu\left\{\left[\boldsymbol{I}_N \otimes \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\right](\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\right\}. \qquad \text{(A.34)}
\end{aligned}$$

In addition, we still need a set of partial derivatives defined by

$$\boldsymbol{D}_t' = \frac{\partial[d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})]}{\partial(\mathrm{vec}\,\boldsymbol{\Phi})d(\mathrm{vec}\,\boldsymbol{\Phi})^\top} = 2\left[\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top}\right)^\top \otimes \boldsymbol{I}_N\right], \qquad \text{(A.35)}$$

$$\boldsymbol{E}_t' = \frac{\partial[d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})]}{\partial(\mathrm{vec}\,\boldsymbol{K})d(\mathrm{vec}\,\boldsymbol{K})^\top} = 2\left[\left(\boldsymbol{\mathcal{C}}_t^\top \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top}\right)^\top \otimes \boldsymbol{I}_N\right], \qquad \text{(A.36)}$$

and finally conclude the derivation with

$$\widehat{\boldsymbol{DE}}_t' = \frac{\partial[d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})]}{\partial(\mathrm{vec}\,\boldsymbol{\Phi})d(\mathrm{vec}\,\boldsymbol{K})^\top} = \left[\left(\boldsymbol{\mathcal{C}}_t^\top \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top}\right)^\top \otimes \boldsymbol{I}_N\right]. \qquad \text{(A.37)}$$

We therefore have obtained a new set of recursions composed by

$$\begin{aligned}
\frac{d^2(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d\nu^2} =& \boldsymbol{X}_t \frac{d^2(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\nu^2} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\nu}\right)^\top \boldsymbol{\mathcal{C}}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\nu}\right) + \boldsymbol{K}\boldsymbol{a}_t', \\
\frac{d^2(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))d(\mathrm{vech}(\boldsymbol{\Omega}))^\top} =& \boldsymbol{X}_t \frac{d^2(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))d(\mathrm{vech}(\boldsymbol{\Omega}))^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top}\right)^\top \boldsymbol{\mathcal{C}}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top}\right) + \boldsymbol{K}\boldsymbol{B}_t', \\
\frac{d^2(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))d\nu} =& \boldsymbol{X}_t \frac{d^2(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))d\nu} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top}\right)^\top \boldsymbol{\mathcal{C}}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\nu}\right) + \boldsymbol{K}\widehat{\boldsymbol{aB}}_t',
\end{aligned}$$

which continue with

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})d(\operatorname{vec}\boldsymbol{\Phi})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})d(\operatorname{vec}\boldsymbol{\Phi})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})^\top}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})^\top}\right) + \boldsymbol{D}_t',$$

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})d(\operatorname{vec}\boldsymbol{K})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})d(\operatorname{vec}\boldsymbol{K})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})^\top}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})^\top}\right) + \boldsymbol{E}_t',$$

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})d(\operatorname{vec}\boldsymbol{K})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})d(\operatorname{vec}\boldsymbol{K})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})^\top}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})^\top}\right) + \widehat{\boldsymbol{DE}}_t',$$

and conclude with

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\nu)d(\operatorname{vec}\boldsymbol{\Phi})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\nu)d(\operatorname{vec}\boldsymbol{\Phi})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})^\top}\right),$$

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\nu)d(\operatorname{vec}\boldsymbol{K})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\nu)d(\operatorname{vec}\boldsymbol{K})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\nu}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})^\top}\right),$$

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\nu)d(\operatorname{vec}\boldsymbol{\Phi})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vech}(\boldsymbol{\Omega}))d(\operatorname{vec}\boldsymbol{\Phi})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vech}(\boldsymbol{\Omega}))^\top}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{\Phi})^\top}\right),$$

$$\frac{d^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})}{d(\operatorname{vech}(\boldsymbol{\Omega}))d(\operatorname{vec}\boldsymbol{K})^\top} = \boldsymbol{X}_t\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vech}(\boldsymbol{\Omega}))d(\operatorname{vec}\boldsymbol{K})^\top} + \boldsymbol{K}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vech}(\boldsymbol{\Omega}))^\top}\right)^\top \boldsymbol{C}_t'\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d(\operatorname{vec}\boldsymbol{K})^\top}\right).$$

The construction of the Hessian can now be completed by simply deriving the remaining second-order partial derivatives of the second differential in (A.22).

Thanks to this representation it is easy to show that

$$\alpha_t' = \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \nu^2} = \frac{1}{2}\left[\frac{1}{2}\psi'\left(\frac{\nu+N}{2}\right) - \frac{1}{2}\psi'\left(\frac{\nu}{2}\right) + \frac{N}{\nu^2} - \frac{N}{\nu^2}b_t - \frac{\nu+N}{\nu^2}b_t(1-b_t) + \frac{1}{\nu}b_t\right],$$

$$\begin{aligned}\boldsymbol{\beta}_t' = \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial(\operatorname{vech}(\boldsymbol{\Omega}))\partial(\operatorname{vech}(\boldsymbol{\Omega}))^\top} &= \left[\frac{\nu+N}{2\nu^2}(1-b_t)^2\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\otimes\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top)(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\boldsymbol{\mathcal{D}}_N\right]\\ &\quad - \left[\frac{\nu+N}{\nu}(1-b_t)\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2})\boldsymbol{\mathcal{D}}_N\right]\\ &\quad + \left[\frac{1}{2}\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{\mathcal{D}}_N\right],\end{aligned}$$

$$\boldsymbol{\varsigma}_t' = \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top} = \left[\frac{\nu+N}{\nu^2}2(1-b_t)^2\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\right] - \left[\frac{\nu+N}{\nu}(1-b_t)\boldsymbol{\Omega}^{-1}\right],$$

and

$$\widehat{\alpha\boldsymbol{\beta}'_t} = \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial(\mathrm{vech}(\boldsymbol{\Omega}))\partial\nu} = \frac{1}{2}\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)\left[\frac{\nu+N}{\nu^2}b_t(1-b_t)-\frac{N}{\nu^2}(1-b_t)\right],$$

$$\widehat{\alpha\boldsymbol{\varsigma}'_t} = \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\nu} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\left[\frac{\nu+N}{\nu^2}b_t(1-b_t)-\frac{N}{\nu^2}(1-b_t)\right],$$

$$\begin{aligned}\widehat{\boldsymbol{\beta}\boldsymbol{\varsigma}'_t} = \frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial(\mathrm{vech}(\boldsymbol{\Omega}))^\top} &= \left[\frac{\nu+N}{\nu^2}(1-b_t)^2(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2})\boldsymbol{\mathcal{D}}_N\right]\\ &+\left[\frac{\nu+N}{\nu}2(1-b_t)(\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{\mathcal{D}}_N\right],\end{aligned}$$

which completes the construction of the Hessian matrix.

A.2.3 The Conditional Information Matrix

Taking conditional expectation of the negative Hessian matrix yields the fundamental conditional information matrix needed for the Fisher's scoring method. Likewise to the score and the Hessian, we start the discussion by taking advantage again from the differentials of the log-likelihood function.

Then we have

$$\begin{aligned}\mathbb{E}_{t-1}[\mathrm{d}^2\ell_t(\boldsymbol{\theta})] = &\left[\frac{1}{4}\psi'\left(\frac{\nu+N}{2}\right)-\frac{1}{4}\psi'\left(\frac{\nu}{2}\right)+\frac{N(\nu+N+4)}{2\nu(\nu+N)(\nu+N+2)}\right](\mathrm{d}^2\nu)\\ &+\left[\frac{1}{2(\nu+N+2)}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\mathrm{vec}\,\boldsymbol{\Omega}^{-1})(\mathrm{vec}\,\boldsymbol{\Omega}^{-1})^\top(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\ &-\left[\frac{\nu+N}{2(\nu+N+2)}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right]\\ &+\left[\frac{1}{(\nu+N)(\nu+N+2)}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\mathrm{vec}\,\boldsymbol{\Omega}^{-1})(\mathrm{d}\nu)\right]\\ &-\left[\frac{\nu+N}{\nu+N+2}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right].\end{aligned}$$

The calculations of this matrix require for the first set $\boldsymbol{\xi} = (\boldsymbol{\omega}^\top, (\mathrm{vech}(\boldsymbol{\Omega}))^\top, \nu)^\top$,

$$\boldsymbol{\mathcal{I}}_t^{(\xi)}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\xi}d\boldsymbol{\xi}^\top}\right] = \boldsymbol{\mathcal{I}}^{(\xi)}(\boldsymbol{\theta}) + \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right)^\top\boldsymbol{\mathcal{I}}^{(\mu)}(\boldsymbol{\theta})\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right), \tag{A.38}$$

For the second vector $\boldsymbol{\psi} = ((\mathrm{vec}\,\boldsymbol{\Phi})^\top, (\mathrm{vec}\,\boldsymbol{K})^\top)^\top$, we have

$$\boldsymbol{\mathcal{I}}_t^{(\psi)}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\psi}d\boldsymbol{\psi}^\top}\right] = \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right)^\top\boldsymbol{\mathcal{I}}^{(\mu)}(\boldsymbol{\theta})\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right), \tag{A.39}$$

in conclusion, the negative conditional expected value of the cross-second derivatives are

$$\boldsymbol{\mathcal{I}}_t^{(\xi,\psi)}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\xi}d\boldsymbol{\psi}^\top}\right] = \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\xi}^\top}\right)^\top\boldsymbol{\mathcal{I}}^{(\mu)}(\boldsymbol{\theta})\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\psi}^\top}\right). \tag{A.40}$$

Now, the recursion the partial derivative for the dynamic location are already available from (A.21) and thus, the calculations boils down to the static terms of the matrix, which are easily retrieved.

We get

$$\mathcal{I}^{(\mu)}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}_{t|t-1}\partial \boldsymbol{\mu}_{t|t-1}^\top}\right] = \frac{\nu+N}{\nu+N+2}\boldsymbol{\Omega}^{-1},$$

while the terms of the static matrix $\mathcal{I}^{(\xi)}(\boldsymbol{\theta})$ are

$$\mathcal{I}^{(\nu)}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \nu^2}\right] = \frac{1}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+N}{2}\right) - \frac{2N(\nu+N+4)}{\nu(\nu+N)(\nu+N+2)}\right],$$

$$\begin{aligned}\mathcal{I}^{(\mathrm{v}(\boldsymbol{\Omega}))}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial(\mathrm{vech}(\boldsymbol{\Omega}))\partial(\mathrm{vech}(\boldsymbol{\Omega}))^\top}\right] &= \frac{\nu+N}{2(\nu+N+2)}\mathcal{D}_N^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\mathcal{D}_N \\ &\quad - \frac{1}{2(\nu+N+2)}\mathcal{D}_N^\top(\mathrm{vech}(\boldsymbol{\Omega}^{-1}))(\mathrm{vech}(\boldsymbol{\Omega}^{-1}))^\top\mathcal{D}_N,\end{aligned}$$

and lastly the cross terms

$$\mathcal{I}^{(\mathrm{v}(\boldsymbol{\Omega}),\nu)}(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial(\mathrm{vech}(\boldsymbol{\Omega}))\partial\nu}\right] = -\frac{1}{(\nu+N)(\nu+N+2)}\mathcal{D}_N^\top(\mathrm{vech}(\boldsymbol{\Omega}^{-1})).$$

With these last derivations, we have completed the derivations for the Fisher's scoring method in the multivariate DCS-t set up.

A.2.4 Third differentials

The end of this appendix present the third differential of the conditional log-likelihood with respect to the dynamic location, since, as it turns out, it is needed for the proof of the asymptotic normality of the MLE, see Lemma A.3.12.

We invoke again equation (A.22), differentiate only with respect $\boldsymbol{\mu}_{t|t-1}$ and obtain

$$\begin{aligned}\mathrm{d}^3_{\boldsymbol{\mu}_{t|t-1}}\ell_t(\boldsymbol{\theta}) = &\left[8\frac{\nu+N}{\nu^3}(1-b_t)^3(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\ &+ \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top[\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\otimes\boldsymbol{I}_N + \boldsymbol{I}_N\otimes\boldsymbol{\epsilon}_t\boldsymbol{\Omega}^{-1/2}](\mathrm{d}\boldsymbol{\mu}_{t|t-1})^2\right] \\ &- \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\ &- \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\right] \\ &- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\ &- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}^3\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\right].\end{aligned} \tag{A.41}$$

A.3 Lemmata

A.3.1 Lemmata for the Proof of Consistency

Lemma A.3.1 (Uniform Integrability and moments of the likelihood). *Consider model (2.1), (2.2) and (2.3). Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence. Assume that*

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,

Then, for all $m \geq 1$, we obtain

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta})|^m\right] < \infty$$

if and only if $\nu > 2$.

Proof. Consider the t-th contribution to the likelihood in equation (2.9). With the c_r-inequality and Lemma 2.2.1 at hand, it should be clear that, if there exist a stationary ergodic sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and $\varrho(\boldsymbol{\Phi}) < 1$ in model (2.1), (2.2) and (2.3), then we have that $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\mu}_{t|t-1}\|^m] < \infty$ and

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ell_t(\boldsymbol{\theta})\right|^m\right] \\
&\leq \bar{C}_1\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ln[1+(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_t)/\nu]\right|^m\right] \leq \bar{C}_1\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\right|^m\right] \\
&= \bar{C}_1\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\operatorname{tr}\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\right|^m\right] \leq \bar{C}_1\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\|\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\|\right|^m\right] \\
&\leq \bar{C}_3\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1}\|^{2m}\right] \leq \bar{C}_3\mathbb{E}\left[\|\boldsymbol{y}_t\|^{2m}\right] + \bar{C}_3\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\mu}_{t|t-1}\|^{2m}\right] < \infty,
\end{aligned}$$

since the second inequality is entailed in the elementary property that $\ln(x) \leq x - 1$ for $x > 1$. The third inequality follows from the fact that $|\operatorname{tr}(\boldsymbol{AB})| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$, see result 4.1.2(2) of Lütkepohl, 1996. The fourth is trivial, since the scale matrix is symmetric and positive definite, thus invertible and the last inequality follows from subadditivity. The moment bounds now can be shown to be satisfied by assumptions. For example, $\mathbb{E}[\|\boldsymbol{y}_t\|^2] < \infty$ is always ensured if $\nu > 2$ and $\bar{C}_3 > 0$ is a finite constant. The result is extended to the whole likelihood as an application of the Continuous Mapping Theorem. □

Lemma A.3.2 (Identifiability Uniqueness of the true parameter vector). *Consider model (2.1), (2.2) and (2.3). Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence. Assume that*

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,
2. *the parameter space $\boldsymbol{\Theta}$ is compact with $2 < \nu < \infty$ and*
3. *the true parameter vector $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.*

Then,

$$\mathbb{E}\Big[|\ell_t(\boldsymbol{\theta}_0)|\Big] < \infty.$$

Furthermore, for every $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$,

$$\mathbb{E}\Big[|\ell_t(\boldsymbol{\theta})|\Big] < \mathbb{E}\Big[|\ell_t(\boldsymbol{\theta}_0)|\Big].$$

Proof. We immediately note that $\mathbb{E}\Big[|\ell_t(\boldsymbol{\theta}_0)|\Big] < \infty$ follows from Lemma A.3.1 and then we can turn to the second statement.

In proving identifiability uniqueness of $\boldsymbol{\theta}_0$ it suffices to consider the sequence $\{\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)\}$, under the assumption that $(\nu, \text{vech}\,\boldsymbol{\Omega})^\top = (\nu_0, \text{vech}\,\boldsymbol{\Omega}_0)^\top$. Thus, denoting with $\boldsymbol{\mu}_{0,t|t-1}$ the dynamic location vector as a function of the true parameter vector, the difference sequence between the two likelihoods is

$$\begin{aligned}
&\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0) \\
&\quad = \ln\Big[1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}_0^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu_0\Big] - \ln\Big[1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{0,t|t-1})^\top \boldsymbol{\Omega}_0^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{0,t|t-1})/\nu_0\Big] \\
&\quad \leq \Big[(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}_0^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu_0\Big] - \Big[(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}_0^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{0,t|t-1})/\nu_0\Big],
\end{aligned}$$

where the inequality is implied by the elementary relation $\ln(x) \leq x - 1$ for $x > 1$, that will be always strict unless $x = 1$, which will be the case if and only if $\boldsymbol{\mu}_{t|t-1} = \boldsymbol{\mu}_{0,t|t-1}$ almost surely since, $\boldsymbol{\Omega}_0$ is symmetric positive definite and $2 < \nu < \infty$. Thus, taking expectation yields

$$\mathbb{E}[\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)] < \mathbb{E}\Big[\text{tr}\Big((\boldsymbol{\Omega}_0^{-1}/\nu_0)\Big((\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top - (\boldsymbol{y}_t - \boldsymbol{\mu}_{0,t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{0,t|t-1})^\top\Big)\Big)\Big].$$

Also, is not hard to see that this in turn implies that we can write the recursion as

$$(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\mu}_{0,t]1|t}) = (\boldsymbol{\omega} - \boldsymbol{\omega}_0) + (\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\omega}_0 + (\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\mu}_{0,t|t-1} + (\boldsymbol{K} - \boldsymbol{K}_0)\boldsymbol{u}_t,$$

and relation above entails the fact that, if $\boldsymbol{\mu}_{t|t-1} = \boldsymbol{\mu}_{0,t|t-1}$ for all t almost surely, then

$$(\boldsymbol{\omega} - \boldsymbol{\omega}_0) + (\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\omega}_0 = (\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\boldsymbol{\mu}_{0,t|t-1} + (\boldsymbol{K} - \boldsymbol{K}_0)\boldsymbol{u}_t,$$

almost surely. Nonetheless, we note that as long as $\det \boldsymbol{K} \neq 0$ the whole multivariate system of equations is indeed stochastic, since one cannot find a nontrivial solution of the system that will cancel out the driving force $\boldsymbol{u}_t$ of the dynamic location vector and as a result the only available option boils down to the equivalence between all the parameters, that is $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $\boldsymbol{\Phi} = \boldsymbol{\Phi}_0$ and $\boldsymbol{K} = \boldsymbol{K}_0$.

Summing up, we have shown that $\mathbb{E}[\ell_t(\boldsymbol{\theta})] < \mathbb{E}[\ell_t(\boldsymbol{\theta}_0)]$ for every $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. □

Lemma A.3.3 (Uniform convergence of the likelihood function). *Consider model (2.1), (2.2) and (2.3). Let* $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence. Assume that*

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,
2. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \prod_{k=1}^{j} \frac{\partial \boldsymbol{\mu}_{t-k+1|t-k}}{\partial \boldsymbol{\mu}_{t-k|t-k-1}^\top} \right\|\right] < 1$, *for some* $j \geq 1$ *large enough and*
3. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \boldsymbol{\mu}_{1|0}}{\partial \boldsymbol{\theta}^\top} \right\|\right] < \infty$,
4. *the parameter space* $\boldsymbol{\Theta}$ *is compact with* $2 < \nu < \infty$.

Then,

$$\sup_{\theta\in\Theta} |\widehat{\mathcal{L}}_T(\theta) - \mathcal{L}_T(\theta)| \xrightarrow{a.s.} 0 \qquad as \qquad t \to \infty,$$

where $\widehat{\mathcal{L}}_T(\theta)$ *is the empirical likelihood started with* $\mu_{1|0}$ *and* $\mathcal{L}(\theta)$ *is the unique stationary ergodic counterpart, defined in* (A.4) *and* (A.5) *respectively.*

Proof. We apply a mean-value expansion of the log-likelihood around $\mu^{\star}_{t|t-1}$ which is on the chord between the started filtered location $\hat{\mu}_{t|t-1}$ and $\mu_{t|t-1}$. We take the supremum over the compact parameter space and see that

$$\sup_{\theta\in\Theta} |\widehat{\mathcal{L}}_T(\theta) - \mathcal{L}_T(\theta)| \le \sup_{\theta\in\Theta} \left\| \frac{\partial \widehat{\mathcal{L}}_T(\theta)}{\partial \mu^{\star\top}_{t|t-1}} \right\| \sup_{\theta\in\Theta} \|\hat{\mu}_{t|t-1} - \mu_{t|t-1}\|,$$

where by direct calculation and by the triangle inequality we have

$$\begin{aligned}\sup_{\theta\in\Theta} \left\| \frac{\partial \widehat{\mathcal{L}}_T(\theta)}{\partial \mu^{\top}_{t|t-1}} \right\| \le & \frac{1}{T}\sum_{t=1}^{T} \sup_{\theta\in\Theta} \left\| \Omega^{-1} \frac{\nu+N}{\nu} \frac{(y_t - \mu_{t|t-1})}{1 + (y_t - \mu_{t|t-1})^{\top}\Omega^{-1}(y_t - \mu_{t|t-1})/\nu} \right\| \\ \le & c_{\Omega} \left(\max_{\theta\in\Theta} \frac{\nu+N}{\nu} \right) \frac{1}{T}\sum_{t=1}^{T} \sup_{\theta\in\Theta} \|y_t - \mu_{t|t-1}\| \\ & \times \sup_{\theta\in\Theta} \left| \left[1 + (y_t - \mu_{t|t-1})^{\top}\Omega^{-1}(y_t - \mu_{t|t-1})/\nu \right]^{-1} \right|.\end{aligned}$$

Assumption 4 is of fundamental importance here. As we can observe, if we treat the dynamic location vector as a fixed parameter with value $\mu^{\star}_{t|t-1}$ and let $y_t \to \infty$ the entire term in the right hand side of the latter inequality will vanishes. Hence, we satisfy the condition that at least

$$\sup_{\theta\in\Theta} \left\| \frac{\partial \widehat{\mathcal{L}}_T(\theta)}{\partial \mu^{\star\top}_{t|t-1}} \right\| = O_p(1),$$

which is enough to imply the existence of log-moments. Furthermore, Assumptions 1 and 2 are needed in order to maintain the filter invertible and thus to apply Lemma 2.3.3 such that

$$\sup_{\theta\in\Theta} \|\hat{\mu}_{t|t-1} - \mu_{t|t-1}\| \xrightarrow{\text{e.a.s.}} 0,$$

In conclusion, an application of Lemma 2.1 in Straumann and Mikosch, 2006 demonstrates that the claimed almost sure convergence holds true. □

Lemma A.3.4 (Uniform Law of Large Numbers of the likelihood). *Consider model* (2.1), (2.2) *and* (2.3). *Let* $\{\epsilon_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence. Assume that*

1. $\varrho(\Phi) < 1$ *and* $\det K \neq 0$,
2. *the parameter space* Θ *is compact with* $2 < \nu < \infty$,
3. *the true parameter vector* $\theta_0 \in \Theta$ *and*
4. $\forall \theta \in \Theta$, *if* $\theta \neq \theta_0$ *then* $\mu_{t|t-1} \neq \mu_{0t}$ *almost surely and* $\forall t \geq 1$.

Then,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \qquad as \qquad t \to \infty,$$

where $\mathcal{L}_T(\boldsymbol{\theta})$ is the stationary ergodic average likelihood and $\mathcal{L}(\boldsymbol{\theta})$ is the limit likelihood are defined in (A.5) and (A.6) respectively.

Proof. The Uniform Law of Large Numbers in its version for ergodic stationary processes is reported on White, 1994 as Theorem A.2.2, applies straightforwardly to our case since

1. the parameter space is compact,
2. the empirical likelihood function $\mathcal{L}_T(\boldsymbol{\theta})$ defined in (A.5) is continuous in $\boldsymbol{\theta}\ \forall\ \boldsymbol{y}_t$ and $\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$ is measurable in $\boldsymbol{y}_t$,
3. By Lemma A.3.2 we obtain the identifiability and the moment bound $\mathbb{E}\left[|\ell_t(\boldsymbol{\theta}_0)|\right] < \infty$ ensure the dominance condition.

Thus, all the conditions of Theoerm A.2.2 in White, 1994 are met and the proof is complete. □

A.3.2 Lemmata for the Proof of Asymptotic Normality

Lemma A.3.5 (Stationarity, Ergodicity and Moments for the First Differentials of the Dynamic Location). *Consider model (2.1), (2.2) and (2.3). Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence. Consider the stochastic difference equation*

$$\mathrm{d}(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega}) = \boldsymbol{X}_t \mathrm{d}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{R}_t,$$

where $\boldsymbol{X}_t$ is defined as in (A.13)

$$\boldsymbol{X}_t = \boldsymbol{\Phi} + \boldsymbol{K}\boldsymbol{\mathcal{C}}_t,$$

where $\boldsymbol{\mathcal{C}}_t$ is in (A.15), $\boldsymbol{R}_t$ as in (A.14),

$$\boldsymbol{R}_t = \boldsymbol{K}\boldsymbol{a}_t \mathrm{d}\nu + \boldsymbol{K}\boldsymbol{B}_t \mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega} + \boldsymbol{D}_t \mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Phi} + \boldsymbol{E}_t \mathrm{d}\,\mathrm{vec}\,\boldsymbol{K},$$

with components in (A.16), (A.17), (A.18), (A.19) and (A.20). If for some integer $t \geq 1$ and $m > 0$ the conditions

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,
2. $\gamma(\boldsymbol{X}_t) = \left\{ \frac{1}{t}\mathbb{E}\left[\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{X}_1 \dots \boldsymbol{X}_t\| \right] \right\} < 0$,
3. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{R}_t\|^m \right] < \infty$,

are satisfied, then the series $\{\mathrm{d}(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})\}$ represented by

$$\mathrm{d}(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega}) = \sum_{j=0}^{t-1} \left(\prod_{k=1}^{j} \boldsymbol{X}_{t-k} \right) \boldsymbol{R}_{t-j}$$

converges almost surely to the unique stationary ergodic solution.

Furthermore, $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathrm{d}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})\|^m] < \infty$ for every $m > 0$.

Proof. The discussion here follows closely the arguments of the proof of Lemma 2.3.1, in fact, if we consider the recursion (A.3.5) as a function of the innovations $\boldsymbol{\epsilon}_t$, we can easily see again that $\boldsymbol{X}_t$ and all the components of $\boldsymbol{R}_t$ are just linear in $\boldsymbol{\mu}_t$, implying that the Lyapunov condition 2 and the moment bounds of condition 3, are enough to ensure the generation of stationary ergodic $\{\boldsymbol{X}_t, \boldsymbol{R}_t\}$

Now, from Lemma 2.3.1, is clear that the first two conditions are used in order to maintain the multivariate system stable and the matrices $\boldsymbol{X}_t$ random, while the well-known Lyapunov condition 2 for linear stochastic difference equations, give us the sufficient condition in order to obtain ergodic sequences $\{\boldsymbol{X}_t\}$, see Vervaat, 1979 and Basrak, Davis, and Mikosch, 2002.

Moreover, because the Hölder and Minkowsky inequalities imply

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathrm{d}(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega})\|^m\right] \leq \sum_{j=0}^{t-1}\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\prod_{k=0}^{j}\boldsymbol{X}_{t-k}\right\|^m\right]^{1/m}\mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathrm{d}\boldsymbol{R}_{t-j}\|^m\right]^{1/m}.$$

In addition, we note that equation (A.18) imply

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}\|\boldsymbol{X}_t\|^m\right] &\leq \sup_{\theta\in\Theta}\|\boldsymbol{\Phi}\|^m + \mathbb{E}\left[\sup_{\theta\in\Theta}\|\boldsymbol{K}\boldsymbol{C}_t\|^m\right] \\
&\leq \bar{\rho}^m + c_{\boldsymbol{K}}\mathbb{E}\left[b_t^{m/2}(1-b_t)^{m/2}\right]\mathbb{E}\left[\|(\mathbf{u}_t\otimes\mathbf{u}_t)\|^m\right] + c_{\boldsymbol{K}}N^{m/2}\mathbb{E}\left[(1-b_t)^{m/2}\right] \\
&= \bar{\rho}^m + c_{\boldsymbol{K}}\mathbb{E}\left[\|\mathbf{u}_t\|^{2m}\right]\frac{\mathcal{B}\left(\frac{N+m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} + c_{\boldsymbol{K}}N^{m/2}\frac{\mathcal{B}\left(\frac{N}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} \\
&= \bar{\rho}^m + \frac{c_{\boldsymbol{K}}}{N^m}\frac{\mathcal{B}\left(\frac{N+m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} + c_{\boldsymbol{K}}N^{m/2}\frac{\mathcal{B}\left(\frac{N}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} < \infty,
\end{aligned}$$

by Lemma 2.2.1. Note that the condition 1 is needed in order to maintain the matrix $\boldsymbol{X}_t$ random and identifiable.

Thus, it remains to prove the moment bounds of $\boldsymbol{R}_t$ for every $m > 0$. We have,

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}\|\boldsymbol{a}_t\|^m\right] &= \mathbb{E}\left[\sup_{\theta\in\Theta} b_t^{3m/2}(1-b_t)^{m/2}/\nu^{m/2}\right]\mathbb{E}\left[\sup_{\theta\in\Theta}\|\boldsymbol{\Omega}^{1/2}\mathbf{u}_t\|^m\right] \\
&\leq \frac{c_{\boldsymbol{\Omega}}}{N^{m/2}}\frac{\mathcal{B}\left(\frac{N+3m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} < \infty,
\end{aligned}$$

In addition, we have

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}\|\operatorname{vec}\boldsymbol{B}_t\|^m\right] &= \mathbb{E}\left[\sup_{\theta\in\Theta}\nu^{m/2}b_t^{3m/2}(1-b_t)^{m/2}\right]\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|(\boldsymbol{\Omega}^{-1/2}\mathbf{u}_t\otimes\boldsymbol{\Omega}^{-1/2}\mathbf{u}_t\otimes\boldsymbol{\Omega}^{1/2}\mathbf{u}_t)\right\|^m\right] \\
&\leq c_{\boldsymbol{\Omega}}\mathbb{E}\left[\|\mathbf{u}_t\|^{3m}\right]\frac{\mathcal{B}\left(\frac{N+3m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} \\
&= \frac{c_{\boldsymbol{\Omega}}}{N^{3m/2}}\frac{\mathcal{B}\left(\frac{N+3m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)} < \infty,
\end{aligned}$$

and

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{D}_t\|^m\right] = \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\left[(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})^\top \otimes \boldsymbol{I}_N\right]\right\|^m\right]$$
$$\leq \left\{\sqrt{N}\bar{c}\sum_{j=0}^{\infty}\bar{\rho}^j\left(\mathbb{E}\left[\|\boldsymbol{u}_{t-j}\|^m\right]\right)^{1/m}\right\}^m < \infty,$$

by Lemma 2.3.1, and finally

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{E}_t\|^m\right] = \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\left[(\boldsymbol{u}_t)^\top \otimes \boldsymbol{I}_N\right]\right\|^m\right]$$
$$\leq N^{m/2}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{u}_t\|^m\right]$$
$$\leq c_{\boldsymbol{\Omega}} \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \nu^{m/2}\frac{\mathcal{B}\left(\frac{N+m}{2},\frac{\nu+m}{2}\right)}{\mathcal{B}\left(\frac{N}{2},\frac{\nu}{2}\right)}. < \infty,$$

by Lemma 2.2.1, which completes the proof. □

Lemma A.3.6 (Stationarity, Ergodicity and Moments for the Second Differentials of the Dynamic Location). *Consider model (2.1), (2.2) and (2.3). Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence. Consider the stochastic difference equation*

$$\mathrm{d}^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega}) = \boldsymbol{X}_t\mathrm{d}^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega}) + \boldsymbol{K}\mathrm{d}(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})^\top\boldsymbol{\mathcal{C}}_t'\mathrm{d}(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega}) + \boldsymbol{Q}_t,$$

where $\boldsymbol{X}_t$ is defined in (A.13), where $\boldsymbol{\mathcal{C}}_t$ is in (A.15), $\boldsymbol{\mathcal{C}}_t'$ in (A.31) and $\boldsymbol{Q}_t$ in (A.28),

$$\boldsymbol{Q}_t = \boldsymbol{K}\boldsymbol{a}_t'\mathrm{d}^2\nu + \boldsymbol{K}\boldsymbol{B}_t'\mathrm{d}^2\operatorname{vec}\boldsymbol{\Omega} + \boldsymbol{K}(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})^\top\widehat{\boldsymbol{aB}}_t'\mathrm{d}\nu$$
$$+ \boldsymbol{D}_t'\mathrm{d}^2\operatorname{vec}\boldsymbol{\Phi} + \boldsymbol{E}_t'\mathrm{d}^2\operatorname{vec}\boldsymbol{K} + (\mathrm{d}\operatorname{vec}\boldsymbol{\Phi})^\top\widehat{\boldsymbol{DE}}_t'(\mathrm{d}\operatorname{vec}\boldsymbol{K}),$$

with components in (A.29), (A.30), (A.32), (A.35), (A.36) and (A.37). If for some integer $t \geq 1$ and $m > 0$ the conditions

1. *$\varrho(\boldsymbol{\Phi}) < 1$ and $\det\boldsymbol{K} \neq 0$,*
2. *$\gamma(\boldsymbol{X}_t) = \left\{\frac{1}{t}\mathbb{E}\left[\log\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{X}_1\ldots\boldsymbol{X}_t\|\right]\right\} < 0$,*
3. *$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{Q}_t\|^m\right] < \infty$*

are satisfied, then the series $\{\mathrm{d}(\boldsymbol{\mu}_{t+1}-\boldsymbol{\omega})\}$ represented by

$$\mathrm{d}^2(\boldsymbol{\mu}_{t+1|t}-\boldsymbol{\omega}) = \sum_{j=0}^{t-1}\left\{\left(\prod_{k=1}^{j}\boldsymbol{X}_{t-k}\right)\left[\boldsymbol{K}\mathrm{d}(\boldsymbol{\mu}_{t-j|t-j-1}-\boldsymbol{\omega})^\top\boldsymbol{\mathcal{C}}_{t-j}'\mathrm{d}(\boldsymbol{\mu}_{t-j|t-j-1}-\boldsymbol{\omega}) + \boldsymbol{Q}_{t-j}\right]\right\}$$

converges almost surely to the unique stationary ergodic solution. Furthermore, $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\mathrm{d}(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})\|^m] < \infty$ for every $m > 0$.

Proof. Again from Lemma 2.3.1, is clear that the first two conditions are used in order to maintain true the same sufficient conditions discussed in the proof of Lemma in order to obtain ergodic sequences $\{X_t\}$, see Vervaat, 1979 and Basrak, Davis, and Mikosch, 2002. Hence, the Hölder and Minkowsky inequalities imply and the independence between each component

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathrm{d}^2(\mu_{t+1|t}-\omega)\|^m\right] \leq \sum_{j=0}^{t-1}\left\{\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\prod_{k=0}^{j}X_{t-k}\right\|^m\right]^{1/m}\right.$$
$$\times\left(c_K\mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathrm{d}(\mu_{t-j|t-j-1}-\omega)\|^{2m}\right]^{1/2m}\mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathcal{C}'_{t-j}\|^m\right]^{1/m}\right.$$
$$\left.\left.+\mathbb{E}\left[\sup_{\theta\in\Theta}\|Q'_{t-j}\|^m\right]^{1/m}\right)\right\},$$

from which we can see that by using Lemma 2.3.1 the first two elements are bounded while the third is the second differential of the driving force with respect to the dynamic location vector. From (A.31) in the same vein of Lemma A.3.5, the c_r-inequality establishes that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\|\mathcal{C}'_t\|^m\right]$$
$$\leq\mathbb{E}\left[\sup_{\theta\in\Theta}[8(1-b_t)^3/\nu^2]^m\right.$$
$$\left.\times\left\|\left\{\left[I_N\otimes(y_t-\mu_{t|t-1})(y_t-\mu_{t|t-1})^\top\right]\mathrm{vec}\,\Omega^{-1}\right\}\right\|^m\left\|\left[(y_t-\mu_{t|t-1})^\top\Omega^{-1}\right]\right\|^m\right]$$
$$+\mathbb{E}\left[\sup_{\theta\in\Theta}[2(1-b_t)^2/\nu]^m\right.$$
$$\left.\times\left\|\left\{\left[\Omega^{-1}\otimes I_N\right]\left[(y_t-\mu_{t|t-1})\otimes I_N+I_N\otimes(y_t-\mu_{t|t-1})\right]\right\}\right\|^m\right]$$
$$+\mathbb{E}\left[\sup_{\theta\in\Theta}[2(1-b_t)^2/\nu]^m\left\|\left\{\left[\Omega^{-1}\otimes I_N\right]\left[(y_t-\mu_{t|t-1})\otimes I_N\right]\right\}\right\|^m\right]$$
$$\leq\bar{C}_4\mathbb{E}\left[\|\mathbf{u}_t\|^{3m}\right]+\bar{C}_3\mathbb{E}\left[\|\mathbf{u}_t\|^{2m}\right]+\bar{C}_3\mathbb{E}\left[\|\mathbf{u}_t\|^{2m}\right]<\infty.$$

It is possible to show with some tedious calculations the analogous results for each component of Q_t, which means that we further satisfy condition 3 □

Lemma A.3.7 (Invertibility for the First and Second Differentials of the Dynamic Location Filter). *Consider model (2.1), (2.2) and (2.3). Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence. Let the conditions of Lemma 2.3.1 hold true and consider further the filtering equation (2.7). Under the condition of Lemmata A.3.5, A.3.6 and 2.3.3 we have that for any initialization of the filter $(\mu_{1|0}-\omega)$ and its first differential $\mathrm{d}(\mu_{1|0}-\omega)$, the perturbed first differential and second differential of the dynamic location filter, namely $\{\mathrm{d}(\hat{\mu}_{t|t-1}-\omega)\}_{t\in\mathbb{N}}$ and $\{\mathrm{d}^2(\hat{\mu}_{t|t-1}-\omega)\}_{t\in\mathbb{N}}$ converge exponentially fast almost surely to the unique stationary ergodic solution $\{\mathrm{d}(\mu_{t|t-1}-\omega)\}_{t\in\mathbb{Z}}$ and $\{\mathrm{d}(\mu_{t|t-1}-\omega)\}_{t\in\mathbb{Z}}$.*

Furthermore, for any $m>0$

1. $\mathbb{E}[\sup_{\theta\in\Theta}\|\mathrm{d}(\hat{\mu}_{t|t-1}-\omega)\|^m]<\infty$ *and* $\mathbb{E}[\sup_{\theta\in\Theta}\|\mathrm{d}^2(\hat{\mu}_{t|t-1}-\omega)\|^m]<\infty$,
2. $\mathbb{E}[\sup_{\theta\in\Theta}\|\mathrm{d}(\mu_{t|t-1}-\omega)\|^m]<\infty$ *and* $\mathbb{E}[\sup_{\theta\in\Theta}\|\mathrm{d}^2(\mu_{t|t-1}-\omega)\|^m]<\infty$.

Proof. We provide a detailed formal discussion for the first case, that is the convergence of the perturbed first differential since the proof for the second case, the convergence of the perturbed second is basically the same and is skipped.

The proof of this Lemma builds upon the arguments of Theorem 2.10 in Straumann and Mikosch, 2006 for perturbed stochastic recurrence equations. In particular, the perturbed stochastic recurrence equation correspond to

$$\mathrm{d}(\hat{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\omega}) = \widehat{\boldsymbol{X}}_t \mathrm{d}(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega}) + \widehat{\boldsymbol{R}}_t,$$

which is a nonlinear function of the started sequence $\{(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega})\}_{t\in\mathbb{N}}$. Under maintained assumptions, we note that the relevant contraction condition 2 of Lemma 2.3.3 hold true and thus, the desired convergence of the recurrence equation is obtained if

$$\|\widehat{\boldsymbol{X}}_t - \boldsymbol{X}_t\| \xrightarrow{\text{e.a.s.}} 0 \qquad \textit{and} \qquad \|\widehat{\boldsymbol{R}}_t - \boldsymbol{R}_t\| \xrightarrow{\text{e.a.s.}} 0 \qquad \textit{as} \qquad t \to \infty. \tag{A.42}$$

In order to verify these convergences, we invoke the mean value theorem which yields the following inequalities

$$\|\widehat{\boldsymbol{X}}_t - \boldsymbol{X}_t\| \le \sup_{\boldsymbol{\mu}} \|\boldsymbol{\mathcal{C}}'_t\| \, \|\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}\|, \tag{A.43}$$

and

$$\|\widehat{\boldsymbol{R}}_t - \boldsymbol{R}_t\| \le \sup_{\boldsymbol{\mu}} \left\| \begin{matrix} \boldsymbol{\mathcal{C}}'_t \\ \widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t \\ \widehat{\boldsymbol{a}\boldsymbol{\mathcal{C}}}'_t \end{matrix} \right\| \|\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}\|,$$

where the expression for $\boldsymbol{\mathcal{C}}'_t$, $\widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t$ and $\widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t$ can be found in (A.31), (A.33) and (A.34) respectively. We can combine the results obtained in Lemma A.3.6, the moment bounds obtained in Lemma 2.3.1 together with the almost sure exponentially fast convergence (A.3) in Lemma 2.3.3, in order to achieve the required convergences in (A.42). Indeed, in the same vein of Lemma A.3.5 we can show by direct (tedious) calculations that the property of uniformly boundedness expands easily to each these derivatives, since they can be rewritten as functions of beta distributed random variables and random vectors uniformly distributed on the unit sphere surface.

As a matter of fact, we obtain

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\mathcal{C}}'_t\| = O_p(1), \quad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \begin{matrix} \boldsymbol{\mathcal{C}}'_t \\ \widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t \\ \widehat{\boldsymbol{a}\boldsymbol{\mathcal{C}}}'_t \end{matrix} \right\| = O_p(1) \quad \textit{and} \quad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}\| = o_{e.a.s.}(1) \quad \textit{as} \quad t \to \infty.$$

Thus, repeated applications of Lemma 2.1 in Straumann and Mikosch, 2006 ensure the required convergence in (A.43).

Summing up, under maintained conditions, it is ensured that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathrm{d}(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega}) - \mathrm{d}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})\| \xrightarrow{\text{e.a.s.}} 0 \qquad \textit{as} \qquad t \to \infty.$$

Since the sequence $\{\mathrm{d}^2(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\}$ is a nonlinear function of both the perturbed recurrence $\{\mathrm{d}(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\}$ and the filter $\{(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\}$ the same arguments apply sequentially, yielding

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\mathrm{d}^2(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})-\mathrm{d}^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})\|\xrightarrow{\text{e.a.s.}}0 \qquad \textit{as} \qquad t\to\infty.$$

As regards the second claim for the moment bounds, it is a consequence of the Continuous Mapping Theorem, since the differentials are nonlinear continuous functions of $\boldsymbol{\mu}_{t|t-1}$, which has unbounded number of moments as we proved in Lemma 2.3.3. □

Lemma A.3.8 (Martingale Difference Property of the Likelihood's First Differential). *Consider model (2.1), (2.2) and (2.3). Let $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence and, further, suppose that*

1. $\varrho(\boldsymbol{\Phi})<1$ *and* $\det\boldsymbol{K}\neq 0$,
2. $\gamma(\boldsymbol{X}_t)=\left\{\frac{1}{t}\mathbb{E}\left[\log\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{X}_1\ldots\boldsymbol{X}_t\|\right]\right\}<0$,
3. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{R}_t\|^2\right]<\infty$,
4. *the parameter space* $\boldsymbol{\Theta}$ *is compact with* $2<\nu<\infty$.

Then, the first differential of the likelihood $\{\ell_t(\boldsymbol{\theta})\}$ is a martingale difference sequence with finite second moments, that is

$$\mathbb{E}_{t-1}[\ell_t(\boldsymbol{\theta})]=0 \qquad \textit{and} \qquad \mathbb{E}[|\ell_t(\boldsymbol{\theta})|^2]<\infty.$$

Proof. The first differential of the log-likelihood contribution at time t with respect to the parameter vector $\boldsymbol{\theta}$ is given by (A.8), that is

$$\begin{aligned}\mathrm{d}\ell_t(\boldsymbol{\theta})=&\frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{N}{\nu}+\frac{\nu+N}{\nu}b_t-\ln w_t\right](\mathrm{d}\nu)\\&+\frac{1}{2}(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}))^\top\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\left[\frac{\nu+N}{\nu}\frac{1}{w_t}(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)-\mathrm{vec}\,\boldsymbol{I}_N\right]\\&+\frac{\nu+N}{\nu}\frac{1}{w_t}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1}),\end{aligned}$$

with $\psi(\cdot)$ the digamma function, b_t and w_t are in (2.4) and (2.5) and $\{\mathrm{d}\boldsymbol{\mu}_{t|t-1}\}$ is the first differential of the filtered location sequence $\{\boldsymbol{\mu}_{t|t-1}\}$ given by (A.11) and (A.12). We have that

$$\begin{aligned}\mathbb{E}_{t-1}[\mathrm{d}\ell_t(\boldsymbol{\theta})]=&\frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{N}{\nu}+\frac{\nu+N}{\nu}\mathbb{E}_{t-1}[b_t]-\mathbb{E}_{t-1}[\ln w_t]\right](\mathrm{d}\nu)\\&+\frac{1}{2}(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}))^\top\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\left[\frac{\nu+N}{\nu}\mathbb{E}_{t-1}[(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)/w_t]-\mathrm{vec}\,\boldsymbol{I}_N\right]\\&+\frac{\nu+N}{\nu}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}\mathbb{E}_{t-1}[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})/w_t],\end{aligned}$$

since $\mathrm{d}\boldsymbol{\mu}_{t|t-1}$ is $\mathscr{F}_{t-1}$-measurable. Straightforward calculations give us

$$\begin{aligned}
\mathbb{E}_{t-1}[b_t] &= \frac{N}{\nu+N}, \\
\mathbb{E}_{t-1}[\ln(1/w_t)] = \mathbb{E}_{t-1}[\ln(1-b_t)] &= \psi\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu+N}{2}\right), \\
\mathbb{E}_{t-1}[(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)/w_t] = \nu\mathbb{E}_{t-1}[(\mathbf{u}_t \otimes \mathbf{u}_t)]\mathbb{E}_{t-1}[b_t] &= \frac{\nu}{\nu+N}\operatorname{vec}\boldsymbol{I}_N, \\
\mathbb{E}_{t-1}[(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/w_t] = \sqrt{\nu}\mathbb{E}_{t-1}[\sqrt{b_t(1-b_t)}]\,\boldsymbol{\Omega}^{1/2}\mathbb{E}_{t-1}[\mathbf{u}_t] &= \mathbf{0},
\end{aligned}$$

where the first equality follows from the property the beta distribution and as Lemma 2.2.1 shows $b_t \sim \mathcal{B}eta\left(\frac{N}{2}, \frac{\nu}{2}\right)$, which further imply the second result. The third and the fourth can be obtained by taking advantage from the different stochastic representations of the model that was given in (2.10) and (2.11). Thus, it is clear that after simple substitutions we obtain the claimed martingale difference property $\mathbb{E}_{t-1}[\mathrm{d}\ell_t(\boldsymbol{\theta})]$.

Now we can concentrate on the second claim. We simply note that the condition imposed on the parameter space with $\nu > 2$ ensures the existence of the second moment of the innovations $\boldsymbol{\epsilon}_t$ exists by the property of the multivariate Student's t and by Lemmata 2.2.1, 2.3.1, A.3.1 and the Continuous Mapping Theorem. □

Lemma A.3.9 (CLT for the Likelihood's First Differential). *Consider model (2.1), (2.2) and (2.3). Let $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence and, further, suppose that*

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,
2. $\gamma(\boldsymbol{X}_t) = \left\{\frac{1}{t}\mathbb{E}\left[\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{X}_1 \ldots \boldsymbol{X}_t\|\right]\right\} < 0$,
3. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{R}_t\|^2\right] < \infty$,
4. *the parameter space* $\boldsymbol{\Theta}$ *is compact with* $2 < \nu < \infty$.

Then, the first differential of the likelihood $\{\ell_t(\boldsymbol{\theta})\}$ obeys the CLT for martingale difference sequences, that is

$$\sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}) \overset{\mathscr{D}}{\Rightarrow} \mathscr{N}(\mathbf{0}, \boldsymbol{V}) \qquad \text{as} \quad t \to \infty, \text{ where} \qquad \boldsymbol{V} = \mathbb{E}\left[\frac{d\mathcal{L}'_T(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\frac{d\mathcal{L}'_T(\boldsymbol{\theta})}{d\boldsymbol{\theta}^\top}\right] \qquad \text{as} \quad t \to \infty.$$

Proof. To see this result, it suffices to consider Lemma A.3.8, where the relevant properties of the likelihood's first differential are obtained. Hence, with the support of the Cramér-Wold device (see Vaart, 1998 pag. 16) it is possible to immediately apply the CLT for martingales of Billingsley, 1961 on the linear combination

$$\sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}) = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} \overset{\mathscr{D}}{\Rightarrow} \mathscr{N}(\mathbf{0}, \boldsymbol{V}).$$

□

Lemma A.3.10 (Almost sure convergence of the Likelihood's First Differential). *Consider model (2.1), (2.2) and (2.3). Let $\{\boldsymbol{\epsilon}_t\}_{t\in\mathbb{Z}}$ be a stationary and ergodic sequence. Assume that*

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det \boldsymbol{K} \neq 0$,
2. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\prod_{k=1}^{j} \frac{\partial\boldsymbol{\mu}_{t-k+1|t-k}}{\partial\boldsymbol{\mu}^\top_{t-k|t-k-1}}\right\|\right] < 1$, *for some* $j \geq 1$ *large enough and*

3. $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\mu_{1|0}}{\partial\theta^{\top}}\right\|\right]<\infty$,

4. *the parameter space* Θ *is compact with* $2<\nu<\infty$,

5. *the true parameter vector* $\theta_0\in\Theta$ *and*

6. $\forall\theta\in\Theta$*, if* $\theta\neq\theta_0$ *then* $\mu_{t|t-1}\neq\mu_{0t}$ *almost surely and* $\forall t\geq 1$.

Then,

$$\sqrt{T}|\mathrm{d}\widehat{\ell}_T(\theta_0)-\mathrm{d}\ell_T(\theta_0)|\xrightarrow{a.s.}0\qquad as\qquad T\to\infty,$$

where $\mathrm{d}\widehat{\ell}_T(\theta)$ *is the sum of the t-th contributions to the likelihood's first differential started with* $\mu_{1|0}$*, while* $\mathrm{d}\ell_T(\theta)$ *is the unique stationary ergodic counterpart.*

Proof. We can demonstrate the above almost sure convergence by appealing again to the invertibility property of the location filter proved in Lemma 2.3.3 and its differentials, established in Lemma A.3.7. As already noted, this property will also ensure that the perturbed first differential of the dynamic location will converge to its unique stationary ergodic solution. Summing up we have that,

$$\|\hat{\mu}_{0,t|t-1}-\mu_{0,t|t-1}\|\xrightarrow{\text{e.a.s.}}0\qquad and\qquad \|\mathrm{d}\hat{\mu}_{0,t|t-1}-\mathrm{d}\mu_{0,t|t-1}\|\xrightarrow{\text{e.a.s.}}0\qquad as\qquad t\to\infty.$$

Hence, we can rely on a multivariate mean value expansion around all the element of the vectors $\mu^{\star}_{t|t-1}$ and $\mathrm{d}\mu^{\star}_{t|t-1}$, which are on the chords between $[\hat{\mu}_{t|t-1},\mu_{t|t-1}]$ and $[\mathrm{d}\hat{\mu}_{t|t-1},\mathrm{d}\mu_{t|t-1}]$ respectively, yielding

$$|\mathrm{d}\widehat{\ell}_T(\theta_0)-\mathrm{d}\ell_T(\theta_0)|\leq\sup_{(\mu,\mathrm{d}\mu)}\left\|\begin{matrix}\frac{\partial(\mathrm{d}\widehat{\ell}_T(\theta_0))}{\partial\mu^{\star\top}_{t|t-1}}\\ \frac{\partial(\mathrm{d}\widehat{\ell}_T(\theta_0))}{\partial(\mathrm{d}\mu^{\star\top}_{t|t-1})})\end{matrix}\right\|\left\|\begin{matrix}(\hat{\mu}_{0,t|t-1}-\mu_{0,t|t-1})\\(\mathrm{d}\hat{\mu}_{0,t|t-1}-\mathrm{d}\mu_{0,t|t-1})\end{matrix}\right\|.$$

Note that the first term on the right hand of the inequality is at least uniformly bounded and hence we can demonstrate the exponentially fast almost sure convergence of the term in the left hand side by appealing to Lemma 2.1 in Straumann and Mikosch, 2006 since the convergence at the same rate of second term in the right hand side as an application of Lemma A.3.7. □

Lemma A.3.11 (Properties of the Likelihood's Second Differential). *Consider model* (2.1)*,* (2.2) *and* (2.3)*. Let* $\{\epsilon_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence and, further, suppose that*

1. $\varrho(\Phi)<1$ *and* $\det K\neq 0$,

2. *the parameter space* Θ *is compact with* $2<\nu<\infty$,

3. $\gamma(X_t)=\left\{\frac{1}{t}\mathbb{E}\left[\log\sup_{\theta\in\Theta}\|X_1\ldots X_t\|\right]\right\}<0$,

4. $\mathbb{E}\left[\sup_{\theta\in\Theta}\|R_t\|^2\right]<\infty$ *and*

5. $\forall\theta\in\Theta$*, if* $\theta\neq\theta_0$ *then* $\mu_{t|t-1}\neq\mu_{0,t|t-1}$ *almost surely and* $\forall t\geq 1$.

Then, the second differential of the likelihood $\{\mathrm{d}^2\ell_t(\boldsymbol{\theta})\}$ *is stationary ergodic with at least two finite moments. In particular, we have that*

$$\mathbb{E}[\mathrm{d}^2\ell_t(\boldsymbol{\theta})] < \infty,$$

and nonsingular.

Proof. Since the complete equation of the second differential is more subtle than the first, we leave it in (A.22) and, for the sake of completeness we report a a general case, where we prove the arguments in a more clear and concise fashion in terms of derivatives rather then differentials. Elementary matrix calculus give us the following result

$$\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top} = \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} + \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)^\top \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right) + \frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}^\top}\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}.$$

Applying the expectation operator we get a finite and static term in the first addend of the right hand side while by using the independence and the martingale difference properties of the score vector the last term becomes null. Thus, we can focus our attention on the middle term. Define

$$\boldsymbol{\mathcal{I}}^{(\boldsymbol{\mu}_{t|t-1})}(\boldsymbol{\theta}) = -\,\mathbb{E}\left[\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)^\top \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right].$$

Note that, by the independence property, we expresse its vectorized counterpart as

$$\operatorname{vec}\boldsymbol{\mathcal{I}}^{(\boldsymbol{\mu}_{t|t-1})}(\boldsymbol{\theta}) = \mathbb{E}\left[\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\otimes\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right]^\top \operatorname{vec}\boldsymbol{\mathcal{I}}^{(\boldsymbol{\mu})}(\boldsymbol{\theta}).$$

We know from Lemmata 2.3.2, 2.3.1, A.3.5, A.3.6, 2.3.3, A.3.7 that the dynamic location filter and its differentials, are invertible and achieve their own unique stationary ergodic solution with an unbounded numer of finite moments. Thus, we obtain the desired result by repeated applications of the Law of iterated expectation (LIE) to the following equivalence

$$\begin{aligned}
&\mathbb{E}_{t-1}\left[\left(\frac{d(\boldsymbol{\mu}_{t+1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\otimes\left(\frac{d(\boldsymbol{\mu}_{t+1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right]^\top\\
&\quad=\mathbb{E}_{t-1}\left[\left(\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}+\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\otimes\left(\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}+\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\right]^\top\\
&\quad=\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\otimes\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)^\top\mathbb{E}_{t-1}\left[\left(\boldsymbol{X}_t\otimes\boldsymbol{X}_t\right)\right]^\top+\mathbb{E}_{t-1}\left[\left(\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\otimes\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\right]^\top\\
&\qquad+\mathbb{E}_{t-1}\left[\left(\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\otimes\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\right]^\top+\mathbb{E}_{t-1}\left[\left(\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\otimes\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right]^\top.
\end{aligned}$$

Note that the Lyapunov condition 3 is more that enough to ensure the stability of the recursions. □

Lemma A.3.12 (**Uniform convergence of the likelihood's second derivatives**). *Consider model* (2.1), (2.2) *and* (2.3). *Let* $\{\epsilon_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence. Assume that*

1. $\varrho(\boldsymbol{\Phi}) < 1$ *and* $\det\boldsymbol{K} \neq 0$,
2. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\prod_{k=1}^{j}\frac{\partial\boldsymbol{\mu}_{t-k+1|t-k}}{\partial\boldsymbol{\mu}_{t-k|t-k-1}^\top}\right\|\right] < 1$, *for some* $j \geq 1$ *large enough and*

3. $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\mu_{1|0}}{\partial\theta^\top}\right\|\right] < \infty$,

4. *the parameter space* Θ *is compact with* $2 < \nu < \infty$.

Then,

$$\sup_{\theta\in\Theta} |\widehat{\mathcal{L}}''_T(\theta) - \mathcal{L}''_T(\theta)| \xrightarrow{a.s.} 0 \qquad as \qquad t\to\infty,$$

where $\widehat{\mathcal{L}}''_T(\theta)$ *is the empirical likelihood's second derivatives started with* $\mu_{1|0}$ *and* $\mathcal{L}''(\theta)$ *is the unique stationary ergodic counterpart.*

Proof. We note that the second derivatives of the likelihood is a nonlinear map of the filtered location vector and its first end second differential and hence, the mean value theorem has to be applied also for each dynamic equation. As a result,

$$\sup_{\theta\in\Theta} \|\widehat{\mathcal{L}}''_T(\theta) - \mathcal{L}''_T(\theta)\| \le \sup_{\theta\in\Theta}\left\|\begin{matrix}\frac{\partial\widehat{\mathcal{L}}''_T(\theta)}{\partial\mu^{\star\top}_{t|t-1}}\\ \frac{\partial\widehat{\mathcal{L}}''_T(\theta)}{\partial(\mathrm{d}\mu^{\star\top}_{t|t-1})}\\ \frac{\partial\widehat{\mathcal{L}}''_T(\theta)}{\partial(\mathrm{d}^2\mu^{\star\top}_{t|t-1})}\end{matrix}\right\| \sup_{\theta\in\Theta}\left\|\begin{matrix}(\hat{\mu}_{t|t-1}-\mu_{t|t-1})\\ (\mathrm{d}\hat{\mu}_{t|t-1}-\mathrm{d}\mu_{t|t-1})\\ (\mathrm{d}^2\hat{\mu}_{t|t-1}-\mathrm{d}^2\mu_{t|t-1})\end{matrix}\right\|.$$

Thus, the proof can proceed following the same arguments of the proof of Lemma A.3.10, obtaining the uniformly boundedness of the first term and the exponentially fast convergence of the second term in the right hand side. Notice also, the last component of the first term in the right hand side involve a third order differential, which can be found in (A.41) and is uniformly bounded. The tools developed in Lemma 2.1 of Straumann and Mikosch, 2006 conclude the proof. □

Lemma A.3.13 (Uniform Law of Large Numbers of the likelihood). *Consider model* (2.1), (2.2) *and* (2.3). *Let* $\{\epsilon_t\}_{t\in\mathbb{Z}}$ *be a stationary and ergodic sequence. Assume that*

1. $\varrho(\Phi) < 1$ *and* $\det K \neq 0$,

2. *the parameter space* Θ *is compact with* $2 < \nu < \infty$,

3. *the true parameter vector* $\theta_0 \in \Theta$ *and*

4. $\forall\theta\in\Theta$, *if* $\theta\neq\theta_0$ *then* $\mu_{t|t-1}\neq\mu_{0,t|t-1}$ *almost surely and* $\forall t\ge 1$.

Then,

$$\sup_{\theta\in\Theta} |\mathcal{L}''_T(\theta) - \mathcal{L}''(\theta)| \xrightarrow{a.s.} 0 \qquad as \qquad t\to\infty,$$

where $\mathcal{L}_T(\theta)$ *is the stationary ergodic empirical likelihood and* $\mathcal{L}(\theta)$ *is the limit likelihood defined in* (A.5) *and* (A.6) *respectively.*

Proof. Since $\mathcal{L}''_T(\theta)$ is a function of $\{y_t, y_{t-1}, \dots\}$, therefore, under mantained assumptions, stationary and ergodic. The Lemma follows straightforwardly from Lemma A.3.11 and the The Uniform Law of Large Numbers in its version for ergodic stationary processes, see for example Theoerm A.2.2 in White, 1994 as in Lemma A.3.4. □

Appendix B

Proofs of Chapter 3

B.1 Proofs

Proof. We note that by making the substitution $g(z_t) = \delta_i$ and $c(z_t) = \phi_i + \kappa_i[(\nu_i+1)\epsilon_{it}^2]/[(\nu_i-2)+\epsilon_{it}^2] - 1$ we can embed the model to the more general class of *GARCH* processes considered by Ling and McAleer, 2002, that is

$$\sigma_{i,t+1}^2 = g(\epsilon_{it}) + c(\epsilon_{it})\sigma_{it}^2.$$

Thus, relying on Theorem 2.1 of Ling and McAleer, 2002, if $\mathbb{E}[c(\epsilon_{it})] < 1$, then there exist a unique stationary and ergodic solution. In our case, it is easy to see that

$$\mathbb{E}[c(\epsilon_{it})] = \mathbb{E}\left[\phi_i + \kappa_i\left(\frac{(\nu_i+1)\epsilon_{it}^2}{(\nu_i-2)+\epsilon_{it}^2} - 1\right)\right] = \phi_i < 1.$$

The second equality follows directly from the properties of the Student's *t* distribution, in fact the random variable

$$b_{it} = \frac{\epsilon_{it}^2}{(\nu_i-2)+\epsilon_{it}^2}, \tag{B.1}$$

is distributed as a beta random variable with shape and scale parameters of $1/2$ and $\nu_i/2$ respectively and its expected value is $1/(\nu_i+1)$, see Harvey, 2013 for more details. The last inequality follows trivially by assumptions. In conclusion, by recursive arguments, it is easy to see that we could also achieve the following almost sure representation of the process,

$$\sigma_{i,t+1}^2 = g(\epsilon_{it}) + g(\epsilon_{it})\sum_{k=1}^{\infty}\prod_{i=0}^{k} c(\epsilon_{it-1}). \tag{B.2}$$

It is important to note that $\{\epsilon_{it}\}_{t\in\mathbb{Z}}$ forms an *IID* sequence of positive random variables.

Having verified that under maintained assumption the *Beta-t-GARCH* generates stationary ergodic paths, we can show that the same holds for the common factor process. To be specific, we consider the same recursion

$$f_{t+1}^2 = g(\epsilon_{it}) + c(\epsilon_{it})f_t^2, \tag{B.3}$$

where $g(\epsilon_{it}) = \omega$ and $c(\epsilon_{it}) = \beta + \alpha(1/N\sum_{i=1}^N \sigma_{it}^2\epsilon_{it}^2 - 1)$. Again, also in this case $\{\epsilon_{it}\}_{t\in\mathbb{Z}}$ forms an *IID* sequence.

Thus, it is clear that

$$\mathbb{E}\left[\beta + \alpha\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_{it}^2\epsilon_{it}^2 - 1\right)\right] = \beta + \alpha\left(\frac{\delta_i}{1-\phi_i} - 1\right) = \beta < 1,$$

where the first equality follow by independence between the idiosyncratic processes and the innovations and then, by unfolding and recalling that $\delta_i = 1 - \phi_i$ for $i = 1, \ldots, N$, the second equality follows. Again, the inequality is ensured by assumption and the implied almost sure representation is the analogous version of (B.2).

In conclusion, Assumption 2, ensure that the sequence composed by $\{(f_t^2, (\sigma_t^2)^\top, \epsilon_t^\top)\}$ converges to the unique stationary ergodic solution $\{(f_t^2, (\sigma_t^2)^\top, \epsilon_t^\top)\}_{t\in\mathbb{Z}}$. Thus, the same old true for $\{(f_t\sigma_{it})^2\}_{t\in\mathbb{Z}}$ and $\{f_t\sigma_{it}\epsilon_{it}\}_{t\in\mathbb{Z}}$ for $i = 1, \ldots, N$ by Proposition 3.36 of White, 2001, implying that the series $\{x_{it}\}_{t\in\mathbb{Z}}$ is second-order stationary. □

Proof of Proposition 2

Proof. By similar arguments as those in the Proof of Prooposition 1, if $\mathbb{E}[z_t^j] < \infty$, we obtain necessary and sufficient conditions for the existence of the moments of the *Beta-t-GARCH* and the common factor process, that is

$$\mathbb{E}[c(z_t)^{j/2}] < 1 \qquad \textit{for} \quad j = 2, 4, \ldots,$$

by Theorem 2.2 of Ling and McAleer, 2002.

To see this, consider the case where $d \in [1, \infty)$, the almost sure representation obtained in equation (B.2) in the Proof of Proposition 1 and recall that $\{z_t\}_{t\in\mathbb{Z}}$ forms an *IID* sequence of positive random variables. Then, by virtue of the Minkowsky's inequality, we have

$$\begin{aligned}\mathbb{E}[\sigma_{i,t+1}^{2d}] &\leq g(z_t)^d + g(z_t)^d \mathbb{E}\left[\sum_{k=1}^{\infty}\prod_{i=0}^{k} c(z_{t-i})^d\right] \\ &\leq g(z_t)^d + g(z_t)^d \sum_{k=1}^{\infty}\left(\mathbb{E}[c(z_{t-i})^d]\right)^{k/d},\end{aligned}$$

thus, if $\mathbb{E}[c(z_t)^{j/2}] < 1$ for $j = 2, 4, \ldots$, we obtain $\mathbb{E}[\sigma_{i,t+1}^j] < \infty$ since it satisfies the Cauchy criteria.

Now, we can easily show that an analogous result hold true for the factor process $\{f_{t+1}^2\}$, since it admits the same representation, displayed in equations (B.3). Thus, following the same recursive argument we get that $\mathbb{E}[f_{t+1}^j] < \infty$.

To conclude the proof, we will prove how the formula in (3.7) may be merely obtained as a straightforward application of the generalized Hölder's inequality.

Suppose that

$$\frac{1}{m} = \frac{1}{n_f} + \frac{1}{n_\sigma} + \frac{1}{n_\epsilon} = \frac{n_\sigma n_\epsilon + n_f n_\epsilon + n_f n_\sigma}{n_f n_\sigma n_\epsilon}.$$

Then, the generalized Hölder's inequality implies that

$$\|x_{it}\|_m = \|f_t\sigma_{it}\epsilon_{it}\|_m \leq \|f_t\|_{n_f}\|\sigma_{it}\|_{n_\sigma}\|\epsilon_{it}\|_{n_\epsilon}.$$

By the arguments above, we know that $\|f_t\|_{n_f} < \infty$, $\|\sigma_{it}\|_{n_\sigma} < \infty$ and $\|\epsilon_{it}\|_{n_\epsilon} < \infty$ for some n_f, n_σ and n_ϵ for $i = 1, \ldots, N$. In particular, this means that

$$f_t \in L^{n_f}, \qquad \sigma_{it} \in L^{n_\sigma}, \qquad \epsilon_{it} \in L^{n_\epsilon},$$

and as a matter of fact

$$(f_t \times \sigma_{it} \times \epsilon_{it}) \in L^m.$$

Therefore, we obtain the desired number of moments

$$m = \frac{n_f n_\sigma n_\epsilon}{n_\sigma n_\epsilon + n_f n_\epsilon + n_f n_\sigma}.$$

□

Proof of Proposition 3

Proof. We start with the filter for the common factor. The linearity of the recursion makes the evaluation of invertibility particularly easy. In fact, by repeated substitution it is possible to see that the filtered common factor

$$\hat{f}^2_{t+1} = \frac{\omega}{1-(\beta-\alpha)} + \alpha \frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{t-1} (\beta-\alpha)^j x^2_{i,t-j} + (\beta-\alpha)^{t-1} \hat{f}^2_1, \tag{B.4}$$

where $\hat{f}^2_1$ is some initial value. It is clear, then, that the condition $|\beta - \alpha| < 1$ ensures that asymptotically, as $t \to \infty$, the impact of this initial value vanishes and hence the filter is invertible. In particular, $\hat{f}_t$ converges to limit

$$f^2_{t+1} = \frac{\omega}{1-(\beta-\alpha)} + \alpha \frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{\infty} (\beta-\alpha)^j x^2_{i,t-j}. \tag{B.5}$$

We now consider the perturbed stochastic recurrence equation

$$\hat{\sigma}^2_{i,t+1} = \hat{\Phi}_t(\hat{\sigma}^2_{it}),$$

which is a perturbed version of the filtered $\hat{\sigma}^2_{i,t+1} = \Phi_t(\hat{\sigma}^2_{it})$ since $\hat{\Phi}_t$ depends on the nonstatonary filtered factor $\hat{f}^2_t$ whereas Φ_t depends on the limit stationary filtered factor f^2_t. The unperturbed version takes the limit $\{f^2_t\}_{t\in\mathbb{Z}}$ and generate $\hat{\sigma}^2_{i,t+1} = \Phi_t(\hat{\sigma}^2_{it})$. We verify the condition of Theorem 2.10 in Straumann and Mikosch, 2006 and proceed step-by-step.

First, the condition of the Bougerol's Theorem must hold for the unperturbed recurrence equation. To start, we note that $\{x_{it}\}$ and $\{f^2_t\}$ are stationary and ergodic and the initial log-moment condition is satisfied since

$$\begin{aligned}
&\mathbb{E}\left[\log^+ \left|\delta_i + \kappa_i \frac{(\nu_i+1)(x^2_{it}/f^2_t)}{(\nu_i-2)\bar{\sigma}^2_i + (x^2_{it}/f^2_t)}\right] \bar{\sigma}^2_i\right|\Big] \\
&\quad \le \log^+|\delta_i| + \log^+|\phi_i - \kappa_i| + \log^+|\kappa_i| \\
&\qquad \mathbb{E}\left[\log^+ \left|\frac{(\nu_i+1)(x^2_{it}/f^2_t)}{(\nu_i-2)\bar{\sigma}^2_i + (x^2_{it}/f^2_t)}\right] \bar{\sigma}^2_i\right|\Big] + \log^+|\bar{\sigma}^2_i|,
\end{aligned}$$

and we can see that

$$\mathbb{E}\left[\log \left|\frac{(\nu_i+1)(x^2_{it}/f^2_t)}{(\nu_i-2)\bar{\sigma}^2_i + (x^2_{it}/f^2_t)}\right] \bar{\sigma}^2_i\right|\Big] < \infty$$

since it is uniformly bounded in $\bar{\sigma}^2_i \ge \delta_i = (1-\phi_i) > 0$ implied by $|\phi| < 1$ from assumption 4, and uniformly bounded for every $x_{it} \in \mathbb{R}$. We also require that the contraction condition of the unperturbed recurrence equation satisfy

$$\mathbb{E}\left[\log^+ \left|\frac{\partial\Phi(\sigma^2_{it})}{\partial\sigma^2_{it}}\right|\right] < 0,$$

which is verified from the fact that

$$\frac{\partial \Phi(\sigma_{it}^2)}{\partial \sigma_{it}^2} = (\phi - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^4/f_t^4)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2},$$

and therefore

$$\sup_{\sigma_i^{2\star}} \left\| (\phi - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^4/f_t^4)}{[(\nu_i - 2)\sigma_i^{2\star} + (x_{it}^2/f_t^2)]^2} \right\| \leq \max_{i=1,\dots,N} \left\{ |\phi - \kappa_i|; |\phi + \kappa_i \nu_i| \right\} < 1,$$

because, as before $\sigma_i^{2\star} \geq \delta_i = (1 - \phi_i) > 0$ by assumption 4 and $f_t \geq \omega > 0$. Moreover, is uniformly bounded for every $x_{it} \in \mathbb{R}$ and then, the result follows from assumption 5. In practice, the contraction condition will ensure that

$$\sup_{\theta \in \Theta} \|\hat{\sigma}_{it}^2 - \sigma_t^2\| \xrightarrow{\text{e.a.s.}} 0 \qquad as \qquad t \to \infty.$$

Second, we verify the logarithmic moment for the stationary solution $\{\sigma_{it}^2\}_{t \in \mathbb{Z}}$, which is cleary implied by Proposition 3.7, in fact

$$\mathbb{E}[\sigma_{i,t+1}^j] = \mathbb{E}[\Phi(\sigma_{it}^2)^j] < \infty$$

for $j = 2, 4, \dots$.

Third, the perturbed stochastic recurrence equation $\hat{\sigma}_{i,t+1}^2 = \hat{\Phi}_t(\hat{\sigma}_{it}^2)$, must converges to the unperturbed counterpart, that is verified as an application of the mean value theorem. We have

$$\sup_{\theta \in \Theta} \|\hat{\Phi}_t(\hat{\sigma}_{it}^2) - \Phi_t(\hat{\sigma}_{it}^2)\| \leq \sup_{\theta \in \Theta} \left\| \frac{\partial \Phi_t^\star(\hat{\sigma}_{it}^2)}{\partial f_t^{2\star}} \right\| \sup_{\theta \in \Theta} \|\hat{f}_t^2 - f_t^2\|,$$

where

$$\frac{\partial \Phi_t(\sigma_{it}^2)}{\partial f_t^2} = \frac{\kappa_i \sigma_{it}^2 (\nu_i + 1)(x_{it}^4/f_t^6)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2} - \frac{\kappa_i \sigma_{it}^2 (\nu_i + 1)(x_{it}^2/f_t^4)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]},$$

is uniformly bounded.

To see this define with $\hat{b}_t$, that is the non-stationary counterpart of

$$b_t = \frac{(x_{it}^2/f_t^2)}{(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)},$$

which takes $\hat{\sigma}_{it}^2$ and $\hat{f}_t^2$ instead of σ_{it}^2 and f_t^2. It can be check that $\hat{b}_t$ is bounded between 0 and 1. So we can write

$$\begin{aligned}
\sup_{\theta \in \Theta} \left\| \frac{\partial \Phi_t^\star(\hat{\sigma}_{it}^2)}{\partial f_t^{2\star}} \right\| &\leq \sup_{\theta \in \Theta} \sup_{f} \left\| \frac{1}{f_t^{2\star}} \right\| \sup_{\theta \in \Theta} \|\kappa_i \hat{\sigma}_{it}^2 (\nu_i + 1)(\hat{b}_{it}^2 - \hat{b}_{it})\| \\
&\leq \sup_{\theta \in \Theta} \left\| \frac{1}{\omega} \right\| \sup_{\theta \in \Theta} \|\kappa_i \hat{\sigma}_{it}^2 (\nu_i + 1)(\hat{b}_{it}^2 - \hat{b}_{it})\| < \infty,
\end{aligned}$$

The last inequality holds because $\omega > 0$, as imposed by assumption 3, and $\hat{\sigma}_{it}^2$ is uniformly bounded.

We obtain

$$\sup_{\theta \in \Theta} \|\hat{\Phi}_t(\hat{\sigma}_{it}^2) - \Phi_t(\hat{\sigma}_{it}^2)\| \xrightarrow{\text{e.a.s.}} 0 \qquad as \qquad t \to \infty.$$

as a straightforward application of Lemma 2.1 in Straumann and Mikosch, 2006, since as $t \to \infty$ the norm $\sup_{\theta\in\Theta} \|\hat{f}_t^2 - f_t^2\|$ will vanish by previous arguments.

Fourth, it only remains to verify that the Lipschitz constant of the perturbed stochastic recurrence equation converges exponentially fast almost surely to the Lipschitz constan of the unperturbed map, that can be verified again by the mean value theorem. We have

$$\sup_{\theta\in\Theta} \|\hat{\Phi}_t(\hat{\sigma}_{it}^2) - \Phi_t(\hat{\sigma}_{it}^2)\| \leq \sup_{\theta\in\Theta} \left\| \frac{\partial^2 \Phi_t^\star(\sigma_{it}^2)}{\partial \sigma_{it}^{2\star} \partial f_t^{2\star}} \right\| \sup_{\theta\in\Theta} \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\|,$$

where

$$\begin{aligned}\frac{\partial^2 \Phi_t(\sigma_{it}^2)}{\partial \sigma_{it}^2 \partial f_t^2} =& \frac{\kappa_i(\nu_i+1)(x_{it}^4/f_t^6)}{[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2} - \frac{(2[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)](\nu_i-2))\kappa_i(\nu_i+1)(x_{it}^4/f_t^6)}{[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2} \\ &- \frac{\kappa_i(\nu_i+1)(x_{it}^2/f_t^4)}{[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]} + \frac{\kappa_i\sigma_{it}^2(\nu_i-2)(\nu_i+1)(x_{it}^2/f_t^4)}{[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]}\end{aligned}$$

Therefore, by the same argument as before, we may rewrite the above equation in terms of the random variables b_t, in fact

$$\begin{aligned}\sup_{\theta\in\Theta} \left\| \frac{\partial^2 \Phi_t^\star(\sigma_{it}^2)}{\partial \sigma_{it}^{2\star} \partial f_t^{2\star}} \right\| \leq& \sup_{\theta\in\Theta} \sup_{f} \left\| \frac{\kappa_i(\nu_i+1)\hat{b}_t}{f_t^{2\star}} \right\| \\ &\times \left[1 + \|\hat{b}_t^2\| + \left(\|(2[(\nu_i-2)\sigma_{it}^{2\star} + (x_{it}^2/f_t^2)](\nu_i-2))\| \right) + \|\sigma_{it}^{2\star}(\nu_i-2)\| \right]\end{aligned}$$

so that, it is possible to verify that

$$\sup_{\theta\in\Theta} \left\| \frac{\partial^2 \Phi_t^\star(\sigma_{it}^2)}{\partial \sigma_{it}^{2\star} \partial f_t^{2\star}} \right\| < \infty,$$

and since $\sup_{\theta\in\Theta} \|\hat{\sigma}_{it}^2 - \sigma_t^2\|$ as $t \to \infty$ we could apply again Lemma 2.1 in Straumann and Mikosch, 2006 and obtain the invertibility of the filter.

Finally, since the multiplicative structure of model (3.5), it remains to show that the filtered $\{(\hat{f}_t\hat{\sigma}_{it})^2\}_{t\in\mathbb{N}}$ is still invertible and hence will converge to the unique stationary and ergodic solution for $i = 1, \ldots, N$ and for any starting points $\hat{f}_1^2$ and $\hat{\boldsymbol{\sigma}}_1^2$, where $\hat{\boldsymbol{\sigma}}_1^2 = (\hat{\sigma}_{11}^2, \ldots, \hat{\sigma}_{N1}^2)^\top$.

We easily achieve this result by using the following elementary decomposition

$$\begin{aligned}\|(\hat{f}_t\hat{\sigma}_{it})^2 - (f_t\sigma_{it})^2\| =& \|(\hat{f}_t^2 - f_t^2)\sigma_{it}^2 + f_t^2(\hat{\sigma}_{it}^2 - \sigma_{it}) + (\hat{f}_t^2 - f_t^2)(\hat{\sigma}_{it}^2 - \sigma_{it})\| \\ \leq& \|\hat{f}_t^2 - f_t^2\| \, \|\hat{\sigma}_{it}^2\| + \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \, \|f_t^2\| + \|\hat{f}_t^2 - f_t^2\| \, \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\|,\end{aligned} \tag{B.6}$$

for $i = 1, \ldots, N$ by virtue of the triangle inequality and the Cauchy-Schwartz's inequality. At this point, we immediately recognize that $\sup_{\theta\in\Theta} \|\hat{f}_t^2 - f_t^2\| \xrightarrow{\text{e.a.s.}} 0$ and $\sup_{\theta\in\Theta} \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \xrightarrow{\text{e.a.s.}} 0$. Furthermore, recall from the previous section that we have $\mathbb{E}[\sup_{\theta\in\Theta} |f_t^2|] < \infty$ and $\mathbb{E}[\sup_{\theta\in\Theta} |\sigma_{it}^2|] < \infty$. Thus, the Jensen's inequality implies again the existence of their log-moments and we are allowed to apply Lemma 2.1 of Straumann and Mikosch, 2006 componentwise to the last inequality in order to obtain

$$\sup_{\theta\in\Theta} \|(\hat{f}_t\hat{\sigma}_{it})^2 - (f_t\sigma_{it})^2\| \xrightarrow{\text{e.a.s.}} 0 \qquad as \qquad t \to \infty,$$

which completes the proof. □

Proof of Lemma 3.4.1

Proof. We begin our proof with the first uniform convergence in (3.11), where $1 \leq N < \infty$ is fixed. From the log-likelihood function (3.10) it is clear that, since we could rewrite the ratios $x_{it}^2/[(\nu_i-2)(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2]$ in terms of the innovations ϵ_{it}^2/ν_i, we only need to verify that $\mathbb{E}[|\log(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2|] < \infty$ and then we have

$$\mathbb{E}[|\log(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2|] \leq \mathbb{E}[|\log f_t(\boldsymbol{\lambda})^2|] + \mathbb{E}[|\log \sigma_{it}(\boldsymbol{\psi}_i)^2|].$$

It is possible to show that $\mathbb{E}[|\log f_t(\boldsymbol{\lambda})^2|] < \infty$, since under Assumption 2, Proposition 2 entails $\mathbb{E}[|\log f_t(\boldsymbol{\lambda})^2|] < \infty$. We note that, $\mathbb{E}[\log f_t(\boldsymbol{\lambda})^{2\varepsilon}] < \infty$ for some $\varepsilon > 0$ due to the compactness of $\boldsymbol{\Theta}$. Additionally, $f_t(\boldsymbol{\lambda})^2$ is bounded away from zero almost surely because $\omega > 0$ and hence $\mathbb{E}[\log f_t(\boldsymbol{\lambda})^2] < \infty$ which implies $\mathbb{E}[|\log f_t(\boldsymbol{\lambda})^2|] < \infty$.

Using similar arguments it is possible to show that $\mathbb{E}[|\log \sigma_{it}(\boldsymbol{\psi}_i)^2|] < \infty$ since under Assumption 3 $\min_{i=1,\ldots,N} \nu_i > 2$ and $\phi < 1$. Note also that the multiplicative recursion $\{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}$ is bounded away from zero for all $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}_i$ with $i = 1, \ldots, N$ almost surely, therefore $\mathbb{E}[\log(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2] < \infty$ which again implies $\mathbb{E}[|\log(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2|] < \infty$.

The first result in (3.11) is then obtained by appelling at the Uniform Strong Law of Large Numbers established by Theorem 2.7 in Straumann and Mikosch, 2006. In fact, this ergodic theorem applies under the moment bound $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_{it}(\boldsymbol{\theta})|] < \infty$, which is clearly satisfied by virtue of the discussion above and standard continuity arguments. We obtain this result by combining Theorem 13.3 in Billingsley, 2012 and Theorem 3.5.8 in Stout, 1974 on the likelihood function over the stationary and ergodic sequence $\{(f_t^2, (\boldsymbol{\sigma}_t^2)^\top, \boldsymbol{\epsilon}_t^\top)\}_{t\in\mathbb{Z}}$.

Now we concentrate on the second uniform convergence in (3.11), where $N \to \infty$. However, it suffices to notice that the same moment bound $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_{it}(\boldsymbol{\theta})|] < \infty$ is enough to ensure that the limit $\mathcal{L}(\boldsymbol{\theta})$ exists and is finite, thus

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_N(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0 \quad as \quad N \to \infty,$$

again by virtue of Theorem 2.7 in Straumann and Mikosch, 2006. □

Proof of Lemma 3.4.2

Proof. Consider the conditional density (3.9) and the likelihood (3.10). Let us define $\ell_{it}(\boldsymbol{\theta}_0) \equiv \ell_{it}(\boldsymbol{\lambda}_0, \boldsymbol{\psi}_{i0})$ and $\ell_{it}(\boldsymbol{\theta}) \equiv \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)$ and note that, if $\nu_{i0} = \nu_i$, then

$$\begin{aligned}
0 =& \ell_{it}(\boldsymbol{\theta}) - \ell_{it}(\boldsymbol{\theta}_0)\\
=& \frac{1}{2}\log\frac{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}{(f_t(\boldsymbol{\lambda}_0)\sigma_{it}(\boldsymbol{\psi}_{i0}))^2}\\
&- \frac{(\nu_{i0}+1)}{2}\log\left[\left(1+\frac{x_{it}^2}{(\nu_{i0}-2)(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}\right)\Bigg/\left(1+\frac{x_{it}^2}{(\nu_{i0}-2)(f_t(\boldsymbol{\lambda}_0)\sigma_{it}(\boldsymbol{\psi}_{i0}))^2}\right)\right],
\end{aligned}$$

where the equation holds if and only if $(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2 = (f_t(\boldsymbol{\lambda}_0)\sigma_{it}(\boldsymbol{\psi}_{i0}))^2$ for all t and $i = 1, \ldots, N$, therefore one only need to prove that this last equality is true if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Moreover, it is clear that the $\{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\}_{t\in\mathbb{Z}}$ and $\{(f_t(\boldsymbol{\lambda}_0)\sigma_{it}(\boldsymbol{\psi}_{i0}))^2\}_{t\in\mathbb{Z}}$ are stationary ergodic sequences, therefore the same holds true for the sequence generated by their differences and hence we are allowed to analyze the difference recursion and to this end we define the following random variables

$$\eta_{it} = \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_{it}^2\epsilon_{it}^2 - 1\right) \qquad u_{it} = \left(\frac{(\nu_{i0}+1)\epsilon_{it}^2}{(\nu_{i0}-2)+\epsilon_{it}^2} - 1\right).$$

Then, after some manipulations and recalling that $\delta_i = (1-\phi_i)$ with $i = 1,\dots,N$, it possible to rewrite the difference recursion as

$$\begin{aligned}(f_{t+1}(\boldsymbol{\lambda})\sigma_{i,t+1}(\boldsymbol{\psi}_i))^2 &- (f_{t+1}(\boldsymbol{\lambda}_0)\sigma_{i,t+1}(\boldsymbol{\psi}_{i0}))^2 \\ &= (\omega-\omega_0) - (\omega\phi_i - \omega_0\phi_{i0}) + \xi_{it} f_t(\boldsymbol{\lambda}_0)^2 + \zeta_{it}\sigma_{it}(\boldsymbol{\psi}_{i0})^2 + \varsigma_{it}(f_t(\boldsymbol{\lambda}_0)\sigma_{it}(\boldsymbol{\psi}_{i0}))^2, \quad \text{(B.7)}\end{aligned}$$

where

$$\varsigma_{it} = (\beta\phi_i - \beta_0\phi_{i0}) + \Big\{(\alpha\kappa_i - \alpha_0\kappa_{i0})\Big\}\eta_{it}u_{it} + \Big\{(\alpha\phi_i - \alpha_0\phi_{i0})\Big\}\eta_{it} + \Big\{(\beta\kappa_i - \beta_0\kappa_{i0})\Big\}u_{it}, \quad \text{(B.8)}$$

$$\xi_{it} = (\beta-\beta_0) - (\beta\phi_i - \beta_0\phi_{i0}) + \Big\{(\alpha-\alpha_0) - (\alpha\phi_i - \alpha_0\phi_{i0})\Big\}\eta_{it}, \quad \text{(B.9)}$$

$$\zeta_{it} = (\omega\phi_i - \omega_0\phi_{i0}) + \Big\{(\omega\kappa_i - \omega_0\kappa_{i0})\Big\}u_{it}. \quad \text{(B.10)}$$

The discussion may start by considering the terms $(\omega-\omega_0)$ and $(\omega\phi_i - \omega_{i0}\phi_{i0})$. It is clear that every random variable involved in equations (B.7) and the related random coefficients are independent from each other and so excluding cases where linear combinations of the variable are null. Hence, in order to obtain $(f_{t+1}(\boldsymbol{\lambda})\sigma_{i,t+1}(\boldsymbol{\psi}_i))^2 - (f_{t+1}(\boldsymbol{\lambda}_0)\sigma_{i,t+1}(\boldsymbol{\psi}_{i0}))^2 = 0$ we should have the equivalence

$$(\omega-\omega_0) - (\omega\phi_i - \omega_0\phi_{i0}) = \xi_{it} f_t(\boldsymbol{\lambda}_0)^2 + \zeta_{it}\sigma_{it}(\boldsymbol{\psi}_{i0})^2 + \varsigma_{it}(f_t(\boldsymbol{\lambda}_0)\sigma_{it}(\boldsymbol{\psi}_{i0}))^2,$$

but clearly, under the conditions $\alpha > 0$ and $\kappa_i > 0$ for every $i = 1,\dots,N$, this is not possible, since the distributions of $f_{t+1}(\boldsymbol{\lambda}_0)$, $\sigma_{i,t+1}(\boldsymbol{\psi}_{i0})$ and then $(f_{t+1}(\boldsymbol{\lambda}_0)\sigma_{i,t+1}(\boldsymbol{\psi}_{i0}))$ are ensured to be nondegenerate almost surely for all t. As a consequence, one must have

$$\omega = \omega_0 \qquad \phi_i = \phi_{i0}.$$

Now consider the equation (B.10). Of course, the result just obtained entails the condition $(\kappa_i - \kappa_{i0})u_{it} = 0$, which is possible if and only if $\kappa_i = \kappa_{i0}$ by continuity of u_{it}.

Finally, equation (B.9) give us

$$(\beta-\beta_0) - (\beta\phi_i - \beta_0\phi_{i0}) = \Big\{(\alpha-\alpha_0) - (\alpha\phi_i - \alpha_0\phi_{i0})\Big\}\eta_{it}.$$

However, we note again that $\beta = \beta_0$ and $\alpha = \alpha_0$ because η_{it} is random. Then, the discussion extends trivially for (B.8). Therefore, the proof is complete. □

Proof of Theorem 3.4.3

Proof. First, we focus on the case for T large. To this end, we may split the likelihood function as follows

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_{NT}(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\|. \quad \text{(B.11)}$$

Let us consider the difference of the first term in the RHS. Applying the mean value expansion around $(\hat{f}^*_t(\boldsymbol{\lambda})\hat{\sigma}^*_{it}(\boldsymbol{\psi}_i))^2$ for $i = 1,\dots,N$ yields

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_{NT}(\boldsymbol{\theta})\| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (f^\star_t(\boldsymbol{\lambda})\sigma^\star_{it}(\boldsymbol{\psi}_i))^2} \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2 - (f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\|. \quad \text{(B.12)}$$

where $(f_t^\star(\boldsymbol{\lambda})\sigma_{it}^\star(\boldsymbol{\psi}_i))^2$ lies between $(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2$ and $(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2$. The partial derivative may be expressed as

$$\frac{\partial \mathcal{L}_{NT}(\boldsymbol{\theta})}{\partial (f_t^\star(\boldsymbol{\lambda})\sigma_{it}^\star(\boldsymbol{\psi}_i))^2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(2f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^{2\star}}[(\nu_i+1)b_{it}^\star(\boldsymbol{\theta})-1], \tag{B.13}$$

where $b_{it}^\star(\boldsymbol{\theta})$ is defined as

$$b_{it}^\star(\boldsymbol{\theta}) = \frac{[x_{it}^2/(f_t^{2\star}(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^{2\star}]}{(\nu_i-2)+[x_{it}^2/(f_t^{2\star}(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^{2\star}]},$$

and it can be easily noted that this is a random variable bounded in the interval $[0,1]$.

Now, considering the fact that both $(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2$ and $(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2$ lie in $[c,+\infty)$ and $[g,+\infty)$, where

$$c = \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left(\frac{\omega}{1-(\beta-\alpha)}\right), \qquad g = \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left(\frac{\delta_i}{1-(\phi_i-\kappa_i)}\right),$$

such that $c>0$ and also $g>0$, we obtain that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial \mathcal{L}_{NT}(\boldsymbol{\theta})}{\partial (f_t^\star(\boldsymbol{\lambda})\sigma_{it}^\star(\boldsymbol{\psi}_i))^2}\right\| \leq \frac{1}{2cg}\left(\frac{\max_{\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_N\in\boldsymbol{\Psi}}(\nu_i+1)}{cg}+1\right) = \bar{M},$$

and therefore

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})-\mathcal{L}_{NT}(\boldsymbol{\theta})\| \leq \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{\infty}\bar{M}\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2-(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\|. \tag{B.14}$$

Clearly, the first term on the RHS is finite and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2-(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\| \xrightarrow{\text{e.a.s.}} 0$ by virtue of Proposition 3. Hence, we satisfy the conditions of Lemma 2.1 in Straumann and Mikosch, 2006 and we obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})-\mathcal{L}_{NT}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0, \forall t\in\mathbb{N}$.

Moreover we can show that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\mathcal{L}_{NT}(\boldsymbol{\theta})-\mathcal{L}_{N}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0$, as an application of Lemma 3.4.1,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})-\mathcal{L}_{N}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0.$$

In Lemma 3.4.2 we have established the identifiability of the unique maximizer $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}$. In addition, we note that $\boldsymbol{\Theta}$ is compact and the continuity of $\mathcal{L}_{NT}(\boldsymbol{\theta})$ in $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, $\forall t\in\mathbb{N}$, implies that the limit $\mathcal{L}_{N}(\boldsymbol{\theta})$ is also a continuous function in $\boldsymbol{\theta}$. It follows that $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}$ is also unique, see White, 1994. The strong consistency thus follows for N fixed and $T\to\infty$.

Now let us turn to the case of $N\to\infty$. Analogously to (3.11), we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\hat{\mathcal{L}}_{N}(\boldsymbol{\theta})-\mathcal{L}(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\hat{\mathcal{L}}_{N}(\boldsymbol{\theta})-\mathcal{L}_{N}(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\mathcal{L}_{N}(\boldsymbol{\theta})-\mathcal{L}(\boldsymbol{\theta})\|. \tag{B.15}$$

Thus, we can follow the same strategy such that, the multivariate mean value theorem can be applied componentwise for the first term in the RHS of (B.15) yileding the same result of (B.12) and (B.13). Notice that, for $i=1,\ldots,N$ they are still bounded independent random variables and the exponentially fast almost sure convergence of the second term holds even if we let $N\to\infty$. Moreover, Lemma 3.4.1 entails the convergence of the second term in the RHS of

(B.15).

Summing up, we may rely again to Lemma 3.4.2 which again guarantees the uniqueness identifiability of the true parameter vector $\boldsymbol{\theta}_0$ and finally, by the compactness of the parameter space $\boldsymbol{\Theta}$ strong consistency follows for N and $T \to \infty$. □

Proof of Proposition 4

Proof. First, note that

$$d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2 = (df_t^2(\boldsymbol{\lambda}))\sigma_{it}^2(\boldsymbol{\psi}_i) + f_t^2(\boldsymbol{\lambda})(d\sigma_{it}^2(\boldsymbol{\psi}_i)).$$

Consider the recursions in (3.6) and notice also that under maintained assumptions the random maps involved in that system are twice continuosly differentiable functions of a stationary ergodic sequence with finite log-moments. Hence, the existence and convergence to the stationary ergodic solution follow directly from Theorems 3.5.3 and 3.5.8 of Stout, 1974.

Second, in the same vein of Proposition 3, we consider (B.4) and by appealing to the Cauchy rule of invariance for differentials, we obtain the unfolded perturbed recursion below

$$\begin{aligned} d\hat{f}_{t+1}^2(\boldsymbol{\lambda}) =& (d\omega)\frac{1}{1-(\beta-\alpha)} \\ &+ (d\alpha)\left\{\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{t-1}(\beta-\alpha)^j - \alpha\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{t-1}j(\beta-\alpha)^{j-1}\right]x_{i,t-j}^2 - (t-1)(\beta-\alpha)^{t-2}\hat{f}_1^2\right\} \\ &+ (d\beta)\left\{\frac{1}{[1-(\beta-\alpha)]^2} + \alpha\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{t-1}j(\beta-\alpha)^{j-1}x_{i,t-j}^2 + (t-1)(\beta-\alpha)^{t-2}\hat{f}_1^2\right\}. \end{aligned} \tag{B.16}$$

Thus, it follows again that the condition $|\beta-\alpha| < 1$ of Assumption 4 is enough to ensure the asymptotic negligibility of the perturbation due to the chosen starting value for the filtering equation of the common factor and therefore, the exponentially fast almost sure convergence to the unique stationary ergodic solution.

Turning to the differential of the idiosyncratic components, we have

$$\begin{aligned} d\hat{\sigma}_{i,t+1}^2(\boldsymbol{\psi}_i) =& \left\{(\phi-\kappa_i) + \kappa_i\frac{(\nu_i+1)(x_{it}^4/\hat{f}_t^4)}{[(\nu_i-2)\hat{\sigma}_{it}^2 + (x_{it}^2/\hat{f}_t^2)]^2}\right\}\left[d\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)\right] \\ &+ \left\{(d\delta_i) + (d\phi_i)\hat{\sigma}_{it}^2 + (d\kappa_i)\hat{\sigma}_{it}^2 u_{it} + \kappa_i\hat{\sigma}_{it}^2\hat{a}_{it}(d\nu_i) + \kappa_i\hat{\sigma}_{it}^2\hat{B}_{it}\left[d\hat{f}_t^2(\boldsymbol{\lambda})\right]\right\}, \end{aligned} \tag{B.17}$$

where, for a lighter notations, we report the unperturbed components

$$u_{it} = \left(\frac{(\nu_i+1)(x_{it}^2/f_t^2)}{(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)} - 1\right), \tag{B.18}$$

$$a_{it} = \left(\frac{(x_{it}^2/f_t^2)}{(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)} - \frac{\sigma_{it}^2(\nu_i+1)(x_{it}^2/f_t^2)}{[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2}\right), \tag{B.19}$$

$$B_{it} = \left(\frac{(\nu_i+1)(x_{it}^4/f_t^6)}{[(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2} - \frac{(\nu_i+1)(x_{it}^2/f_t^4)}{(\nu_i-2)\sigma_{it}^2 + (x_{it}^2/f_t^2)}\right). \tag{B.20}$$

At this point we shall proceed as in Proposition 3 and first show the asymptotic negligibility of the perturbation due to the fact that the random maps of the idiosyncratic processes are functions of the perturbed random map of the differential of the common factor. Here it suffice to note that, for certain constant $c > 0$, there exist a stationary sequence $\{C(t)\}_{t\in\mathbb{N}}$, such that

$$\sup_{\theta\in\Theta}\|d\hat{\sigma}_{it}^2(\psi_i)-d\sigma_{it}^2(\psi_i)\|\leq c\times C(t)\sup_{\theta\in\Theta}\|d\hat{f}_t^2(\lambda)-df_t^2(\lambda)\|.$$

Therefore, we satisfy again the conditions of Lemma 2.1 in Straumann and Mikosch, 2006 since the second term vanishes almost surely exponentially fast. Finally, the compact parameter space imposed by the given assumptions also ensure that the random map (B.17) is a contraction on average, hence we further satisfy the conditions of Theorem 2.10 in Straumann and Mikosch, 2006, which demonstrate $\|d\hat{\sigma}_{it}^2(\psi_i)-d\sigma_{it}^2(\psi_i)\|\xrightarrow{\text{e.a.s.}}0$ as $t\to\infty$ for $i=1,\dots,N$.

With a similar decomposition used in (B.6) we obtain the desired result

$$\sup_{\theta\in\Theta}\|d(\hat{f}_t(\lambda)\hat{\sigma}_{it}(\psi_i))^2-d(f_t(\lambda)\sigma_{it}(\psi_i))^2\|\xrightarrow{\text{e.a.s.}}0\qquad as\qquad t\to\infty\qquad and\qquad i=1,\dots,N.\tag{B.21}$$

□

Proof of Proposition 5

Proof. Again, we note that

$$\sup_{\theta\in\Theta}\|d(f_t(\lambda)\sigma_{it}(\psi_i))^2\|\leq\sup_{\theta\in\Theta}\|df_t^2(\lambda)\|\sup_{\theta\in\Theta}\|\sigma_{it}^2(\psi_i)\|+\sup_{\theta\in\Theta}\|f_t^2(\lambda)\|\sup_{\theta\in\Theta}\|d\sigma_{it}^2(\psi_i)\|.$$

As the recursion for σ_{it}^2 in equation (3.3), equation (B.17), together with its components (B.18), (B.19) and (B.20) shows the same uniformly boundedness, hence the arbitrary large number of bounded moments n'_σ may be obtained by following closely the arguments of Proposition 2. Trivially, the same extends with respect to the the stationary counterpart of the perturbed $d\hat{f}_t^2$, giving us the arbitrary large n'_f number of bounded moments and the conclusion follows as an application of the generalized Hölder's inequality. □

Proof of Lemma 3.4.4

Proof. First, we note that for $i=1,\dots,N$ and $t=1,\dots,T$,

$$\frac{d\ell_{it}(\lambda,\psi_i)}{d(\lambda,\psi_i)}=\frac{\partial\ell_{it}(\lambda,\psi_i)}{\partial(\lambda,\psi_i)}+\frac{\partial\ell_{it}(\lambda,\psi_i)}{\partial(f_t(\lambda)\sigma_{it}(\psi_i))^2}\frac{d(f_t(\lambda)\sigma_{it}(\psi_i))^2}{d(\lambda,\psi_i)}.\tag{B.22}$$

The first differential in equation (B.22) is a continuous function of strictly stationary and ergodic processes, hence Theorem 13.3 of Billingsley, 2012 applies straightforwardly. Moreover, because $\{(f_t(\lambda)\sigma_{it}(\psi_i))^2\}$ and the differential processes $\{d(f_t(\lambda)\sigma_{it}(\psi_i))^2\}$ are adapted to $\mathscr{F}_{t-1}$ we have that

$$\mathbb{E}\left[\frac{d\ell_{it}(\lambda,\psi_i)}{d(\lambda,\psi_i)}\middle|\mathscr{F}_{t-1}\right]=\mathbb{E}\left[\frac{\partial\ell_{it}(\lambda,\psi_i)}{\partial(\lambda,\psi_i)}\middle|\mathscr{F}_{t-1}\right]+\mathbb{E}\left[\frac{\partial\ell_{it}(\lambda,\psi_i)}{\partial(f_t(\lambda)\sigma_{it}(\psi_i))^2}\middle|\mathscr{F}_{t-1}\right]\frac{d(f_t(\lambda)\sigma_{it}(\psi_i))^2}{d(\lambda,\psi_i)}.$$

The only component of the first tem in the RHS of (B.22) is

$$\frac{\partial\ell_{it}(\lambda,\psi_i)}{\partial\nu_i}=\frac{1}{2}\left[\psi\left(\frac{\nu_i+1}{2}\right)-\psi\left(\frac{\nu_i}{2}\right)-\frac{1}{\nu_i-2}+\frac{\nu_i+1}{\nu_i-2}b_{it}(\theta)-\log(1-b_{it}(\theta))\right],\tag{B.23}$$

while the first component of the second term is

$$\frac{\partial\ell_{it}(\lambda,\psi_i)}{\partial(f_t(\lambda)\sigma_{it}(\psi_i))^2}=\frac{1}{2}\left[(\nu_i+1)b_{it}(\theta)-\frac{1}{(f_t(\lambda)\sigma_{it}(\psi_i))^2}\right].\tag{B.24}$$

Hence, we just need to show that by using the properties of the beta distributed random variables

$$\mathbb{E}[b_{it}(\boldsymbol{\theta})|\mathscr{F}_{t-1}] = 1/(\nu_i+1) \quad \textit{and}$$
$$\mathbb{E}[\log(1-b_{it}(\boldsymbol{\theta}))|\mathscr{F}_{t-1}] = \psi(\nu_i/2) - \psi((\nu_i+1)/2),$$

it is straightforward to see that for $i=1,\dots,N$

$$\mathbb{E}[d\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)|\mathscr{F}_{t-1}] = 0.$$

Note also that this property entails the uniformly boundedness of (B.23) and (B.24), which means that by using the property that $\|XY\|_p \leq \|X\|_{2p}\|Y\|_{2p}$ for any random variables X and Y, we apply Proposition 5 and get the claimed bound of second moment

$$\mathbb{E}[|d\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)|^2] < \infty.$$

□

Proof of Lemma 3.4.5

Proof. The normality of the score function follows smoothly. The mean value theorem can be invoked to highlights once more that the perturbations due to the starting values of every processes are negligible as we go further in the direction of T. Indeed,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathcal{L}}'_{NT}(\boldsymbol{\theta}) - \mathcal{L}'_{NT}(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \hat{\mathcal{L}}'_{NT}(\boldsymbol{\theta})}{\partial(\hat{f}^*_t(\boldsymbol{\lambda})\hat{\sigma}^*_{it}(\boldsymbol{\psi}_i))^2} \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|d(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2 - d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\|,$$

where $(\hat{f}^*_t(\boldsymbol{\lambda})\hat{\sigma}^*_{it}(\boldsymbol{\psi}_i))^2$ lies between $(\hat{f}_t(\boldsymbol{\lambda})\hat{\sigma}_{it}(\boldsymbol{\psi}_i))^2$ and $(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2$. The second partial derivative is

$$\begin{aligned}\frac{\partial \mathcal{L}'_{NT}(\boldsymbol{\theta})}{\partial(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(2f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2} \\ \times\left[\frac{1}{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2} - (\nu_i+1)(\nu_i-2)b_{it}(\boldsymbol{\theta}) - \frac{1}{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}(\nu_i+1)b_{it}(\boldsymbol{\theta})\right],\end{aligned} \tag{B.25}$$

which is uniformly bounded, while the second term of the RHS of the inequality above converges almost surely exponentially fast to zero from Proposition 4. Hence, we satisfy the conditions of Lemma 2.1 in Straumann and Mikosch, 2006 and we obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\mathcal{L}}'_{NT}(\boldsymbol{\theta}) - \mathcal{L}'_{NT}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0$ as both the dimensions N and T increase towards infinity.

Therefore, the score function obeys the Central Limit Theorem for Martingales of Billingsley, 1961, since the existence of the variance covariance matrix $\boldsymbol{V}$ is entailed by Proposition 5. The claimed convergence in distribution is achieved by appealing to Theorem 18.10 (iv) of Vaart, 1998. □

Proof of Proposition 6

Proof. We have

$$d^2(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2 = (d^2 f_t^2(\boldsymbol{\lambda}))\sigma^2_{it}(\boldsymbol{\psi}_i) + 2d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2 + f_t^2(\boldsymbol{\lambda})(d^2\sigma^2_{it}(\boldsymbol{\psi}_i)). \tag{B.26}$$

It is immediate to note that differentiating again (B.16) we obtain a new in which under the same condition, that is $|\beta-\alpha|<1$ will ensure again the usual required convergence to the unique stationary ergodic solution.

As regards the second differential of the idiosyncratic components, we can show that thee same discussion in the proof of Proposition 4 applies sequentially to the higher-order derivatives. In addition, similar arguments of those in Proposition 5 can be extended here, in order to show the number of bounded moments of the second differential processes. A related discussion can be found in Harvey, 2013 pag. 40. All the derivatives for the first-order *Beta-t-GARCH* can be found in the comprehensive analysis of Ito, 2016.

To conclude, we relay again to Proposition 3.36 of White, 2001, in order to extend the arguments to the whole multiplicative process. □

Proof of Lemma 3.4.6

Proof. The first part of the proof follows the same arguments given in the proof of the consistency theorem 3.4.3 and Lemma 3.4.4. Roughly speaking, we can first show the asymptotic irrelevance of the chosen starting values which perturb the second derivatives of the empirical likelihood function and then proceed with the convergence to they respective limits, as $T\to\infty$ for the first one and both $NT\to\infty$ for the second. Of course, this is a consequence of the fact that $\mathrm{d}^2_{it}\ell(\boldsymbol{\theta})$ is a function of $\{x_{it},x_{i,t-1}\}$ which is stationary and ergodic. Hence, it remains to show that the latter limit is finite. In particular, we need to show that

$$\mathbb{E}[d^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)]<\infty,$$

for $i=1,\dots,N$ and every $t=1,\dots,T$.

We note that the

$$\begin{aligned}\frac{d^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{d(\boldsymbol{\lambda},\boldsymbol{\psi}_i)d(\boldsymbol{\lambda},\boldsymbol{\psi}_i)^\top}=&\frac{\partial^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{\partial(\boldsymbol{\lambda},\boldsymbol{\psi}_i)\partial(\boldsymbol{\lambda},\boldsymbol{\psi}_i)^\top}+\frac{\partial\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{\partial(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}\frac{d^2(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}{d(\boldsymbol{\lambda},\boldsymbol{\psi}_i)d(\boldsymbol{\lambda},\boldsymbol{\psi}_i)^\top}\\&+\frac{\partial^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{\partial(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^4}\frac{d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}{d(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}\frac{d(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}{d(\boldsymbol{\lambda},\boldsymbol{\psi}_i)^\top},\end{aligned}\tag{B.27}$$

is a continuous function of strictly stationary and ergodic processes, hence we can apply Theorem 13.3 of Billingsley, 2012. We need,

$$\begin{aligned}\frac{\partial^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{\partial\nu_i^2}=&\frac{1}{4}\left[\psi'\left(\frac{\nu_i+1}{2}\right)-\psi'\left(\frac{\nu_i}{2}\right)\right]\\&+\frac{1}{2}\left[\frac{1}{(\nu_i-2)^2}-\frac{\nu_i+1}{(\nu_i-2)^2}b_{it}(\boldsymbol{\theta})(1-b_{it}(\boldsymbol{\theta}))-\frac{3}{(\nu_i-2)}b_{it}(\boldsymbol{\theta})\right],\end{aligned}\tag{B.28}$$

$$\frac{\partial^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{\partial(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2\partial\nu_i}=\frac{1}{2}\left[b_{it}(\boldsymbol{\theta})-(\nu_i+1)b_{it}(\boldsymbol{\theta})(1-b_{it}(\boldsymbol{\theta}))\right],\tag{B.29}$$

and finally

$$\begin{aligned}\frac{\partial^2\ell_{it}(\boldsymbol{\lambda},\boldsymbol{\psi}_i)}{\partial(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^4}=&\frac{1}{(2f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}\\&\times\left[\frac{1}{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}-(\nu_i+1)(\nu_i-2)b_{it}(\boldsymbol{\theta})-\frac{1}{(f_t(\boldsymbol{\lambda})\sigma_{it}(\boldsymbol{\psi}_i))^2}(\nu_i+1)b_{it}(\boldsymbol{\theta})\right].\end{aligned}\tag{B.30}$$

In line with previous arguments we can easily check the uniformly boundedness of (B.28), (B.29) and (B.30). By using the property that $\|XY\|_p \leq \|X\|_{2p}\|Y\|_{2p}$ for any random variables X and Y, we apply Proposition 5 and Proposition 6. The claimed bound is then obtained. □

Proof of Theorem 3.4.7

Proof. We discuss the proof for $N, T \to \infty$. Standard arguments for the asymptotic normality proof and the Taylor's theorem, lead to the expansion of the conditional likelihood's score around a neighborhood of $\boldsymbol{\theta}_0$, which yields

$$\begin{aligned}\mathbf{0} = \sqrt{NT}\mathcal{L}'_T(\hat{\boldsymbol{\theta}}_T) =& \sqrt{NT}\left[\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0) - \mathcal{L}'_{NT}(\boldsymbol{\theta}_0)\right] + \sqrt{NT}\mathcal{L}'_{NT}(\boldsymbol{\theta}_0)\\ &+ \left[\Big(\mathcal{L}''_{NT}(\boldsymbol{\theta}_0) - \mathcal{L}''(\boldsymbol{\theta}_0)\Big) + \Big(\widehat{\mathcal{L}}''_{NT}(\boldsymbol{\theta}^\star) - \mathcal{L}''_{NT}(\boldsymbol{\theta}_0)\Big) + \mathcal{L}''(\boldsymbol{\theta}_0)\right]\\ &\times \left[\sqrt{NT}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)\right],\end{aligned} \tag{B.31}$$

where $\boldsymbol{\theta}^\star$ is on the cord between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$, componentwise.

From the first line of (B.31), the convergence of the first difference in square brackets is ensured by Proposition 4 and following arguments as in the proof of Lemma 3.4.5, which further entails the fact that $\sqrt{NT}\mathcal{L}'_{NT}(\boldsymbol{\theta}_0)$ obeys the CLT for martingales. Now consider the second line, Lemma 3.4.6 demostrates that the inizial conditions for likelihood's second derivatives are asymptotically irrelevant and the consistency theorem further ensures that the convergence in the same point by continuity arguments. In addition, the Uniform Laqw of Large Numbers guarantee that $\|\mathcal{L}''_{NT}(\boldsymbol{\theta}_0) - \mathcal{L}''(\boldsymbol{\theta}_0)\| \xrightarrow{\text{a.s.}} 0 \quad as \quad NT \to \infty$, see again Lemma 3.4.6 where it is also discussed the existence and invertibility of $\mathbb{E}[d^2\ell_{it}(\boldsymbol{\theta})]$. Thus we can solve the above equation since $\mathcal{L}''(\boldsymbol{\theta}_0)$ is non-singular. Finally we apply the Slusky's Lemma (see Lemma 2.8 (iii) of Vaart, 1998) and complete the proof. □

Appendix C

Proofs of Chapter 4

C.1 Proofs of the Stochastic Properties of the Model

Proof of Theorem 4.3.1

Gaussian Case

Proof. As regards the recursions with the Gaussian specifications, we are able to write the recursions as follows

$$\begin{bmatrix} f_{1,t+1} \\ f_{2,t+1} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \alpha \begin{bmatrix} \frac{1}{(1-\rho^2)}\eta_{1t}^2 - \frac{\rho}{(1-\rho^2)}\eta_{1t}\eta_{2t} - 1 \\ \frac{1}{(1-\rho^2)}\eta_{2t}^2 - \frac{\rho}{(1-\rho^2)}\eta_{1t}\eta_{2t} - 1 \end{bmatrix} + \beta \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}, \qquad t \in \mathbb{Z}, \tag{C.1}$$

where the innovations η_{1t} and η_{2t} are Gaussian with $\mathbb{E}[\eta_{it}] = 0$, $\mathbb{E}[\eta_{it}^2] = 1$ for $i = 1, 2$ and $\mathbb{E}[\eta_{1t}\eta_{2t}] = \rho$.

Thus, the first step is to show the initial log moment condition, which may be easily achieved by using Lemma 2.5.3 in Straumann, 2005. So, for some point $\bar{f}_i \in \mathbb{R}$ and $i = 1, 2$, we have

$$\begin{aligned} \mathbb{E}\left\{ \log^+ \left| \omega_i + \alpha \left[\frac{1}{(1-\rho^2)}\eta_{it}^2 - \frac{\rho}{(1-\rho^2)}\eta_{1t}\eta_{2t} - 1 \right] + \beta \bar{f}_i \right| \right\} \\ \leq 8 \log 2 + \log^+ |\omega_i| + \log^+ |\alpha| + \log^+ |K_\rho| + \log^+ |K_\rho^\star| \\ + \mathbb{E}[\log^+ |\eta_{it}^2|] + \mathbb{E}[\log^+ |\eta_{1t}|] + \mathbb{E}[\log^+ |\eta_{2t}|] + \log^+ |\beta \bar{f}_i| < \infty, \end{aligned}$$

since $\mathbb{E}[\log^+ |\eta_{it}|] < \infty$ and $\mathbb{E}[\log^+ |\eta_{it}^2|] < \infty$ are trivially satisfied from the Gaussianity assumption.

Next, we note that by the c_r-inequality

$$\begin{aligned} \mathbb{E}\left\{ \left| \omega_i + \alpha \left[\frac{1}{(1-\rho^2)}\eta_{it}^2 - \frac{\rho}{(1-\rho^2)}\eta_{1t}\eta_{2t} - 1 \right] + \beta \bar{f}_{it} \right|^n \right\} \\ \leq c\,|\omega_i|^n + c\,|\alpha|^n \left[\mathbb{E}[|K_\rho \eta_{it}|^{2n}] + \mathbb{E}[|K_\rho^\star \eta_{1t}\eta_{2t}|^n] + 1 \right] + c\,|\beta|^n\,|\bar{f}_i|^n < \infty, \end{aligned}$$

since there exists some $n > 0$ such that $\mathbb{E}[|\eta_{it}|^{2n}] < \infty$ for $i = 1, 2$, and $\mathbb{E}[|\eta_{1t}\eta_{2t}|^n] < \infty$. Note also that we could rewrite these moment conditions in vector form, that is, there exists $n > 0$ such that $\mathbb{E}[\|\boldsymbol{\eta}_t \otimes \boldsymbol{\eta}_t\|^n] < \infty$.

The last condition that needs to be verified is the so-called contraction condition, which ensures the strict stationarity and ergodicity property of the score-driven model as a data generating process. As it stands, the recursions in the bivariate system (C.1) are indeed linear for a given $\boldsymbol{\eta}_t$ and so the contraction condition boils down to

$$\mathbb{E}\left[\sup_{f_{it}} |\beta|^n \right] = |\beta|^n \leq |\beta| < 1,$$

which is implied by assumption. □

Student's *t* Case

Proof. The analogous recursions with the Student's *t* specifications are

$$\begin{bmatrix} f_{1,t+1} \\ f_{2,t+1} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \alpha \begin{bmatrix} \frac{(\nu+2)}{(1-\rho^2)(\nu-2)} \left(\frac{\eta_{1t}^2 - \rho\eta_{1t}\eta_{2t}}{1+\frac{1}{(\nu-2)(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)} \right) - 1 \\ \frac{(\nu+2)}{(1-\rho^2)(\nu-2)} \left(\frac{\eta_{2t}^2 - \rho\eta_{1t}\eta_{2t}}{1+\frac{1}{(\nu-2)(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)} \right) - 1 \end{bmatrix} + \beta \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}, \qquad t \in \mathbb{Z}, \qquad \text{(C.2)}$$

where the innovations η_{1t} and η_{2t} Student's *t* distributed with $\nu > 2$ degrees of freedom, $\mathbb{E}[\eta_{it}] = 0$, $\mathbb{E}[\eta_{it}^2] = 1$ for $i = 1,2$ and $\mathbb{E}[\eta_{1t}\eta_{2t}] = \rho$.

Now we have

$$\begin{aligned} &\mathbb{E}\left\{ \log^+ \left| \omega_i + \alpha \left\{ \frac{(\nu+2)}{(1-\rho^2)(\nu-2)} \left[\frac{\eta_{it}^2 - \rho\eta_{1t}\eta_{2t}}{1+\frac{1}{(\nu-2)(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)} \right] - 1 \right\} + \beta \bar{f}_i \right| \right\} \\ &\qquad \le \log^+|\omega_i| + \log^+|\alpha| + \log^+|K_\rho| + \log^+|K_\nu| \\ &\qquad\quad + \mathbb{E}\left[\log^+ \left(\frac{|\eta_{it}\eta_{jt}|}{1+\frac{1}{(\nu-2)(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)} \right) \right] + \log^+|\beta\bar{f}_i|, \end{aligned}$$

where we know that $\mathbb{E}[\log^+|\eta_{it}\eta_{jt}|] < \infty$ for $i,j \in \{1,2\}$ is ensured by the condition on the degrees of freedom, $\nu > 2$, and as regards the denominator, it is a well-known fact that $\log(x) \le x - 1$ for $x > 1$. Hence, we could write

$$\begin{aligned} &\mathbb{E}\left[\log\left(1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right) \right) \right] \\ &\quad \le \frac{1}{(\nu-2)(1-\rho^2)} \mathbb{E}\left[\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2 \right] \\ &\quad = \frac{2}{\nu-2} < \infty. \end{aligned}$$

The c_r-inequality implies that

$$\begin{aligned} &\mathbb{E}\left\{ \left| \omega_i + \alpha \left\{ \frac{(\nu+2)}{(1-\rho^2)(\nu-2)} \left[\frac{\eta_{it}^2 - \rho\eta_{1t}\eta_{2t}}{1+\frac{1}{(\nu-2)(1-\rho^2)}\left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)} \right] - 1 \right\} + \beta \bar{f}_i \right|^n \right\} \\ &\le c\,|\omega_i|^n + c\,|\alpha|^n \left(\mathbb{E}\left[\left| \frac{\eta_{it}\eta_{jt}}{(\nu-2)(1-\rho^2) + \left(\eta_{1t}^2 - 2\rho\eta_{1t}\eta_{2t} + \eta_{2t}^2\right)} \right|^n \right] + 1 \right) + c\,|\beta|^n\,|\bar{f}_i|^n < \infty, \end{aligned}$$

for any n and $i,j \in \{1,2\}$.

Lastly, also for the Student's *t* specifications, the contraction condition boils down to

$$\mathbb{E}\left[\sup_{f_{it}} |\beta|^n \right] = |\beta|^n \le |\beta| < 1,$$

implied by assumption. □

Proof of Theorem 4.3.2

Gaussian Case

Proof. In the Gaussian setting the aforementioned filtering equations are

$$\begin{bmatrix} \hat{f}_{1,t+1}(\boldsymbol{\theta}) \\ \hat{f}_{2,t+1}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \alpha \begin{bmatrix} \frac{1}{(1-\rho^2)}\left(\epsilon_{1t}^2 \exp\{-2\hat{f}_{1t}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}\right) - 1 \\ \frac{1}{(1-\rho^2)}\left(\epsilon_{2t}^2 \exp\{-2\hat{f}_{2t}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}\right) - 1 \end{bmatrix} + \beta \begin{bmatrix} \hat{f}_{1t}(\boldsymbol{\theta}) \\ \hat{f}_{2t}(\boldsymbol{\theta}) \end{bmatrix}, \tag{C.3}$$

with $t \in \mathbb{N}$.

As done in the proof of Theorem 4.3.1, we can evaluate the initial log moment condition componentwise. Relying again on Lemma 2.5.3 in Straumann, 2005, for some fixed point $(f_{11}, f_{21})^\top \in \mathbb{R}^2$, we have

$$\begin{aligned} &\mathbb{E}\left\{ \log^+ \left| \omega_i + \alpha \left[\frac{1}{(1-\rho^2)} \left(\epsilon_{it}^2 \exp\{-2f_{i1}\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(f_{11}+f_{21})\} \right) - 1 \right] + \beta f_{i1} \right| \right\} \\ &\qquad \le 8\log 2 + \log^+|\omega_i| + \log^+|\alpha| + \log^+|K_\rho| + \log^+|K_\rho^\star| \\ &\qquad\quad + \mathbb{E}[\log^+|\epsilon_{it}^2|] + \mathbb{E}[\log^+|\epsilon_{1t}|] + \mathbb{E}[\log^+|\epsilon_{2t}|] \\ &\qquad\quad - 2f_{i1} - f_{11} - f_{21} + \log^+|\beta f_{i1}| < \infty, \end{aligned}$$

for $i = 1, 2$, since $\mathbb{E}[\log^+|\epsilon_{it}^2|] < \infty$ are implied by the assumption and also $\mathbb{E}[|\epsilon_{it}^2|^n] < \infty$ for some $n > 0$.

Let us determine the required terms for the matrix of partial derivatives in (4.13), where for $i, j = 1, 2$ we have,

$$\frac{\partial \hat{s}_{it}(\boldsymbol{\theta})}{\partial \hat{f}_{it}(\boldsymbol{\theta})} = -\frac{1}{(1-\rho^2)}\left[2\epsilon_{it}^2 \exp\{-2f_{it}\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\} \right], \tag{C.4}$$

while for $i \neq j$

$$\frac{\partial \hat{s}_{it}(\boldsymbol{\theta})}{\partial \hat{f}_{jt}(\boldsymbol{\theta})} = \frac{1}{(1-\rho^2)}\left[\rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\} \right]. \tag{C.5}$$

In conclusion, the desired exponentially fast almost sure convergence is obtained by using equations (C.4) and (C.5) to verify condition (4.15). □

Student's *t* Case

Proof. The filtering recursions with the Student's *t* specifications are

$$\begin{bmatrix} \hat{f}_{1,t+1}(\boldsymbol{\theta}) \\ \hat{f}_{2,t+1}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \alpha \begin{bmatrix} \frac{(\nu+2)}{(\nu-2)(1-\rho^2)} \left(\frac{\epsilon_{1t}^2 \exp\{-2\hat{f}_{1t}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}}{1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}} \right)} \right) - 1 \\ \frac{(\nu+2)}{(\nu-2)(1-\rho^2)} \left(\frac{\epsilon_{2t}^2 \exp\{-2\hat{f}_{2t}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}}{1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{2t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}} \right)} \right) - 1 \end{bmatrix} + \beta \begin{bmatrix} \hat{f}_{1t}(\boldsymbol{\theta}) \\ \hat{f}_{2t}(\boldsymbol{\theta}) \end{bmatrix}, \tag{C.6}$$

with $t \in \mathbb{N}$.

Here, evaluating the initial log moment condition for some fixed point $(f_{11}, f_{21})^\top \in \mathbb{R}^2$ yields

$$\begin{aligned}
&\mathbb{E}\left\{ \log^+ \left| \omega_i + \alpha \left\{ \frac{(\nu+2)}{(\nu-2)(1-\rho^2)} \right.\right.\right. \\
&\quad \times \left[\frac{\epsilon_{it}^2 \exp\{-2\hat{f}_{it}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}}{1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)} \right] - 1 \Bigg\} + \beta f_{i1} \Bigg| \Bigg\} \\
&\quad \leq \log^+ |\omega_i| + \log^+ |\alpha| \\
&\qquad + \mathbb{E}\left[\log^+ \left(\frac{|\epsilon_{it}\epsilon_{jt}|}{(\nu-2)(1-\rho^2) + \left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)} \right) \right] \\
&\qquad\quad + \log^+ |\beta f_{i1}| < \infty,
\end{aligned}$$

for $i, j \in \{1, 2\}$. The result here follows because $\mathbb{E}[\log^+ |\epsilon_{it}\epsilon_{jt}|] < \infty$ is implied by the assumption and moreover $\mathbb{E}[|\epsilon_{it}\epsilon_{jt}|^n] < \infty$ for some $n > 0$ is ensured by the already proved existence of second moments in the data generating process, in fact $2 < \nu < \infty$.

The terms needed for the construction of (4.13) are, for $i = 1, 2$

$$\begin{aligned}
\frac{\partial \hat{s}_{it}(\boldsymbol{\theta})}{\partial \hat{f}_{it}(\boldsymbol{\theta})} =& \frac{(\nu+2)}{(\nu-2)(1-\rho^2)} \left[\frac{\rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\} - 2\epsilon_{it}^2 \exp\{-2\hat{f}_{it}(\boldsymbol{\theta})\}}{1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)} \right] \\
&- \frac{(\nu+2)}{(\nu-2)^2(1-\rho^2)^2} \left[\frac{2\rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\} - 2\epsilon_{it}^2 \exp\{-2\hat{f}_{it}(\boldsymbol{\theta})\}}{\left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)\right]^2} \right] \\
&- \frac{(\nu+2)}{(\nu-2)^2(1-\rho^2)^2} \left[\frac{\epsilon_{it}^2 \exp\{-2\hat{f}_{it}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}}{\left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{it}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)\right]^2} \right]
\end{aligned} \tag{C.7}$$

while for $i \neq j$

$$\begin{aligned}
\frac{\partial \hat{s}_{it}(\boldsymbol{\theta})}{\partial \hat{f}_{jt}(\boldsymbol{\theta})} =& \frac{(\nu+2)}{(\nu-2)(1-\rho^2)} \left[\frac{\rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}}{1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)} \right] \\
&- \frac{(\nu+2)}{(\nu-2)^2(1-\rho^2)^2} \left[\frac{2\rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\} - 2\epsilon_{jt}^2 \exp\{-2\hat{f}_{jt}(\boldsymbol{\theta})\}}{\left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{1t}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)\right]^2} \right] \\
&- \frac{(\nu+2)}{(\nu-2)^2(1-\rho^2)^2} \left[\frac{\epsilon_{it}^2 \exp\{-2\hat{f}_{it}(\boldsymbol{\theta})\} - \rho\epsilon_{1t}\epsilon_{2t}\exp\{-(\hat{f}_{1t}(\boldsymbol{\theta}) + \hat{f}_{2t}(\boldsymbol{\theta}))\}}{\left[1 + \frac{1}{(\nu-2)(1-\rho^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2\hat{f}_{1t}(\boldsymbol{\theta})\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{\hat{f}_{it}(\boldsymbol{\theta})+\hat{f}_{2t}(\boldsymbol{\theta})\}} + \frac{\epsilon_{2t}^2}{\exp\{2\hat{f}_{2t}(\boldsymbol{\theta})\}}\right)\right]^2} \right]
\end{aligned} \tag{C.8}$$

In the same vein, the desired exponentially fast almost sure convergence is obtained by using equations (C.7) and (C.8) to verify condition (4.15). □

C.2 Analytical Formulae for a Time-Varying Correlation

Gaussian Case

$$u_t = \left[\frac{\rho_t}{1-\rho_t^2} - \left(\frac{\rho_t}{(1-\rho^2)^2}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right) - \frac{1}{1-\rho_t^2}\frac{\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}}\right]\frac{\partial\rho_t}{\gamma_t}.$$

Differentiating the score with respect to the dynamic conditional correlation coefficient yields

$$\begin{aligned}\frac{\partial u_t}{\partial\rho_t} = &\left[\frac{1+\rho_t^2}{(1-\rho_t^2)^2} - \frac{1+3\rho_t^2}{(1-\rho_t^2)^3}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right]\left(\frac{\partial\rho_t}{\partial\gamma_t}\right)^2 \\ &+ \left[\frac{\rho_t}{1-\rho_t^2} - \frac{\rho_t}{(1-\rho_t^2)^2}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right. \\ &\left. - \frac{1}{1-\rho_t^2}\frac{\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}}\right]\frac{\partial^2\rho_t}{\partial\gamma_t^2}.\end{aligned}$$

Student's t Case

$$u_t = \left[\frac{\rho_t}{1-\rho_t^2} - (\nu+2)\left(\frac{\frac{\rho_t}{(\nu-2)(1-\rho_t^2)^2}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right) - \frac{1}{(\nu-2)(1-\rho_t^2)}\frac{\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}}}{1+\frac{1}{(\nu-2)(1-\rho_t^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)}\right)\right]\frac{\partial\rho_t}{\partial\gamma_t}.$$

Analogously to the Gaussian case, we differentiate with respect to the dynamic conditional correlation coefficient, so that

$$\begin{aligned}\frac{\partial u_t}{\partial\rho_t} = &\left\{\frac{1+\rho_t^2}{(1-\rho_t^2)^2} - (\nu+2)\right. \\ &\times\left(\left\{\left[\left(\frac{1}{(\nu-2)(1-\rho_t^2)} + \frac{4\rho_t^2}{(\nu-2)(1-\rho_t^2)^3}\right)\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right]\right.\right. \\ &\left. - \left(\frac{4\rho_t^2}{(\nu-2)(1-\rho_t^2)^2}\frac{\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}}\right)\right\} \\ &\times\left[1+\frac{1}{(\nu-2)(1-\rho_t^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right]^{-1} \\ &- \left\{\frac{\rho_t}{(\nu-2)(1-\rho_t^2)^2}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right. \\ &\left. - \left(\frac{1}{(\nu-2)(1-\rho_t^2)}\frac{\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}}\right)\right\} \\ &\left.\left.\times\left[1+\frac{1}{(\nu-2)(1-\rho_t^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)\right]^{-2}\right)\right\}\left(\frac{\partial\rho_t}{\partial\gamma_t}\right)^2 \\ &+ \left\{\frac{\rho_t}{1-\rho_t^2} - (\nu+2)\left(\frac{\frac{\rho_t}{(\nu-2)(1-\rho_t^2)^2}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right) - \frac{1}{(\nu-2)(1-\rho_t^2)}\frac{\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}}}{1+\frac{1}{(\nu-2)(1-\rho_t^2)}\left(\frac{\epsilon_{1t}^2}{\exp\{2f_{1t}\}} - \frac{2\rho_t\epsilon_{1t}\epsilon_{2t}}{\exp\{f_{1t}+f_{2t}\}} + \frac{\epsilon_{2t}^2}{\exp\{2f_{2t}\}}\right)}\right)\right\}\frac{\partial^2\rho_t}{\partial\gamma_t^2}.\end{aligned}$$

Bibliography

Alexander, Carol (2001). "Orthogonal GARCH". In: *Mastering Risk* Financial Times–Prentice Hall: London.2, 21–38.

Amado, Cristina and Timo Teräsvirta (2013). "Modelling volatility by variance decomposition". In: *Journal of Econometrics* 175.2, pp. 142 –153. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2013.03.006. URL: http://www.sciencedirect.com/science/article/pii/S030440761300064X.

Baillie, Richard T. and Robert J. Myers (1991). "Bivariate garch estimation of the optimal commodity futures Hedge". In: *Journal of Applied Econometrics* 6.2, pp. 109–124. DOI: 10.1002/jae.3950060202. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.3950060202. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.3950060202.

Barigozzi, Matteo et al. (2014). "Disentangling systematic and idiosyncratic dynamics in panels of volatility measures". In: *Journal of Econometrics* 182.2, pp. 364 –384. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2014.05.017. URL: http://www.sciencedirect.com/science/article/pii/S0304407614001390.

Basrak, Bojan, Richard A. Davis, and Thomas Mikosch (2002). "Regular variation of GARCH processes". In: *Stochastic Processes and their Applications* 99.1, pp. 95 –115. ISSN: 0304-4149. DOI: https://doi.org/10.1016/S0304-4149(01)00156-9. URL: http://www.sciencedirect.com/science/article/pii/S0304414901001569.

Bauwens, Luc, Sébastien Laurent, and Jeroen V. K. Rombouts (2006). "Multivariate GARCH models: a survey". In: *Journal of Applied Econometrics* 21.1, pp. 79–109. DOI: 10.1002/jae.842. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.842. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.842.

Billingsley, P. (2012). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9781118341919. URL: https://books.google.it/books?id=a3gavZbxyJcC.

Billingsley, Patrick (1961). "The Lindeberg-Levy Theorem for Martingales". In: *Proceedings of the American Mathematical Society* 12.5, pp. 788–792. ISSN: 00029939, 10886826. URL: http://www.jstor.org/stable/2034876.

Blasques, F., S. J. Koopman, and A. Lucas (Mar. 2015a). "Information-theoretic optimality of observation-driven time series models for continuous responses". In: *Biometrika* 102.2, pp. 325–343. ISSN: 0006-3444. DOI: 10.1093/biomet/asu076. eprint: http://oup.prod.sis.lan/biomet/article-pdf/102/2/325/682829/asu076supp.pdf. URL: https://doi.org/10.1093/biomet/asu076.

— (Mar. 2015b). "Information-theoretic optimality of observation-driven time series models for continuous responses". In: *Biometrika* 102.2, pp. 325–343. ISSN: 0006-3444. DOI: 10.1093/biomet/asu076. eprint: http://oup.prod.sis.lan/biomet/article-pdf/102/2/325/682829/asu076supp.pdf. URL: https://doi.org/10.1093/biomet/asu076.

Blasques, Francisco et al. (2018). "Feasible invertibility conditions and maximum likelihood estimation for observation-driven models". In: *Electron. J. Statist.* 12.1, pp. 1019–1052. DOI: 10.1214/18-EJS1416. URL: https://doi.org/10.1214/18-EJS1416.

Bollerslev, Tim (1986). “Generalized autoregressive conditional heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307 –327. ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(86)90063-1. URL: http://www.sciencedirect.com/science/article/pii/0304407686900631.

— (1990). “Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model”. In: *The Review of Economics and Statistics* 72.3, pp. 498–505. ISSN: 00346535, 15309142. URL: http://www.jstor.org/stable/2109358.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge (1988). “A Capital Asset Pricing Model with Time-Varying Covariances”. In: *Journal of Political Economy* 96.1, pp. 116–131. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1830713.

Bollerslev, Tim and Jeffrey M. Wooldridge (1992). “Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances”. In: *Econometric Reviews* 11.2, pp. 143–172. DOI: 10.1080/07474939208800229. eprint: https://doi.org/10.1080/07474939208800229. URL: https://doi.org/10.1080/07474939208800229.

Bougerol, P. (1993). “Kalman Filtering with Random Coefficients and Contractions”. In: *SIAM Journal on Control and Optimization* 31.4, pp. 942–959. DOI: 10.1137/0331041. eprint: https://doi.org/10.1137/0331041. URL: https://doi.org/10.1137/0331041.

Breusch, T. S., J. C. Robertson, and A. H. Welsh (1997). “The emperor’s new clothes: a critique of the multivariate t regression model”. In: *Statistica Neerlandica* 51.3, pp. 269–286. DOI: 10.1111/1467-9574.00055. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9574.00055. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9574.00055.

Broda, Christian, Ephraim Leibtag, and David E. Weinstein (2009). “The Role of Prices in Measuring the Poor’s Living Standards”. In: *Journal of Economic Perspectives* 23, pp. 77–97. ISSN: 0895-3309. DOI: 10.1257/jep.23.2.77.

Caivano, Michele, Andrew Harvey, and Alessandra Luati (2016). “Robust time series models with trend and seasonal components”. In: *SERIEs* 7.1, pp. 99–120. ISSN: 1869-4195. DOI: 10.1007/s13209-015-0134-1. URL: https://doi.org/10.1007/s13209-015-0134-1.

Calvet, Laurent E., Veronika Czellar, and Elvezio Ronchetti (2015). “Robust Filtering”. In: *Journal of the American Statistical Association* 110.512, pp. 1591–1606. DOI: 10.1080/01621459.2014.983520. eprint: https://doi.org/10.1080/01621459.2014.983520. URL: https://doi.org/10.1080/01621459.2014.983520.

Capacci, Sara et al. (2019). “The impact of the French soda tax on prices, purchases and tastes. An ex post evaluation”. In: *Plos One*.

Comte, F. and O. Lieberman (2003). “Asymptotic theory for multivariate GARCH processes”. In: *Journal of Multivariate Analysis* 84.1, pp. 61 –84. ISSN: 0047-259X. DOI: https://doi.org/10.1016/S0047-259X(02)00009-X. URL: http://www.sciencedirect.com/science/article/pii/S0047259X0200009X.

Conrad, Christian and Menelaos Karanasos (2010). “NEGATIVE VOLATILITY SPILLOVERS IN THE UNRESTRICTED ECCC-GARCH MODEL”. In: *Econometric Theory* 26.3, 838–862. DOI: 10.1017/S0266466609990120.

Conrad, Christian and Enzo Weber (2013). “Measuring Persistence in Volatility Spillovers”. In: *Discussion Paper Series / University of Heidelberg, Department of Economics* 0543, p. 52.

Conrad, Jennifer, Mustafa N. Gultekin, and Gautam Kaul (1991). “Asymmetric Predictability of Conditional Variances”. In: *The Review of Financial Studies* 4.4, pp. 597–622. ISSN: 08939454, 14657368. URL: http://www.jstor.org/stable/2962151.

Cox, D. R. et al. (1981). "Statistical Analysis of Time Series: Some Recent Developments [with Discussion and Reply]". In: *Scandinavian Journal of Statistics* 8.2, pp. 93–115. ISSN: 03036898, 14679469. URL: http://www.jstor.org/stable/4615819.

Creal, Drew, Siem Jan Koopman, and André Lucas (2011). "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations". In: *Journal of Business & Economic Statistics* 29.4, pp. 552–563. DOI: 10.1198/jbes.2011.10070. eprint: https://doi.org/10.1198/jbes.2011.10070. URL: https://doi.org/10.1198/jbes.2011.10070.

— (2013). "GENERALIZED AUTOREGRESSIVE SCORE MODELS WITH APPLICATIONS". In: *Journal of Applied Econometrics* 28.5, pp. 777–795. DOI: 10.1002/jae.1279. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.1279. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1279.

Croux, Christophe and Kristel Joossens (2008). "Robust Estimation of the Vector Autoregressive Model by a Least Trimmed Squares Procedure". In: *COMPSTAT 2008*. Ed. by Paula Brito. Heidelberg: Physica-Verlag HD, pp. 489–501. ISBN: 978-3-7908-2084-3.

Darolles, Serge, Christian Francq, and Sébastien Laurent (2018). "Asymptotics of Cholesky GARCH models and time-varying conditional betas". In: *Journal of Econometrics* 204.2, pp. 223 –247. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2018.02.003. URL: http://www.sciencedirect.com/science/article/pii/S0304407618300307.

Davies, Laurie and Ursula Gather (1993). "The Identification of Multiple Outliers". In: *Journal of the American Statistical Association* 88.423, pp. 782–792. ISSN: 01621459. URL: http://www.jstor.org/stable/2290763.

Diewert, W. Erwin, Kevin J. Fox, and Jan de Haan (2016). "A newly identified source of potential CPI bias: Weekly versus monthly unit value price indexes". In: *Economics Letters* 141, pp. 169–172.

Doz, Catherine and Eric Renault (2006). "Factor Stochastic Volatility in Mean Models: A GMM Approach". In: *Econometric Reviews* 25.2-3, pp. 275–309. DOI: 10.1080/07474930600713325. eprint: https://doi.org/10.1080/07474930600713325. URL: https://doi.org/10.1080/07474930600713325.

Durbin, J. and S.J. Koopman (2012). *Time series Analysis by State Space Methods*. Second. Oxford University Press.

Embrechts, Paul, Thomas Mikosch, and Claudia Klüppelberg (1997). *Modelling Extremal Events: For Insurance and Finance*. Berlin, Heidelberg: Springer-Verlag. ISBN: 3-540-60931-8.

Engle, Robert (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4, pp. 987–1007. URL: https://EconPapers.repec.org/RePEc:ecm:emetrp:v:50:y:1982:i:4:p:987-1007.

— (2002). "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models". In: *Journal of Business & Economic Statistics* 20.3, pp. 339–350. ISSN: 07350015. URL: http://www.jstor.org/stable/1392121.

Engle, Robert and Jose Rangel (2008). "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes". In: *Review of Financial Studies* 21.3, pp. 1187–1222. URL: https://EconPapers.repec.org/RePEc:oup:rfinst:v:21:y:2008:i:3:p:1187-1222.

Engle, Robert and Mark Watson (1981a). "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates". In: *Journal of the American Statistical Association* 76.376, pp. 774–781. DOI: 10.1080/01621459.1981.10477720. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1981.10477720. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477720.

Engle, Robert and Mark Watson (1981b). "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates". In: *Journal of the American Statistical Association* 76.376, pp. 774–781. DOI: 10.1080/01621459.1981.10477720. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1981.10477720. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477720.

Engle, Robert F. and Kenneth F. Kroner (1995). "Multivariate Simultaneous Generalized Arch". In: *Econometric Theory* 11.1, pp. 122–150. ISSN: 02664666, 14694360. URL: http://www.jstor.org/stable/3532933.

Engle, Robert F and Gary Lee (1999). "A long-run and short-run component model of stock return volatility". In: *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, pp. 475–497.

Engle, Robert F., Victor K. Ng, and Michael Rothschild (1990). "Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills". In: *Journal of Econometrics* 45.1, pp. 213 –237. ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(90)90099-F. URL: http://www.sciencedirect.com/science/article/pii/030440769090099F.

Fang, Hong-Bin, Kai-Tai Fang, and Samuel Kotz (2002). "The Meta-elliptical Distributions with Given Marginals". In: *Journal of Multivariate Analysis* 82.1, pp. 1 –16. ISSN: 0047-259X. DOI: https://doi.org/10.1006/jmva.2001.2017. URL: http://www.sciencedirect.com/science/article/pii/S0047259X01920172.

Fang, Kai-Tai, Samuel Kotz, and Kai Wang Ng (1990). *Symmetric multivariate and related distributions*. Monographs on statistics and applied probability 36. London [u.a.]: Chapman & Hall. X, 220. ISBN: 0412314304. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+027504395&sourceid=fbw_bibsonomy.

Feenstra, Robert C. and Matthew D. Shapiro (2003a). "High-Frequency Substitution and the Measurement of Price Indexes". In: *Scanner Data and Price Indexes*. Ed. by Robert C. Feenstra and Matthew D. Shapiro. Vol. 64. Studies in Income and Wealth. The University of Chicago Press, pp. 123–150. DOI: 10.7208/chicago/9780226239668.003.0007.

— eds. (2003b). *Scanner Data and Price Indexes*. Vol. 64. Studies in Income and Wealth. Chicago and London: The University of Chicago Press. DOI: 10.7208/chicago/9780226239668.001.0001.

Fiorentini, Gabriele, Enrique Sentana, and Giorgio Calzolari (2003). "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroscedastic Dynamic Regression Models with Student t Innovations". In: *Journal of Business & Economic Statistics* 21.4, pp. 532–546. ISSN: 07350015. URL: http://www.jstor.org/stable/1392326.

Francq, C. and J.M. Zakoian (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley. ISBN: 9781119313489. URL: https://books.google.it/books?id=eaiODwAAQBAJ.

Francq, Christian and Genaro Sucarrat (Jan. 2017). "An Equation-by-equation Estimator of a Multivariate log-GARCH-X Model of Financial Returns". In: *J. Multivar. Anal.* 153.C, pp. 16–32. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2016.09.010. URL: https://doi.org/10.1016/j.jmva.2016.09.010.

Francq, Christian, Olivier Wintenberger, and Jean-Michel Zakoïan (2013). "GARCH models without positivity constraints: Exponential or log GARCH?" In: *Journal of Econometrics* 177.1, pp. 34 –46. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2013.05.004. URL: http://www.sciencedirect.com/science/article/pii/S0304407613001267.

Francq, Christian and Jean-Michel Zakoïan (2012). "QML ESTIMATION OF A CLASS OF MULTIVARIATE ASYMMETRIC GARCH MODELS". In: *Econometric Theory* 28.1, 179–206. DOI: 10.1017/S0266466611000156.

Gallant, A. Ronald, Peter E. Rossi, and George Tauchen (1993). "Nonlinear Dynamic Structures". In: *Econometrica* 61.4, pp. 871–907. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2951766.

Gonzalo, Jesus and Clive Granger (1995). "Estimation of Common Long-Memory Components in Cointegrated Systems". In: *Journal of Business & Economic Statistics* 13.1, pp. 27–35. ISSN: 07350015. URL: http://www.jstor.org/stable/1392518.

Hafner, Christian M. and Helmut Herwartz (2008). "Testing for Causality in Variance Using Multivariate GARCH Models". In: *Annales d'Économie et de Statistique* 89, pp. 215–241. ISSN: 0769489X, 22726497. URL: http://www.jstor.org/stable/27715168.

Hafner, Christian M. and Oliver Linton (2010). "Efficient estimation of a multivariate multiplicative volatility model". In: *Journal of Econometrics* 159.1, pp. 55 –73. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2010.04.007. URL: http://www.sciencedirect.com/science/article/pii/S0304407610001132.

Hamilton, James Douglas (1994). *Time series analysis*. Princeton, NJ: Princeton Univ. Press. XIV, 799. ISBN: 0691042896. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+126800421&sourceid=fbw_bibsonomy.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9780471348054. URL: https://books.google.nl/books?id=3n87CwAAQBAJ.

Hannan, E. J. and M. Deistler (1987). *The Statistical Theory of Linear Systems*. New York, NY, USA: John Wiley & Sons, Inc. ISBN: 0-471-80777-X.

Hansen, Peter Reinhard and Matthias Schmidtblaicher (2019). "A Dynamic Model of Vaccine Compliance: How Fake News Undermined the Danish HPV Vaccine Program". In: *Journal of Business & Economic Statistics* 0.0, pp. 1–21. DOI: 10.1080/07350015.2019.1623045. eprint: https://doi.org/10.1080/07350015.2019.1623045. URL: https://doi.org/10.1080/07350015.2019.1623045.

Harvey, Andrew and Rutger-Jan Lange (2017). "Volatility Modeling with a Generalized t Distribution". In: *Journal of Time Series Analysis* 38.2, pp. 175–190. DOI: 10.1111/jtsa.12224. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jtsa.12224. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12224.

Harvey, Andrew and Stephen Thiele (2016). "Testing against changing correlation". In: *Journal of Empirical Finance* 38. Recent developments in financial econometrics and empirical finance, pp. 575 –589. ISSN: 0927-5398. DOI: https://doi.org/10.1016/j.jempfin.2015.09.003. URL: http://www.sciencedirect.com/science/article/pii/S0927539815000973.

Harvey, Andrew C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Great Britain: Cambridge University Press.

— (2013). *Dynamic models for Volatility and Heavy Tails*. Econometric Society Monograph - Cambridge University Press.

Harvey, Andrew C. and Alessandra Luati (2014). "Filtering With Heavy Tails". In: *Journal of the American Statistical Association* 109.507, pp. 1112–1122. DOI: 10.1080/01621459.2014.887011. eprint: http://dx.doi.org/10.1080/01621459.2014.887011. URL: http://dx.doi.org/10.1080/01621459.2014.887011.

Herwartz, Helmut and Helmut Lütkepohl (2000). "Multivariate volatility analysis of VW stock prices". In: *Int. Syst. in Accounting, Finance and Management* 9, pp. 35–54.

Ito, R. (2016). "Asymptotic Theory for Beta-t-GARCH". In: *Cambridge Working Papers in Economics 1607*.

Ivancic, Lorraine, W. Erwin Diewert, and Kevin J. Fox (2011). "Scanner data, time aggregation and the construction of price indexes". In: *Journal of Econometrics* 161, pp. 24–35. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2010.09.003.

Jeantheau, Thierry (1998). "Strong Consistency of Estimators for Multivariate Arch Models". In: *Econometric Theory* 14.1, pp. 70–86. ISSN: 02664666, 14694360. URL: http://www.jstor.org/stable/3532977.

Jiang, Zhichao and Peng Ding (2016). "Robust modeling using non-elliptically contoured multivariate t distributions". In: *Journal of Statistical Planning and Inference* 177, pp. 50 –63. ISSN: 0378-3758. DOI: https://doi.org/10.1016/j.jspi.2016.04.004. URL: http://www.sciencedirect.com/science/article/pii/S0378375815301026.

Kalman, Rudolph Emil (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D, pp. 35–45.

Kaplan, Greg and Sam Schulhofer-Wohl (2017). "Inflation at the household level". In: *Journal of Monetary Economics* 91, pp. 19–38. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2017.08.002.

Kohn, Robert and Craig F. Ansley (1989). "A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models". In: *Biometrika* 76.1, pp. 65–79. ISSN: 00063444. URL: http://www.jstor.org/stable/2336370.

Koopman, Siem Jan, André Lucas, and Marcel Scharth (2016). "Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models". In: *The Review of Economics and Statistics* 98.1, pp. 97–110. DOI: 10.1162/REST_a_00533. eprint: https://doi.org/10.1162/REST_a_00533. URL: https://doi.org/10.1162/REST_a_00533.

Kotz, S. and S. Nadarajah (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press. ISBN: 9780521826549. URL: https://books.google.it/books?id=dmxtU-TxTi4C.

Ling, Shiqing and Michael McAleer (2002). "Stationarity and the existence of moments of a family of GARCH processes". In: *Journal of Econometrics* 106.1, pp. 109 –117. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(01)00090-2. URL: http://www.sciencedirect.com/science/article/pii/S0304407601000902.

Lucas, Andre, Julia Schaumburg, and Bernd Schwaab (2019). "Bank Business Models at Zero Interest Rates". In: *Journal of Business & Economic Statistics* 37.3, pp. 542–555. DOI: 10.1080/07350015.2017.1386567. eprint: https://doi.org/10.1080/07350015.2017.1386567. URL: https://doi.org/10.1080/07350015.2017.1386567.

Lütkepohl, H. (1996). *Handbook of Matrices*. Handbook of Matrices. Wiley. ISBN: 9780471970156. URL: https://books.google.it/books?id=FuVGbQS2lZoC.

— (2007). *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg. ISBN: 9783540262398. URL: https://books.google.it/books?id=muorJ6FHIiEC.

Magnus, J.R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9781119541202. URL: https://books.google.it/books?id=8sOKDwAAQBAJ.

Martinet, Guillaume Gaetan and Michael McAleer (2018). "On the invertibility of EGARCH(p, q)". In: *Econometric Reviews* 37.8, pp. 824–849. DOI: 10.1080/07474938.2016.1167994. eprint: https://doi.org/10.1080/07474938.2016.1167994. URL: https://doi.org/10.1080/07474938.2016.1167994.

McCaffrey, Daniel F. et al. (1992). "Estimating the Lyapunov Exponent of a Chaotic System With Nonparametric Regression". In: *Journal of the American Statistical Association* 87.419, pp. 682–695. ISSN: 01621459. URL: http://www.jstor.org/stable/2290206.

Melser, Daniel (2018). "Scanner Data Price Indexes: Addressing Some Unresolved Issues". In: *Journal of Business & Economic Statistics* 36.3, pp. 516–522. ISSN: 0735–0015. DOI: 10.1080/07350015.2016.1218339.

Muler, Nora and Victor J. Yohai (2013). “Robust estimation for vector autoregressive models”. In: *Computational Statistics & Data Analysis* 65. Special issue on Robust Analysis of Complex Data, pp. 68 –79. ISSN: 0167-9473. DOI: https://doi.org/10.1016/j.csda.2012.02.011. URL: http://www.sciencedirect.com/science/article/pii/S016794731200093X.

Nakatani, Tomoaki and Timo Teräsvirta (2009). “Testing for volatility interactions in the Constant Conditional Correlation GARCH model”. In: *The Econometrics Journal* 12.1, pp. 147–163. DOI: 10.1111/j.1368-423X.2008.00261.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1368-423X.2008.00261.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1368-423X.2008.00261.x.

Nijman, Theo and Enrique Sentana (1996). “Marginalization and contemporaneous aggregation in multivariate GARCH processes”. In: *Journal of Econometrics* 71.1, pp. 71 –87. ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(94)01695-X. URL: http://www.sciencedirect.com/science/article/pii/030440769401695X.

Oh, Dong Hwan and Andrew J. Patton (2017). “Modeling Dependence in High Dimensions With Factor Copulas”. In: *Journal of Business & Economic Statistics* 35.1, pp. 139–154. DOI: 10.1080/07350015.2015.1062384. eprint: https://doi.org/10.1080/07350015.2015.1062384. URL: https://doi.org/10.1080/07350015.2015.1062384.

Pan, Jianxin and Gilbert Mackenzie (2003). “On modelling Mean-Covariance Structures in Longitudinal Studies”. In: *Biometrika* 90.1, pp. 239–244. ISSN: 00063444. URL: http://www.jstor.org/stable/30042033.

Pedersen, Rasmus Søndergaard (2017). “Inference and testing on the boundary in extended constant conditional correlation GARCH models”. In: *Journal of Econometrics* 196.1, pp. 23 –36. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2016.09.004. URL: http://www.sciencedirect.com/science/article/pii/S0304407616301658.

Pötscher, B.M. and I.R. Prucha (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer Berlin Heidelberg. ISBN: 9783662034866. URL: https://books.google.nl/books?id=pGnmCAAAQBAJ.

Prucha, Ingmar R. and Harry H. Kelejian (1984). “The Structure of Simultaneous Equation Estimators: A Generalization Towards Nonnormal Disturbances”. In: *Econometrica* 52.3, pp. 721–736. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1913473.

Resnick, Sidney (2004). “On the Foundations of Multivariate Heavy-Tail Analysis”. In: *Journal of Applied Probability* 41, pp. 191–212. ISSN: 00219002. URL: http://www.jstor.org/stable/3215977.

Silver, Mick (1995). “Elementary aggregates, micro-indices and scanner data: Some issues in the compilation of consumer price indices”. In: *Review of Income and Wealth* 41.4, pp. 427–438. ISSN: 0034-6586. DOI: 10.1111/j.1475-4991.1995.tb00136.x.

Silver, Mick and Saeed Heravi (2001). “Scanner Data and the Measurement of Inflation”. In: *The Economic Journal* 111.June, pp. 383–404. ISSN: 0013-0133. DOI: 10.1111/1468-0297.00636.

Stout, W.F. (1974). *Almost sure convergence*. Probability and mathematical statistics. Academic Press. ISBN: 9780126727500. URL: https://books.google.it/books?id=QwnvAAAAMAAJ.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics. Springer Berlin Heidelberg. ISBN: 9783540269786. URL: https://books.google.nl/books?id=A0MhAuXDzMcC.

Straumann, Daniel and Thomas Mikosch (2006). “Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach”. In: *The Annals of Statistics* 34.5, pp. 2449–2495. ISSN: 00905364. URL: http://www.jstor.org/stable/25463515.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Non-Linear Time Series. Clarendon Press. ISBN: 9780198522249. URL: https://books.google.it/books?id=w1NyNsXLSlsC.

Tsay, Ruey S. (2005). *Analysis of financial time series*. 2. ed. Wiley series in probability and statistics. Hoboken, NJ: Wiley-Interscience. XXI, 605. ISBN: 978-0-471-69074-0. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+483463442&sourceid=fbw_bibsonomy.

Vaart, A. W. van der (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN: 0-521-49603-9.

Vervaat, Wim (1979). "On a Stochastic Difference Equation and a Representation of Non-Negative Infinitely Divisible Random Variables". In: *Advances in Applied Probability* 11.4, pp. 750–783. ISSN: 00018678. URL: http://www.jstor.org/stable/1426858.

Wang, Fangfang and Eric Ghysels (2015). "ECONOMETRIC ANALYSIS OF VOLATILITY COMPONENT MODELS". In: *Econometric Theory* 31.2, 362–393. DOI: 10.1017/S0266466614000334.

Weide, Roy van der (2002). "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model". In: *Journal of Applied Econometrics* 17.5, pp. 549–564. ISSN: 08837252, 10991255. URL: http://www.jstor.org/stable/4129271.

West, Mike and Jeff Harrison (1997). *Bayesian Forecasting and Dynamic Models (2Nd Ed.)* Berlin, Heidelberg: Springer-Verlag. ISBN: 0-387-94725-6.

White, H. (2001). *Asymptotic Theory for Econometricians*. Economic Theory, Econometrics, and Mathematical Economics. Emerald Group Publishing Limited. ISBN: 9780127466521. URL: https://books.google.nl/books?id=FyG3AAAAIAAJ.

White, Halbert (1994). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. Cambridge University Press. DOI: 10.1017/CCOL0521252806.

Wintenberger, Olivier (2013). "Continuous Invertibility and Stable QML Estimation of the EGARCH(1,1) Model". In: *Scandinavian Journal of Statistics* 40.4, pp. 846–867. DOI: 10.1111/sjos.12038. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjos.12038. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12038.