

Common and Idiosyncratic Conditional Volatility Factors: Theory and Empirical Evidence ^{*}

Francisco Blasques ^{1,2}, Enzo D’Innocenzo ^{1,3†}, Siem Jan Koopman ^{1,2}

¹ *School of Business and Economics, Vrije Universiteit Amsterdam, The Netherlands*

² *Tinbergen Institute Amsterdam, The Netherlands*

³ *Department of Statistical Sciences, University of Bologna, Italy*

This version: June 21, 2021

Abstract

We propose a multiplicative dynamic factor structure for the conditional modelling of the variances of an N -dimensional vector of financial returns. We identify common and idiosyncratic conditional volatility factors. The econometric framework is based on an observation-driven time series model that is simple and parsimonious. The common factor is modeled by a normal density and is robust to fat-tailed returns as it averages information over the cross-section of the observed N -dimensional vector of returns. The idiosyncratic factors are designed to capture the erratic shocks in returns and therefore rely on fat-tailed densities. Our model is potentially of a high-dimension, is parsimonious and it does not necessarily suffer from the curse of dimensionality. The relatively simple structure of the model leads to simple computations for the estimation of parameters and signal extraction of factors. We derive the stochastic properties of our proposed dynamic factor model, including bounded moments, stationarity, ergodicity, and filter invertibility. We further establish consistency and asymptotic normality of the maximum likelihood estimator. The finite sample properties of the estimator and the reliability of our method to track the common conditional volatility factor are investigated by means of a Monte Carlo study. Finally, we illustrate our approach with two empirical studies. The first study is for a panel of financial returns from ten stocks of the S&P100. The second study is for the panel of returns from all S&P100 stocks.

Key words: Financial econometrics, observation-driven models, conditional volatility, common factor.

JEL classification: C32, C52, C58.

^{*}The authors are thankful to Alessandra Luati for helpful comments and suggestions.

[†]Corresponding author: enzo.dinnocenzo2@unibo.it.

1 Introduction

Many different approaches for the econometric and statistical analysis of multiple time series exist. Multivariate extensions of autoregressive models and other linear dynamic models are well established. In case the dimension of the vector of observations for each time period increases, we encounter the *curse of dimensionality* problem which refers to the corresponding exponential increase in the number of parameters. To alleviate this dimension problem, rank-reduction techniques have been considered, including principal components and dynamic factor methods. Such approaches are widely adopted in empirical studies in macroeconomics, finance and macro-finance; see, for example, Forni et al. (2000), Stock and Watson (2002, 2012) and Doz et al. (2011). In these studies, the rank-reduction methods are applied to linear dynamic model structures and for modeling the location or mean function of the observed time series vector. In this study, we will focus on the dynamic modeling of the scale or variance function of the observation vector by means of parsimonious model specifications. In particular, we jointly analyse conditional variances within a vector of financial returns.

In empirical financial studies, conditional volatility in a time series of financial returns is usually analyzed by the generalized autoregressive conditional heteroskedasticity (GARCH) model as proposed in the seminal work of Engle (1982) and Bollerslev (1986). Given the presence of cross-sectional dependence and common features in financial returns, it is widely acknowledged that it is useful to analyze multiple time series of financial returns jointly within a multivariate model. For example, the first multivariate extension of the GARCH model has been proposed by Bollerslev et al. (1988) and is known as the (D)VECH-GARCH model. This is a typical example of a multivariate extension that suffers from the curse of dimensionality. Another multivariate extension of the GARCH model is known as the BEKK-GARCH model and is proposed by Engle and Kroner (1995). While the scalar BEKK specification is a parsimonious formulation of a multivariate GARCH model, it is also a rather restrictive setting in which the dynamic properties of all variances and covariances are the same. More general

specifications of the VEC- and BEKK-GARCH models suffer from infeasible estimation methods due to the exponentially increasing number of parameters and the intractable number of time-varying conditional variances and covariances that need to be analyzed. When the focus is more concerned with time-varying conditional correlations between time series of financial returns, a parsimonious and practical approach is offered by the dynamic conditional correlation (DCC) model of Engle (2002). The DCC offers only a disjoint analysis of variances and correlations. In any case, all multivariate extensions of the GARCH model have enjoyed much popularity due to their ability to capture covariance spillovers and feedbacks such that key interactions between financial returns can be uncovered. A complete review of multivariate GARCH models is provided by Bauwens et al. (2006).

Despite the advance of computing power and efficient estimation methods, an analysis based on the multivariate GARCH model is still regarded as involved due to its curse of dimensionality. To counter this dimension problem for multivariate GARCH models, the aforementioned reduced rank methods have been considered in this context. The factor GARCH model is originally proposed by Engle et al. (1990) and it has proven to be a popular method as it reduces the dimension problem substantially. This approach has been developed further by Alexander (2001) and van der Weide (2002) with their respective orthogonal GARCH (O-GARCH) model and its generalized version, the GO-GARCH model.

In our proposed dynamic conditional variance factor model, the variances of a high-dimensional time series vector of financial returns are decomposed into common and idiosyncratic factors. It allows for an insightful and parsimonious way of simultaneously analyzing dependence structures in a time series panel of financial returns. The common factor in the model should naturally be robust to fat-tailed innovations due to averaging of information over the cross-section. The idiosyncratic factors are typically sensitive to particular events and outliers and should be represented by fat-tailed innovations. In this paper, we propose a multiplicative dynamic conditional variance factor model with a robust updating equation for the idiosyncratic factors.

Our model can be regarded as a multiplicative version of the one-factor (or index) multivariate

time series models which are developed by Engle and Watson (1981) and Gonzalo and Granger (1995), amongst others. However, in adopting the terminology of Cox (1981), we propose an observation-driven dynamic specification for the factors. In this class of dynamic models, we have the main advantage that the predictive likelihood is available in closed-form. Hence, parameter estimation can be carried out using standard likelihood-based methods, even when we drop the Gaussian assumption and introduce nonlinearities in the model. In particular, relying on a simple system of non-linear recursions, our model decomposes the conditional variance of each series of daily returns into the product of common and idiosyncratic components.

Our proposed model can further be regarded 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), where the aforementioned one-factor multivariate time series model approach is considered within the volatility context. Even though this approach has shown to be a powerful tool for capturing complex volatility structures and dynamics, a comprehensive statistical and probabilistic analysis of the multiplicative component models has not been available until the recent work of Wang and Ghysels (2015).

In a different context, Barigozzi et al. (2014) propose a vector multiplicative error model for the identification of common and idiosyncratic dynamic factors in the time series. The common trend factor is extracted using a Nadaraya-Watson estimator, while idiosyncratic factors are individually modeled as scalar and asymmetric GARCH processes in a parametric setting. Hence, a semi-nonparametric analysis is developed. In contrast, we develop a fully parametric framework by specifying a simple univariate GARCH filter for the common factor and we formulate a robust score-driven filter for the idiosyncratic factors based on the Student's t distribution; see the discussions in Creal et al. (2013) and Harvey (2013).

The remainder of the paper is organized as follows. Section 2 introduces our dynamic conditional variance factor model. Section 3 studies its stochastic properties. Section 4 establishes the asymptotic properties of the maximum likelihood estimator (MLE). Section 5 presents a Monte Carlo study in

which we analyze the small sample properties of the MLE and the accuracy of our method in extracting the common conditional volatility factor. Section 6 provides two empirical illustrations using a panel of 10 stocks from the S&P100 and a panel of all S&P100 stocks. Section 7 concludes.

2 Dynamic Conditional Variance Factor Model

Let $\mathbf{x}_t = (x_{1t}, \dots, x_{Nt})^\top \in \mathbb{R}^N$ denote an N -dimensional vector of stock returns at time t and let $\mathcal{F}_{t-1} = \mathcal{F}\{\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots\}$ be the *sigma-field* generated by the past values of \mathbf{x}_t . For each asset $i = 1, \dots, N$, we consider a volatility model, in which the time-varying conditional variance is defined as the product of a dynamic common factor $\{f_t^2\}_{t \in \mathbb{Z}}$, which captures the commonalities between assets, and an idiosyncratic volatility process of the considered asset, $\{\sigma_{it}^2\}_{t \in \mathbb{Z}}$. The model for the stock return of each individual asset i is given by the equation

$$x_{it} = f_t \sigma_{it} \epsilon_{it}, \quad (1)$$

where $\{\epsilon_{it}\}_{t \in \mathbb{Z}}$ for $i = 1, \dots, N$, is an independent and identically distributed (*IID*) sequence of random variables. The dynamic specifications of the common factor variable f_t and each idiosyncratic factor σ_{it}^2 are specified as autoregressive processes with innovations that are resulting from the score of the conditional predictive density as a function of past observations. In case of the i -th idiosyncratic variance, for $i = 1, \dots, N$, we have the following updating recursion

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

where $\delta_i > 0$ is a scalar intercept, $0 < \phi_i < 1$ is an autoregressive coefficient, and $\kappa_i > 0$ is the weight coefficient for the score function s_{it} , for $t = 1, \dots, T$. For the class of score-driven models introduced by Creal et al. (2013), the score s_{it} is defined as the conditional scaled score of the predictive log-density. In particular, we consider the scaled observations $\{x_{it}/f_t\}_{t \in \mathbb{Z}}$, such that we have

$$s_{it} = S_{it} \nabla_{it}(\sigma_{it}), \quad \nabla_{it}(\sigma_{it}) = \frac{\partial \log p(x_{it}/f_t | \sigma_{it}^2, \delta_i, \phi_i, \kappa_i)}{\partial \sigma_{it}^2},$$

where $\nabla_{it}(\sigma_{it})$ is the conditional score and S_{it} is the scaling for the score. Common choices for the latter component are the unit scaling, that is $S_{it} = 1$, the square root of inverse of the Fisher information scaling, that is $S_{it} = \mathcal{I}(\sigma_{it})^{-1/2}$, and the inverse of the Fisher information scaling, $S_{it} = \mathcal{I}(\sigma_{it})^{-1}$, where

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

For a more detailed discussion we refer to Creal et al. (2013) and Harvey (2013). We further assume that the *IID* sequence $\{\epsilon_{it}\}_{t \in \mathbb{Z}}$ for $i = 1, \dots, N$ in model (1), follows a standard Student's t distribution with ν_i degrees of freedom, zero mean and unit scale, that is

$$\epsilon_{it} \sim t_{\nu_i}(0, 1),$$

for $i = 1, \dots, N$. The specifications for each of the i -th idiosyncratic sequence $\{\sigma_{it}\}_{t \in \mathbb{Z}}$ is then completed. It follows that the idiosyncratic sequences are positive conditionally predictable processes that evolve as a first-order *Beta-t-GARCH*, see Harvey (2013). An updating specification for the common factor can be obtained in a similar way.

In our current study, we propose the following updating equations for the common factor f_t^2 and the idiosyncratic factor σ_i^2 , for $i = 1, \dots, N$,

$$f_{t+1}^2 = \omega + \alpha \left(\frac{1}{N} \sum_{i=1}^N x_{it}^2 - f_t^2 \right) + \beta f_t^2, \quad (2)$$

$$\sigma_{i,t+1}^2 = \delta_i + \phi_i \sigma_{it}^2 + \kappa_i \sigma_{it}^2 \left[\frac{(\nu_i + 1)(x_{it}^2/f_t^2)}{(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)} - 1 \right], \quad (3)$$

where $\omega > 0$ is a scalar intercept, $0 < \beta < 1$ is an autoregressive coefficient and $\alpha > 0$ is a scaling parameter. The updating equation (2) is simply based on the common prediction error, defined as the difference between the cross-sectional sample variance $\frac{1}{N} \sum_{i=1}^N x_{it}^2$ and the common factor f_t^2 . The updating equation (3) for the idiosyncratic factor σ_{it}^2 is based on the score-driven mechanism as described above. Our dynamic conditional variance factor model is defined by (1), (2) and (3).

Assumption 1 below imposes constraints which ensure the positivity of the common conditional volatility f_{t+1}^2 .

Assumption 1. *The dynamic parameters of the recursion (2) satisfy $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\beta - \alpha \geq 0$.*

Analogously, with Assumption 2, we give the positivity constraints for the static parameters in (3). For identification purposes, we need to fix the unknown intercepts as $\delta_i = (1 - \phi_i)$ for $i = 1, \dots, N$, in order to ensure that the unconditional mean of the idiosyncratic conditional volatility in (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. *The dynamic parameters of the recursion (3) satisfy $\delta_i = (1 - \phi_i) > 0$, $\kappa_i \geq 0$, $\phi_i \geq 0$ and $\phi_i - \kappa_i \geq 0$, for $i = 1, \dots, N$. Moreover, we have $\min_{i=1, \dots, N} \nu_i > 2$.*

We can represent our dynamic conditional variance factor model as a set of filtering equations, that is

$$\begin{aligned} x_{it} &= f_t \sigma_{it} \epsilon_{it}, \\ f_{t+1}^2 &= \omega + \alpha \frac{1}{N} \sum_{i=1}^N x_{it}^2 + (\beta - \alpha) f_t^2, \\ \sigma_{i,t+1}^2 &= \delta_i + \left[(\phi_i - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^2 / f_t^2)}{(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2 / f_t^2)} \right] \sigma_{it}^2. \end{aligned} \quad (4)$$

The set of recursions in (4) are those actually implemented for computational purposes and requires a proper statistical treatment which we provide in Subsection 3.2 and Section 4 below.

We define the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\lambda}^\top, \boldsymbol{\psi}_1^\top, \dots, \boldsymbol{\psi}_N^\top)^\top$, 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 $\boldsymbol{\psi}_i \cap \boldsymbol{\psi}_j = \emptyset$ 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. Therefore, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$, where $\boldsymbol{\Theta}$ is the parameter space and $p = 3 + 4N$.

3 Statistical properties of the model

3.1 Stationarity, ergodicity & moments

We explore the stationary properties of the time series data generated by model (1). For this purpose, we reformulate the equations in (4). In particular, we study the statistical properties for the system of equations for the asset $i = 1, \dots, N$ at time t given by

$$\begin{aligned} x_{it} &= f_t \sigma_{it} \epsilon_{it}, \\ f_{t+1}^2 &= \omega + \left[\alpha \left(\frac{1}{N} \sum_{i=1}^N \sigma_{it}^2 \epsilon_{it}^2 - 1 \right) + \beta \right] f_t^2, \\ \sigma_{i,t+1}^2 &= \delta_i + \left[\kappa_i \left(\frac{(\nu_i + 1) \epsilon_{it}^2}{(\nu_i - 2) + \epsilon_{it}^2} - 1 \right) + \phi_i \right] \sigma_{it}^2. \end{aligned} \quad (5)$$

Proposition 1 establishes the strict stationarity and ergodicity of the data generation process of the dynamic conditional variance factor model. These results can be obtained after we have shown that both the common factor volatility $\{f_t^2\}_{t \in \mathbb{Z}}$ and the idiosyncratic volatilities $\{\sigma_{it}^2\}_{t \in \mathbb{Z}}$, for $i = 1, \dots, N$, are themselves stationary and ergodic under the following parameter restrictions.

Assumption 3. *The autoregressive parameters of the recursions in (5) satisfy $|\phi_i| < 1$ for $i = 1, \dots, N$, and $|\beta| < 1$.*

Proposition 1. *Let Assumptions 1–3 be satisfied. Then, the common 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 for $i = 1, \dots, N$. Moreover, the multiplicative components sequence $\{f_t \sigma_t^2\}_{t \in \mathbb{Z}}$, with $\sigma_t^2 = (\sigma_{1t}^2, \dots, \sigma_{Nt}^2)^\top$, 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.*

Next we concentrate on the number of bounded moments of model (1). Proposition 2 shows that the data simulated from this model has m bounded moments, where m is a function of the bounded moments n_f of the common factor, the moments $n_\sigma = \min_{i=1, \dots, N} n_{\sigma_i}$, the minimum number of bounded moments between the N -idiosyncratic processes, and the moments n_ϵ of the IID sequence.

Proposition 2. *Under assumptions 1–3, the stationary ergodic series $\{x_{it}\}_{t \in \mathbb{Z}}$ has m bounded moments, that is $\mathbb{E}[|x_{it}|^m] < \infty$ where*

$$m = \frac{n_f n_\sigma n_\epsilon}{n_\sigma n_\epsilon + n_f n_\epsilon + n_f n_\sigma}. \quad (6)$$

Clearly, $m > 2$.

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 $\mathbb{E}[x_{it} | \mathcal{F}_{t-1}] = 0$ and $\mathbb{V}[x_{it} | \mathcal{F}_{t-1}] = (f_t \sigma_{it})^2$. In addition, $\mathbb{V}[x_{it}] = f_t^2$.

3.2 Invertibility

Next we focus on the invertibility property of the updating equations in model (4), and consider 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. 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 $i = 1 \dots, N$. From these starting points we retrieve the estimated paths $\{\hat{f}_t^2\}_{t \in \mathbb{N}}$ and $\{\hat{\sigma}_{it}^2\}_{t \in \mathbb{N}}$ for $i = 1 \dots, N$. It is clear then these initializations will crucially affect the whole filtering process, since the actual rescaled 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 (5). Assumption 4 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 4. *The parameter space $\Theta = \{\Lambda \times \Psi\}$ is compact and satisfies*

1. $|\beta - \alpha| < 1$ and
2. $\max_{i=1, \dots, N} \{|\phi_i - \kappa_i|, |\phi_i + \kappa_i \nu_i|\} < 1$.

Unlike other recent articles which also deal with robust nonlinear filtering methods for time-varying locations (e.g. Harvey and Luati (2014) or 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 with the contraction theorem for perturbed stochastic recursions stated in Theorem 2.10 of 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 4. Following the notation of Straumann and Mikosch (2006) we shall use e.a.s. as a shorthand for the exponentially fast almost surely convergence.

Proposition 3. *Consider the N -dimensional vector process $\{\mathbf{x}_t\}_{t \in \mathbb{Z}}$ be generated from model (1) under assumptions 1–3, such that is stationary and ergodic. In addition, impose Assumption 4, 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 $i = 1 \dots, N$ converge e.a.s. and uniformly to their respective stationary and ergodic solutions $\{f_t\}_{t \in \mathbb{Z}}$ and $\{\sigma_{it}^2\}_{t \in \mathbb{Z}}$ for $i = 1 \dots, N$, that is*

$$\sup_{\theta \in \Theta} \|\hat{f}_t^2 - f_t^2\| \xrightarrow{e.a.s.} 0 \quad \text{and} \quad \sup_{\theta \in \Theta} \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \xrightarrow{e.a.s.} 0 \quad \text{as} \quad t \rightarrow \infty, \quad (7)$$

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

$$\sup_{\theta \in \Theta} \|\hat{f}_t^2 \hat{\boldsymbol{\sigma}}_t^2 - f_t^2 \boldsymbol{\sigma}_t^2\| \xrightarrow{e.a.s.} 0 \quad \text{as} \quad t \rightarrow \infty.$$

4 Maximum likelihood estimation

Having established the stochastic properties of our factor model both as a data generating process (Subsection 3.1) and as a filter (Subsection 3.2), we can now turn to the sample properties of the maximum likelihood estimators (MLE). In this section, we introduce the maximum likelihood procedure for the estimation of the parameters in model (1), we write the time-varying common factor and the idiosyncratic terms as explicit functions of the unknown parameter vectors, that is $f_t^2 = f_t^2(\boldsymbol{\lambda})$ and $\sigma_{it}^2 = \sigma_{it}^2(\boldsymbol{\psi}_i)$ for $i = 1, \dots, N$.

The observation density of x_{it} for $i = 1, \dots, N$ is given by

$$p(x_{it} | \mathcal{F}_{t-1}, \boldsymbol{\lambda}, \boldsymbol{\psi}_i) = \frac{\Gamma[(\nu_i + 1)/2]}{\Gamma[\nu_i/2] \sqrt{\pi(\nu_i - 2)} (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} \left[1 + \frac{x_{it}^2}{(\nu_i - 2)(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} \right]^{-\frac{\nu_i + 1}{2}}, \quad (8)$$

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

$$\begin{aligned} \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) = & \log \Gamma \left[\frac{\nu_i + 1}{2} \right] - \log \Gamma \left[\frac{\nu_i}{2} \right] - \frac{1}{2} \log(\nu_i - 2) - \frac{1}{2} \log \pi \\ & - \frac{1}{2} \log [f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)] - \frac{\nu_i + 1}{2} \log \left[1 + \frac{x_{it}^2}{(\nu_i - 2)(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} \right]. \end{aligned} \quad (9)$$

Therefore, the MLE boils down as the solution of

$$\hat{\boldsymbol{\theta}}_{NT} = (\hat{\boldsymbol{\lambda}}_T^\top, \hat{\boldsymbol{\psi}}_{1T}^\top, \dots, \hat{\boldsymbol{\psi}}_{NT}^\top)^\top = \arg \max_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N \in \boldsymbol{\Psi}} \sum_{i=1}^N \sum_{t=1}^T \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i). \quad (10)$$

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 $i = 1, \dots, N$ and the infeasible log-likelihood function where its starting values are drawn from the stationary distributions, respectively. However, every marginal likelihood depends on the common factor when $N > 1$, hence we have

$$\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{\ell}_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \quad \text{and} \quad \mathcal{L}_{NT}(\boldsymbol{\theta}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i). \quad (11)$$

Then, we define the limit

$$\mathcal{L}_N(\boldsymbol{\theta}) = \frac{1}{N} \mathbb{E} \left[\sum_{i=1}^N \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \right]. \quad (12)$$

4.1 Consistency

This section establishes the consistency of the MLE as $T \rightarrow \infty$. Following White (1994), we obtain the strong consistency of the MLE in (10) from the uniform convergence of the log-likelihood function $\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})$ in (11) to its limit $\mathcal{L}_N(\boldsymbol{\theta})$ as defined in equation (12) and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \Theta$. To this end, we need a further assumption which ensure that the unknown true parameter vector $\boldsymbol{\theta}_0$ belongs to the compact parameter space Θ . Moreover, Assumption 5 below ensures that the criterion can be differentiated around $\boldsymbol{\theta}_0$.

Assumption 5. *The true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\lambda}_0^\top, \boldsymbol{\Psi}_0^\top)^\top$ belongs to the interior of Θ , i.e. $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$.*

To prove strong consistency for T large, we make use of two preliminary Lemmas which facilitate the discussion in proving the main consistency theorem. Lemma 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 compact parameter space Θ . Lemma 4.2 obtains the identification of the true parameter $\boldsymbol{\theta}_0 \in \Theta$.

Lemma 4.1. *Let Assumptions 1–5 hold. Then, the limit $\mathcal{L}_N(\boldsymbol{\theta})$ is well defined and $\mathbb{E}[|\ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)|] < \infty$ for $i = 1, \dots, N$ and $t = 1, \dots, T$. Then,*

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\mathcal{L}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0 \quad \text{as } T \rightarrow \infty. \quad (13)$$

Lemma 4.2. *Under the Assumptions of Lemma 4.1, $\mathcal{L}_N(\boldsymbol{\theta})$ is uniquely maximized at the true parameter $\boldsymbol{\theta}_0$ i.e. $\mathcal{L}_N(\boldsymbol{\theta}_0) > \mathcal{L}_N(\boldsymbol{\theta}) \forall \boldsymbol{\theta} = (\boldsymbol{\lambda}^\top, \boldsymbol{\Psi}^\top)^\top \in \Theta$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.*

We are now in position to prove the main result of this Subsection. Theorem 4.3 establishes the consistency of the MLE for the well-specified dynamic factor model, as $T \rightarrow \infty$.

Theorem 4.3 (Strong consistency). *Consider model (1), satisfying Assumptions 1–5. Then,*

$$\hat{\boldsymbol{\theta}}_{NT} \rightarrow \boldsymbol{\theta}_0 \text{ almost surely as } T \rightarrow \infty. \quad (14)$$

4.2 Asymptotic normality

We finally turn to the asymptotic normality of the MLE. The proof of this result builds upon Theorem 6.2 of White (1994). To this end, let us define $\nabla = \frac{\partial}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial (\boldsymbol{\lambda}, \boldsymbol{\Psi})}$. Then, we prove the asymptotic normality of the MLE in (10) by ensuring that the following conditions hold true: (a) the strong consistency of $\hat{\boldsymbol{\theta}}_{NT}$, (b) the empirical log-likelihood $\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})$ defined in (11) is twice continuously differentiable in $\boldsymbol{\theta} \in \Theta$ almost surely, (c) the asymptotic normality of the score function $\nabla \mathcal{L}_{NT}(\boldsymbol{\theta})$, (d) the uniform convergence of the second derivative of the log-likelihood function $\nabla \mathcal{L}_{NT}(\boldsymbol{\theta})$ over the compact parameter space Θ and finally, (e) the non-singularity of the Fisher's information matrix.

As noted before, the proof of asymptotic normality of the MLE require several steps. We begin by exploring the limit behaviour of the derivative processes $\{\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ and its perturbed version $\{\nabla(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$. The perturbed derivative process $\{\nabla(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ arises from the fact that the derivatives in $\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))$ are nonlinear functions of the nonstationary filtered $\{\hat{f}_t^2(\boldsymbol{\lambda})\}_{t \in \mathbb{N}}$ and $\{\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{N}}$ for $i = 1 \dots, N$, which will be stationary only in the limit, as shown in Proposition 3.

Therefore, to ensure that the derivative processes the derivative processes $\{\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ for $i = 1, \dots, N$, provide stationary approximations to $\{\nabla(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ for $i = 1, \dots, N$, we introduce the following Proposition.

Proposition 4. *Under Assumptions 1–5, the derivative sequences $\{\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ converges e.a.s. to a unique stationary ergodic solution $\{\nabla(f_t^2 \boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{Z}}$. Moreover, for any initializations of the filters $(\hat{f}_1^2 \hat{\boldsymbol{\sigma}}_1^2) \in \mathbb{R}_+^{N+1}$, the perturbed derivatives $\{\nabla(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ will converge e.a.s. to the same stationary ergodic solution $\{\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{Z}}$ for $i = 1, \dots, N$.*

The number of bounded moments 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. In addition to the notation of Proposition 2, we denote with n'_f the number of bounded moments of the derivatives of the common factor $\nabla f_t^2(\boldsymbol{\lambda})$ and with $n'_\sigma = \min_{i=1, \dots, N} n'_{\sigma_i}$ the minimum between the number of bounded moments of the derivatives of the idiosyncratic terms $\nabla \sigma_{it}^2(\boldsymbol{\psi}_i)$.

Proposition 5. *Under Assumptions 1–5, $\{\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{Z}}$ has m' bounded moments uniformly over the parameter space, that is $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\|^{m'}] < \infty$ where*

$$m' = \frac{n'_f n_\sigma}{n_\sigma + n'_f} + \frac{n_f n'_\sigma}{n'_\sigma + n_f}. \quad (15)$$

Clearly, $m' > 2$.

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

Lemma 4.4. *Under Assumptions 1–5, the derivative of the conditional likelihood $\{\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)\}$ is a martingale difference sequence, i.e. $\mathbb{E}[\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) | \mathcal{F}_{t-1}] = 0$ and moreover $\mathbb{E}[|\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)|^2]$ exists and is finite.*

Now we establish the asymptotic distribution of the score function.

Lemma 4.5. *Under Assumptions 1–5, we obtain*

$$\sqrt{NT} \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{V}), \quad \text{as } T \rightarrow \infty, \quad \text{where}$$

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

Next, we enter in the realm of the second derivatives and present a new proposition.

Proposition 6. *Under Assumptions 1–5, the second derivative $\{\nabla^2(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ converges e.a.s. to a unique stationary ergodic solution $\{\nabla^2(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{Z}}$. Moreover, for any initializations of the filter $(\hat{f}_1^2 \hat{\sigma}_1^2) \in \mathbb{R}_+^{N+1}$ and $\nabla(\hat{f}_1^2 \hat{\sigma}_1^2) \in \mathbb{R}_+^{4N+3}$, the perturbed second derivatives $\{\nabla^2(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{N}}$ will converge e.a.s. to the same stationary ergodic solution $\{\nabla^2(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{Z}}$ for $i = 1, \dots, N$.*

Finally, we analyze the second differential of the conditional log-likelihood function. 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 4.6. *Let Assumptions 1–4 hold. Then, the limit $\nabla^2 \mathcal{L}_N(\boldsymbol{\theta})$ is well defined and moreover it is a nonsingular matrix for $i = 1, \dots, N$ and every $t = 1, \dots, T$. Then,*

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta}) - \nabla^2 \mathcal{L}_N(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0 \quad \text{as } T \rightarrow \infty. \quad (16)$$

In conclusion, we present the last theorem which shows the asymptotic Gaussianity of the MLE.

Theorem 4.7 (Asymptotic normality of the MLE). *Consider model (1), satisfying Assumptions 1–5. Then,*

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

where for $i = 1, \dots, N$

$$\mathcal{I}_i(\boldsymbol{\theta}_0) = -\mathbb{E} \left[\frac{\partial^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \partial(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)^\top} \Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0, \boldsymbol{\psi}_i=\boldsymbol{\psi}_{i0}} \right]$$

is the Fisher information matrix evaluated at the true parameter vectors $\boldsymbol{\lambda}_0$ and $\boldsymbol{\psi}_{i0}$.

5 Monte Carlo experiments

In this Section we present two Monte Carlo simulation experiments. The first study is concerned with the finite sample behaviour of the MLE. In order to assess the reliability of the estimation method described in Section 4, we generate time series which mimic the properties of the dataset used in the empirical application in Subsection 6.1. The second study deals with the accuracy of the dynamic factor model specified by recursion (2) in tracking a common volatility pattern when the panel of time series is large. In this second experiment we aim to replicate the data from the whole panel of time series of S&P100, used in Subsection 6.2.

5.1 Finite sample properties of the MLE

In order to study the finite sample properties of the MLE, we perform a Monte Carlo experiment. With this simulation study, we aim to investigate and assess the reliability of the asymptotic theory developed in Section 4. To this end, we chose to simulate a panel of time series directly from our dynamic factor model with Student's t errors, introduced in Section 2, which satisfies the condition given in Assumptions 1-4.

In particular, we generate time series from a fixed data generating process which mimic the dynamic structure of a panel of stock returns that shows strong factor structure. With these experiments we aim to establish the accuracy of the proposed estimation procedure for empirically relevant cases. The simulation scheme is as follows. We fix the dimension of the panel of the generated data to $N = 10$, and the real data generating process for $i = 1, \dots, 10$ is

$$\begin{aligned}x_{it} &= f_t \sigma_{it} \epsilon_{it} & \epsilon_{it} &\sim t_5(0, 1), \\f_{t+1}^2 &= 0.10 + \left[0.05 \left(\frac{1}{10} \sum_{i=1}^{10} \sigma_{it}^2 \epsilon_{it}^2 - 1 \right) + 0.95 \right] f_t^2, \\ \sigma_{i,t+1}^2 &= (1 - 0.90) + \left[0.10 \left(\frac{(5+1)\epsilon_{it}^2}{(5-2) + \epsilon_{it}^2} - 1 \right) + 0.90 \right] \sigma_{it}^2.\end{aligned} \tag{17}$$

We replicate $M = 1000$ simulations and we generate two different set of 10 time series, where the

first set has $T = 1000$ observations while the second has $T = 2000$ observations. Note that the data generating process in (17) satisfy Assumptions 1–3, which ensure that the data generated by that model are stationary and ergodic by Proposition 1.

In our simulations the data generated by the model in (17) will inherits the empirical properties of financial returns with cross-sectional dependence and persistent idiosyncratic terms. Moreover, by fixing the innovations ϵ_{it} to be Student’s t IID random variables with $\nu_i = 5$ degrees of freedom for $i = 1, \dots, N$, the generated data will show with high probability extreme values, which mimic the well-known fat-tailed nature of financial stock returns.

The results are reported in Table 1 in terms of empirical bias, root mean square error (*RMSE*) and the coverage rate of the asymptotic 90% confidence interval (CI), calculated by adding and removing to the Gaussian quantile the estimated standard errors. For the idiosyncratic terms, we report the average of the same measures across all the time series in the panel.

Table 1: Monte Carlo simulations. The table reports the *bias*, *RMSE* and coverage of the 90% confidence interval of the common factor and idiosyncratic parameter estimators. $M = 1000$ replications for $N = 10$ and $T = \{1000, 2000\}$.

	<i>Bias</i>		<i>RMSE</i>		90% <i>CI coverage</i>	
	$T = 1000$	$T = 2000$	$T = 1000$	$T = 2000$	$T = 1000$	$T = 2000$
$f_t^2(\lambda)$						
ω	0.048	0.028	0.105	0.090	0.915	0.902
α	0.068	0.008	0.057	0.018	0.909	0.904
β	-0.042	-0.012	0.087	0.026	0.889	0.894
$\sigma_{it}^2(\psi_i)$						
ϕ_i	-0.076	-0.039	0.064	0.047	0.884	0.894
κ_i	0.015	0.005	0.023	0.016	0.917	0.910
ν_i	0.102	0.081	0.553	0.501	0.913	0.909

Results suggest that the MLE deliver satisfactory results in both the common factor and the idiosyncratic terms since the estimated biases are very small for all parameters. Notably, as the length of the

time series T increases from 1000 to 2000 the empirical biases and root mean square errors show a sensible reduction, which is in line with the asymptotic theory of the MLE discussed in Section 4. However, we note downward biases for the estimated autoregressive coefficients, namely β for the common factor and ϕ_i for $i = 1, \dots, N$ which is also reflected in the coverage rate of the asymptotic 90% confidence interval. In fact, the coverage rate, is slightly smaller than the fixed 90% level. Therefore, the upward biases of all the other coefficient may be explained as a compensation for the downward biases of the autoregressive coefficients. Nevertheless, the estimated coverage rate of the asymptotic confidence interval is really close to the nominal 90% level and therefore, we can conclude that the maximum likelihood procedure provides reliable results for medium-high dimensions of panel of fat-tailed stock returns with strong factor structure and persistent idiosyncratic volatilities.

5.2 Tracking the common factor

In this second Monte Carlo experiment we aim to show that our dynamic factor model, provides a satisfactory and parsimonious approach to track the common fluctuation of volatility in a large panel of time series. To this end, we consider the following stochastic pattern for the common factor:

$$\begin{aligned} f_t^2 &= \exp \tau_t \\ \tau_{t+1} &= 0.003 + 0.90\tau_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, 0.065), \end{aligned} \tag{18}$$

We fix $N = 100$, and the data-generation process for $i = 1, \dots, 100$ is given by

$$x_{it} = f_t \epsilon_{it} \quad \epsilon_{it} \sim t_{\nu_i}(0, 1), \tag{19}$$

where we let the degrees of freedom ν_i for $i = 1, \dots, 100$ range in the following set $\nu = \{5, 10, 15, \infty\}$. Clearly, the case where $\nu = \infty$ will cover the Gaussian case, i.e. $\epsilon_{it} \sim \mathcal{N}(0, 1)$. From the data generating process defined by equations 18 and 19, we generate $M = 1000$ time series each with $T = 1000$ observations. We benchmark our dynamic factor recursion in equation (3) against an alternative approach, that consists in the following two-step procedure: First, we take the average of of the simulated

returns, and second we apply a standard GARCH model on the average. In practice, the procedure can be summarized by two equations,

$$y_t = \frac{1}{100} \sum_{i=1}^{100} x_{it}$$

$$f_{t+1}^2 = \omega + \alpha y_t^2 + \beta f_t^2. \quad (20)$$

We label the procedure in (20) as *sGARCH*. As measures of accuracy we use the mean absolute error (*MAE*) and the mean squared error (*MSE*), defined as

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{f}_t^2 - f_t^2|, \quad MSE = \frac{1}{T} \sum_{t=1}^T (\hat{f}_t^2 - f_t^2)^2,$$

where for each simulation, with \hat{f}_t^2 we denote the estimated paths of the common factor with the recursion in equation (3) and (20), while f_t^2 denotes the true paths generated by (18). Moreover, to determine the procedure between the recursions in (3) and (20) that provides the best approximation of the true process in (18), we regress the true paths f_t^2 against the estimated paths \hat{f}_t^2 , and report the resulting R^2 .

The results are summarized in Table 2, where we report averages across all the 1000 Monte Carlo replications.

Table 2: *MAE*, *MSE* and R^2 results. We perform $M = 1000$ replications for $N = 100$ and $T = 1000$.

	$\nu = 5$			$\nu = 10$			$\nu = 15$			$\nu = \infty$		
	<i>MAE</i>	<i>MSE</i>	R^2	<i>MAE</i>	<i>MSE</i>	R^2	<i>MAE</i>	<i>MSE</i>	R^2	<i>MAE</i>	<i>MSE</i>	R^2
<i>Dynamic Factor Model</i>	0.007	0.000	0.002	0.062	0.004	0.001	0.079	0.006	0.011	0.102	0.010	0.047
<i>sGARCH</i>	0.780	0.609	0.000	0.781	0.610	0.000	0.864	0.747	0.001	0.804	0.647	0.006

Results show that the proposed method to track the common volatility factor in a panel of time series appears to be largely preferred with respect to the approach in (20). In particular, we note that for all the different specifications considered, that is, the data-generation process in (19) with $\nu_i = \{5, 10, 15, \infty\}$ for $i = 1, \dots, 100$, the common factor tracked with the recursion in equation (3) gives the best improvements in terms of all the measures of accuracy, the *MSE*, *MAE* and R^2 . It is interesting

to note that as the degrees of freedom becomes larger, that is the Student's t distribution approaches to the Gaussian distribution, the R^2 gets larger, while the MSE and the MAE tend to increase. The reason is that the estimated paths \hat{f}_t^2 from the simulated x_{it} will always be higher in levels than the true f_t^2 as it also includes the idiosyncratic noise variance.

6 Empirical applications

In this Subsection, we illustrate two applications of our dynamic factor model to the panel of financial returns from S&P100. The data are gathered from Yahoo! Finance¹ and the returns are calculated from closing prices. The observed financial time series span over a period from April 19, 1999 to April 17, 2019, leading to a total of $T = 5,033$ observations. In a first application we consider financial returns of ten selected companies, see Table 3. Then, we show that our parsimonious parametric dynamic factor model is also able to analyze the full panel of S&P100 financial returns.

6.1 Ten stocks from S&P100

We consider time series returns from Exxon Mobile Corporation (XOM), Apple Inc. (AAPL), Amazon.com Inc. (AMZN), International Business Machines (IBM), Microsoft Corporation (MSFT), A.O. Smith Corporation (AOS), Trust Financial Corporation (BBT), Carnival Corporation (CCL), Capital One Financial Corporation (COF) and Region Financial Corporation (RF). Descriptive statistics of the considered returns are detailed in Table 3. Except XOM, AMZN and BBT, the distribution of the returns are negatively skewed, while for all the series, these distributions are highly leptokurtic. In Figure 1 we display each of the series described in Table 3, where it is possible to graphically explore the fat-tailed nature of the considered time series. Also, by checking the data in the graphs, we note there is some common pattern in the volatilities, there is much idiosyncratic behaviour and there are jumps/outliers. Therefore, with this application, we aim to show the capability of our parsimonious parametric model in robustly capturing a great amount of cross-sectional dependence and idiosyncratic behaviour.

¹<https://finance.yahoo.com/recent-quotes>

Table 3: Descriptive statistics. Stock index returns are calculated from closing prices. The time series extend from April 19, 1999 to April 17, 2019 for a total of 5033 observations

Companies	Ticker acronym	Mean	Std. Dev.	Skewness	Exc. Kurtosis
Exxon Mobile Corporation	XOM	0.01	1.52	0.02	9.86
Apple Inc.	AAPL	0.10	2.72	-3.95	106.46
Amazon.com Inc.	AMZN	0.06	3.42	0.45	10.85
International Business Machines	IBM	0.01	1.68	-0.16	9.16
Microsoft Corporation	MSFT	0.02	1.93	-0.09	9.21
A.O. Smith Corporation	AOS	0.05	2.12	-0.07	5.75
Trust Financial Corporation	BBT	0.00	2.09	0.01	5.72
Carnival Corporation	CCL	0.00	2.28	-1.05	21.75
Capital One Financial Corporation	COF	0.01	3.01	-1.24	26.43
Region Financial Corporation	RF	-0.01	3.07	-0.54	41.62

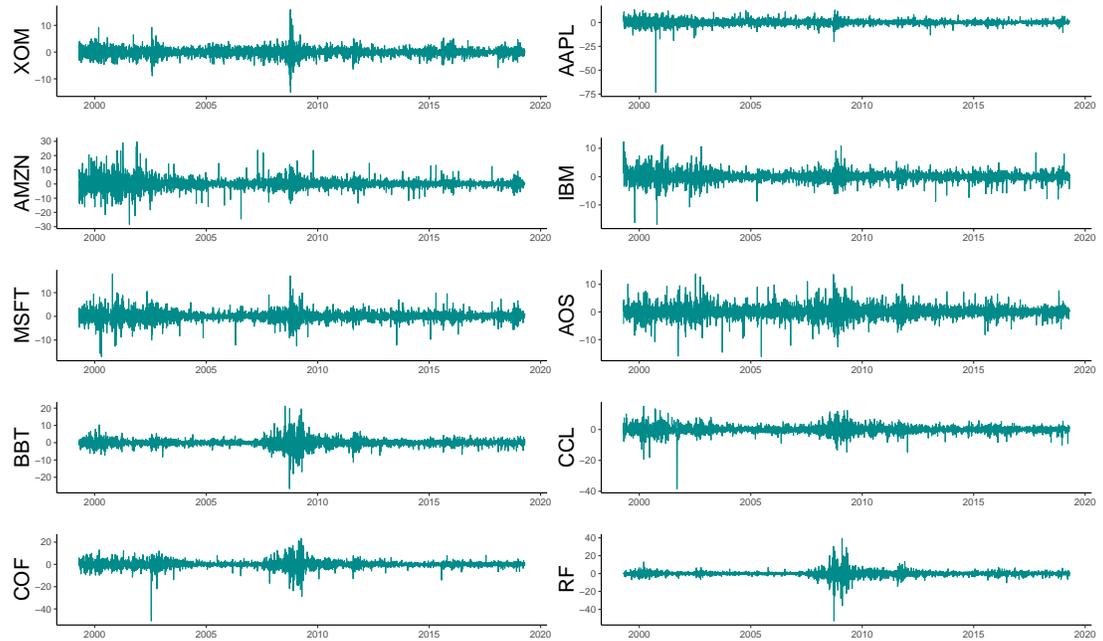


Figure 1: Panel of the ten financial returns from S&P100 separately. The time series extend from April 19, 1999 to April 17, 2019 giving a total of $T = 5033$ observations.

To show that correlations among of our panel of stock returns are changing over time, we split the full data sample in subsamples. The subsamples are constructed from each available year, that is, we obtain from the 1999 to 2019, 21 different subsamples of the panel of stock returns. After this, we

compute the cross-correlation matrices for each subsamples. Although the sample cross-correlation matrix is a consistent estimator, it can be seriously biased in finite sample. Therefore, to overcome this issue, we follow the suggestion of Tsay (2005) and perform a bootstrap resampling method. The correlations matrices of the ten stock returns of S&P100 are displayed in Figure 2 and it clearly unveils time variations in the dependence structure and clustering among stocks.

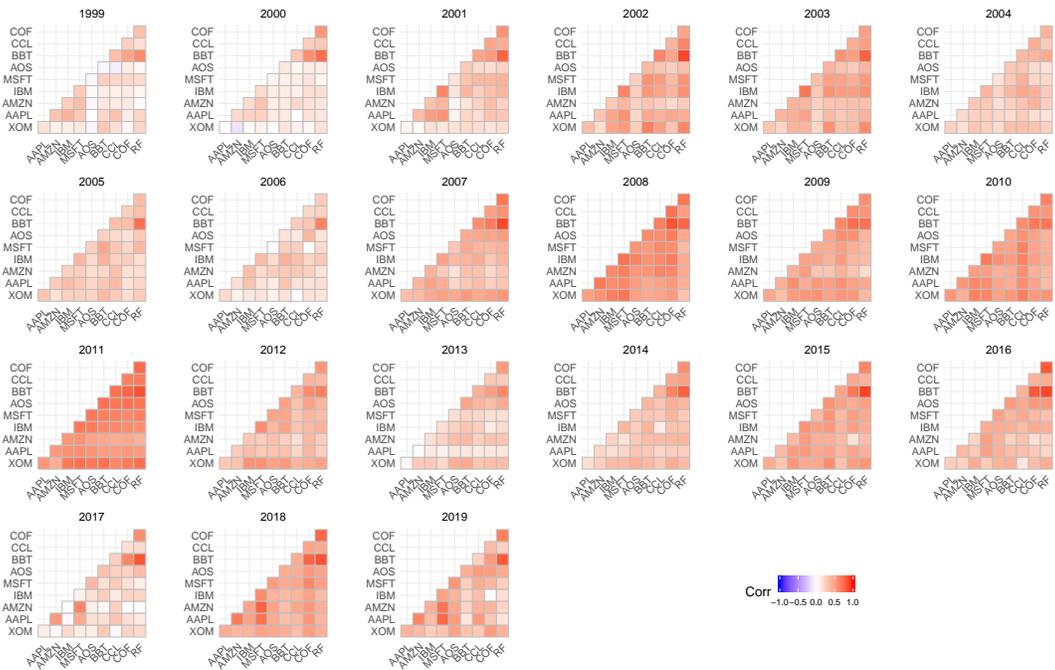


Figure 2: Yearly sample cross-correlation matrices of the daily returns of the ten financial returns from S&P100 estimated from April 19, 1999 to April 17, 2019.

In Table 4 we report the maximum likelihood estimates of the dynamic factor model described in Section 2. In order to show the robustness properties of the proposed model against jumps/outliers, we fit to the ten returns from S&P100 described in Table 3 the model specified by the equations in (4) with two different distributional assumptions, namely the Gaussian (\mathcal{N}) and the Student's $t(t_{\nu_i})$ with ν_i

degrees of freedom for $i = 1, \dots, 10$. The univariate model, i.e.

$$x_{it} = \sigma_{it}\epsilon_{it}$$

$$\sigma_{i,t+1}^2 = \delta_i + \phi_i\sigma_{it}^2 + \kappa_i\sigma_{it}^2 \left[\frac{(\nu_i + 1)(x_{it}^2/f_t^2)}{(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)} - 1 \right],$$

for $i = 1, \dots, 10$, provides estimation benchmark results. We note that the Gaussian version of the model in (4) is simply obtained by letting $\nu \rightarrow \infty$.

As concerns the estimates reported in Table 4, the autoregressive coefficients β and ϕ_i for $i = 1, \dots, 10$, confirm the presence of high persistence in both the common and idiosyncratic volatilities respectively. However, we note that while the estimated α are almost the same, the estimated κ_i for $i = 1, \dots, 10$ are systematically lower in the Gaussian specification. This difference between the two specification can be explained by the fact that when $\nu \rightarrow \infty$ the recursion of the idiosyncratic terms in (4) is not uniformly bounded and therefore, extreme observations can have an unbounded effect on next volatility values hence reducing the values of the coefficients κ_i . Thus, the robustness of the Student's t distribution provides a good alternative to reduce the downward or upward biases of the maximum likelihood estimators.

To better understand the efficacy of the proposed multivariate model and the adequacy of the proposed multivariate volatility model, in Table 5 we compare three likelihood based information criteria, namely the Akaike information criterion (AIC), the Bayes information criteria (BIC) and the Hannan-Quinn information criterion (HQC), and we perform a multivariate ARCH test to check for remaining auto-correlation in the squared residuals to our dynamic factor model, against the univariate models and two competing models belonging to the class of multivariate GARCH models, i.e. the *BEKK* and the dynamic conditional correlation (*DCC*), introduced by Engle and Kroner (1995) and Engle (2002) respectively.

First, from the obtained results it is evident that the introduction of a dynamic common factor during the estimation procedures significantly improve the in-sample fitting performance, in both the distribu-

tional assumptions, namely the Gaussian and Student's t . In particular, we note a drastic reduction of all the likelihood based information criteria, which provide clear evidence of the presence of common pattern in the volatilities. Second, by comparing the three likelihood based information criteria our dynamic factor model with Gaussian and Student's t distributed disturbances, results also reveal that our dynamic factor model give a better in-sample fit compared with both the *BEKK* and the *DCC* models. In addition, the diagnostic analysis suggests that our model is the only one that greatly reduce the amount of dependence in the ten series, as it is the only model where the multivariate ARCH statistic fails to reject the null hypothesis of no serial correlation.

To get deeper insights, Figure 3 displays the daily sample standard deviation of the considered ten returns from S&P100, against the common volatility factor $\hat{f}_t(\boldsymbol{\lambda})$, filtered with the two distributional assumptions for the disturbances, Gaussian and the Student's t . From that figure, one can appreciate the robustness property of common factor inherited from the prediction error $\frac{1}{N} \sum_{i=1}^N x_{it}^2 - \hat{f}_t^2(\boldsymbol{\lambda})$, for $i = 1, \dots, N$. In fact, while the daily sample standard deviation clearly overestimate the common volatility shared among the panel of stock returns, the dynamic common factor seems to well capture the substantial amount of dependence across the panel. Moreover, in Figure 4 we also report the filtered idiosyncratic terms, comparing the estimated volatilities for the Gaussian and Student's t specifications. The volatility estimates from the two models are substantially different for this panel of time series. We see that extreme observations have a strong effect on filtered volatilities for the Gaussian specification, whereas the Student's t based recursions provide a more realistic representation of the idiosyncratic volatilities. We conclude that our dynamic factor model with robust idiosyncratic components provide a satisfactory alternative to model large panels of time series that show strong factor structures and fat-tailed behaviours.

Table 4: Ten stocks of the S&P100 estimated parameters and standard errors (in parenthesis) of the marginals and dynamic factor models based on the Gaussian (\mathcal{N}) and the Student's t (t_{ν_i}) distributed disturbances. Estimation period is April 19, 1999 - April 17, 2019.

	Dynamic Factor Model											Univariate Model				
	\mathcal{N}					t_{ν_i}						\mathcal{N}		t_{ν_i}		
	ω	α	β	ϕ_i	κ_i	ω	α	β	ϕ_i	κ_i	ν_i	ϕ_i	κ_i	ϕ_i	κ_i	ν_i
$f_T^2(\lambda)$	0.053 (0.166)	0.101 (0.023)	0.974 (0.008)			0.054 (0.154)	0.102 (0.012)	0.974 (0.005)								
XOM				0.997 (0.002)	0.040 (0.031)				0.996 (0.000)	0.063 (0.008)	5.464 (0.422)	0.988 (0.001)	0.063 (0.041)	0.991 (0.001)	0.085 (0.011)	7.851 (0.485)
APPL				0.998 (0.003)	0.013 (0.016)				0.996 (0.001)	0.038 (0.008)	4.959 (0.134)	0.999 (0.002)	0.037 (0.027)	0.999 (0.001)	0.075 (0.010)	4.843 (0.253)
AMZN				0.999 (0.001)	0.008 (0.021)				0.998 (0.000)	0.037 (0.010)	4.614 (0.143)	0.989 (0.002)	0.019 (0.014)	0.999 (0.001)	0.092 (0.014)	4.059 (0.244)
IBM				0.999 (0.001)	0.009 (0.033)				0.996 (0.000)	0.058 (0.007)	4.820 (0.206)	0.997 (0.001)	0.035 (0.025)	0.989 (0.000)	0.097 (0.015)	4.811 (0.267)
MSFT				0.999 (0.001)	0.007 (0.015)				0.990 (0.001)	0.066 (0.010)	4.856 (0.223)	0.994 (0.002)	0.030 (0.020)	0.996 (0.002)	0.092 (0.010)	4.624 (0.224)
AOS				0.688 (0.010)	0.094 (0.019)				0.902 (0.000)	0.091 (0.015)	4.788 (0.259)	0.999 (0.001)	0.020 (0.017)	0.999 (0.000)	0.072 (0.017)	4.405 (0.279)
BBT				0.999 (0.002)	0.022 (0.047)				0.998 (0.000)	0.040 (0.014)	5.244 (0.358)	0.998 (0.002)	0.054 (0.035)	0.995 (0.000)	0.081 (0.025)	6.389 (0.460)
CCL				0.824 (0.007)	0.050 (0.026)				0.932 (0.001)	0.059 (0.010)	5.002 (0.289)	0.989 (0.001)	0.032 (0.020)	0.999 (0.001)	0.072 (0.020)	4.890 (0.320)
COF				0.996 (0.002)	0.010 (0.032)				0.997 (0.000)	0.026 (0.014)	5.100 (0.254)	0.999 (0.001)	0.038 (0.024)	0.999 (0.000)	0.085 (0.024)	5.098 (0.312)
RF				0.999 (0.001)	0.027 (0.028)				0.998 (0.001)	0.049 (0.015)	5.284 (0.401)	0.998 (0.002)	0.063 (0.047)	0.999 (0.002)	0.095 (0.017)	6.470 (0.489)

Table 5: Information criteria and diagnostic tests for the residuals of the competitor models based on the Gaussian (\mathcal{N}) and the Student's t (t_{ν_i}) distributed disturbances. ARCH LM refers to the p-values of the test for remaining auto-correlation in the squared residuals.

	AIC		BIC		HQC		ARCH LM	
	\mathcal{N}	t_{ν_i}	\mathcal{N}	t_{ν_i}	\mathcal{N}	t_{ν_i}	\mathcal{N}	t_{ν_i}
<i>Univariate Model</i>	205423.6	198352.9	205227.1	198548.7	205611.3	198421.5	0.001	0.002
<i>Dynamic Factor Model</i>	134899.5	128979.4	135114.8	129194.9	134975.2	129055.1	0.017	0.026
<i>BEKK</i>	190378.0	189405.6	190945.6	190045.0	190576.9	189629.6	0.001	0.001
<i>DCC</i>	188853.9	180219.5	189421.5	180858.8	189052.8	180443.1	0.001	0.001

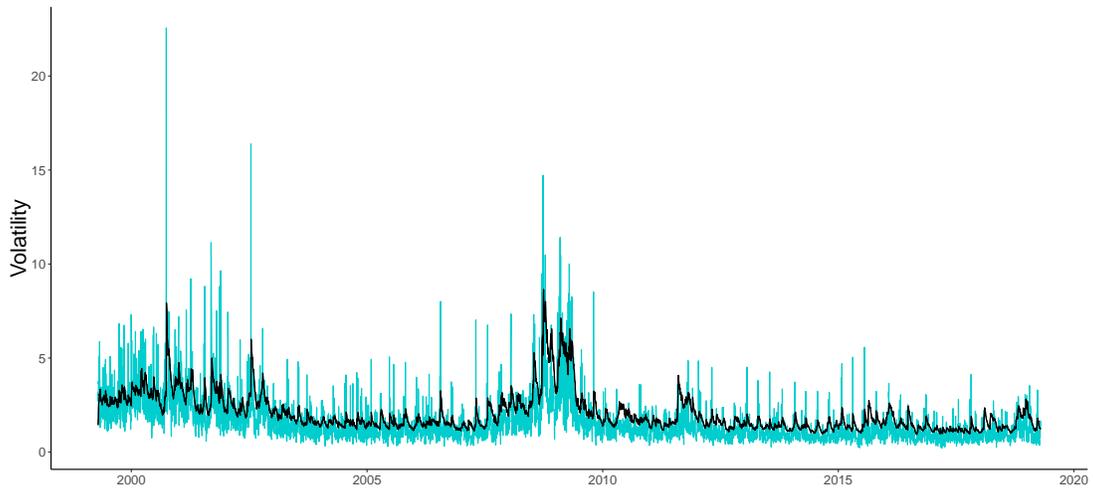


Figure 3: In solid cyan it is reported the daily sample standard deviation, computed from the ten stock returns from S&P100, and in solid thick black, the filtered common volatility $f_t(\lambda)$ with both the Gaussian and Student's t dynamic factor models.

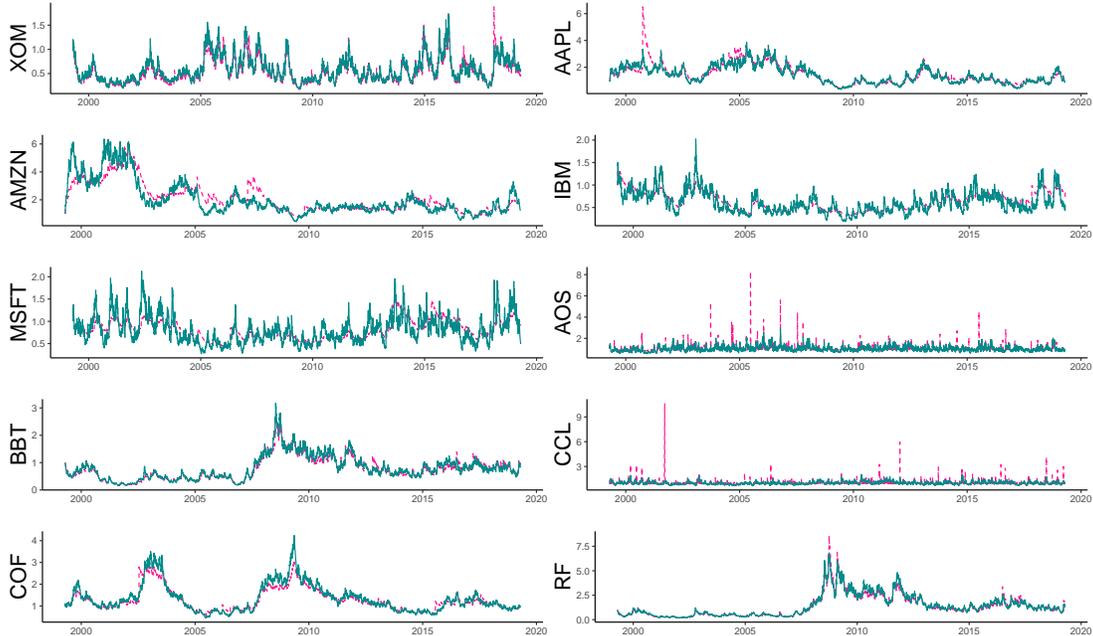


Figure 4: Filtered idiosyncratic volatilities $\sigma_{it}(\psi_i)$ of the ten stock returns from S&P100. In dashed pink and solid cyan we display those filtered with the Gaussian and the Student's t specifications respectively.

6.2 All stocks from S&P100

In the previous empirical application we shown that our methods can work with real data, and was able to robustly retrieve both the common and the idiosyncratic conditional volatility patterns. However, with this second exercise, we aim to show that our parsimonious model can work also with all stock returns from the S&P100. Since the dynamic factor model with Student's t appeared to be the most suitable choice to model the stock returns from the S&P100, we provide estimation details only for this specification. The MLE and their relative standard errors (in parenthesis) are reported in Table 6. To improve readability, we list the median and the first and third quartile of all the coefficient estimates and their standard errors, labelled as $q_{0.50}$, $q_{0.25}$ and $q_{0.75}$ respectively.

Table 6: S&P100 estimated parameters and standard errors (in parenthesis) of the dynamic factor models based on the Student's t (t_{ν_i}) distributed disturbances. Estimation period is April 19, 1999 - April 17, 2019.

<i>Dynamic Factor Model</i>						
t_{ν_i}						
	ω	α	β	ϕ_i	κ_i	ν_i
$f_t^2(\boldsymbol{\lambda})$	0.000 (0.054)	0.169 (0.035)	0.997 (0.007)			
$q_{0.25}$				0.995 (0.000)	0.030 (0.012)	4.999 (0.356)
$q_{0.50}$				0.997 (0.002)	0.039 (0.055)	5.001 (0.458)
$q_{0.75}$				0.998 (0.004)	0.058 (0.078)	5.020 (0.499)

Figure 5 shows the filtered common conditional volatility and the idiosyncratic components, which provides a further visual evidence of the substantial difference between the two source of price fluctuations. In particular, the presence of strong common conditional volatility seems obvious from the graph and moreover, the idiosyncratic volatilities are stationary and vary around the unconditional means, that is equal to one, see Assumption 2.

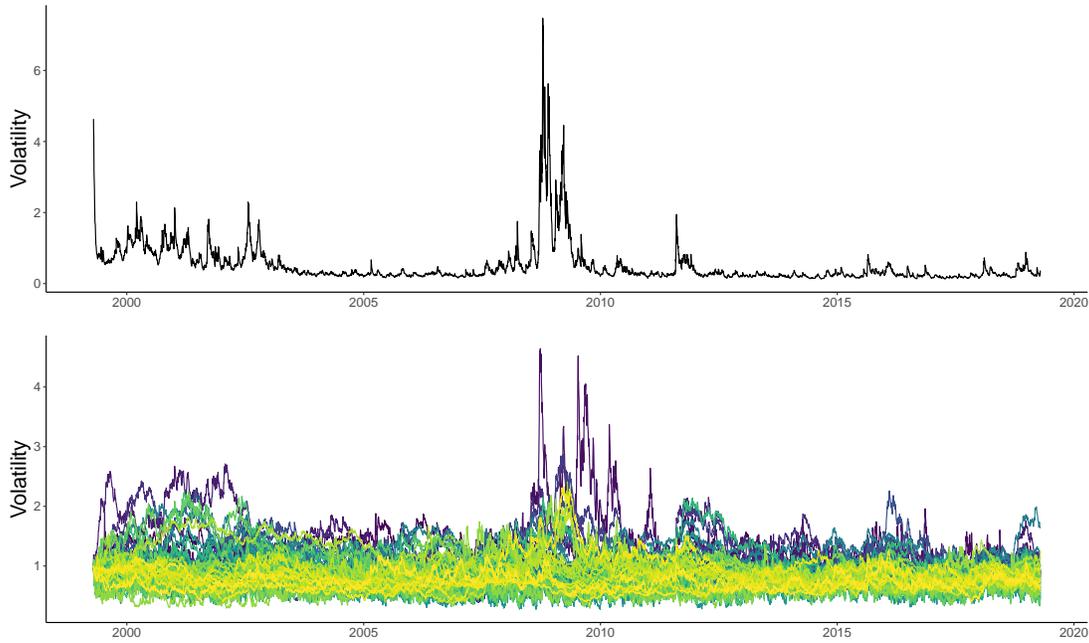


Figure 5: Filtered common volatility $f_t(\lambda)$ (upper panel), and filtered idiosyncratic volatilities $\sigma_{it}(\psi_i)$ (lower panel) of stock returns from S&P100. The time series extend from April 19, 1999 to April 17, 2019 giving a total of $T = 5033$ observations.

In analogy with the Monte Carlo experiment presented in Subsection 5.2, we also filter the common volatility with the procedure as implied by equation (20), that is the standard GARCH (*sGARCH*) model applied to the cross-sectional average. The two filtered paths can be visually compared as in Figure 6, where it is evident the inadequacy of the *sGARCH* in capturing most of the common variance in price fluctuations. The deterioration of the filtered common volatility is clearly visible. While our model suggests that the increase in volatility in the stocks in S&P100 can be attributed mostly to a common component, the *sGARCH* struggle to catch this empirical feature in the data. These empirical results further emphasize the efficiency of our multivariate, yet simple approach.

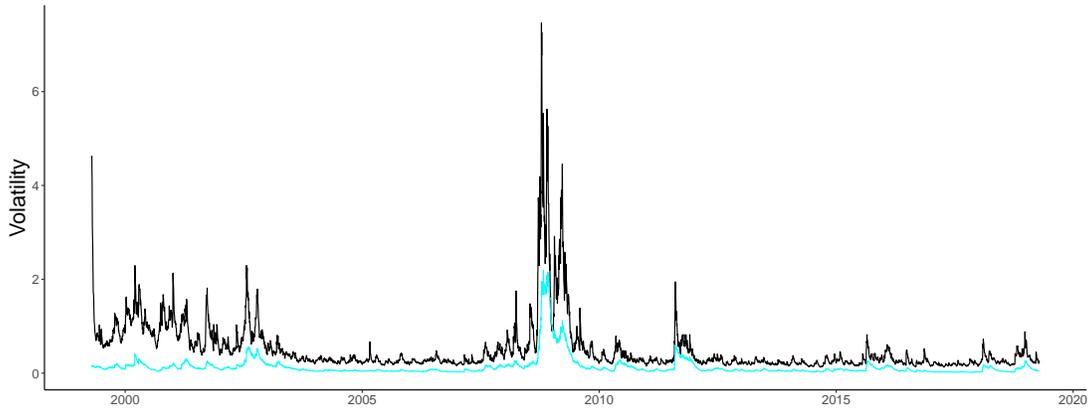


Figure 6: Filtered common volatilities with the dynamic factor model in (4) (black), and *sGARCH* in (20) (cyan) of stock returns from S&P100. The time series extend from April 19, 1999 to April 17, 2019 giving a total of $T = 5033$ observations.

7 Conclusion

We have introduced a new dynamic conditional variance factor model which features a basic updating equation for the common factor and a robust score-driven updating equation for the idiosyncratic factors. 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 (MLE) in large samples. A Monte Carlo study has shown that the MLE has good finite sample properties and the reliability of the proposed method to track the common volatility pattern from a large panel of time series. Finally, two empirical illustrations have shown for a panel of ten selected stocks from the S&P100 and a panel of all 100 stocks that our parsimonious dynamic multiplicative factor structure have various advantages over other competing models. The empirical results have shown the key importance of robust filtering methods to facilitate the idiosyncratic factors in financial data sets.

A Proofs

A.1 Proof of Proposition 1

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

$$b_{it} = \frac{\epsilon_{it}^2}{\nu_i - 2 + \epsilon_{it}^2}, \tag{21}$$

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,

$$\sigma_{i,t+1}^2 = g(\epsilon_{it}) + g(\epsilon_{it}) \sum_{k=1}^{\infty} \prod_{i=0}^k c(\epsilon_{it-1}). \tag{22}$$

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

$$f_{t+1}^2 = g(\epsilon_{it}) + c(\epsilon_{it})f_t^2, \quad (23)$$

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 $i = 1, \dots, N$, the second equality follows. Again, the inequality is ensured by assumptions and the implied almost sure representation is the analogous version of (22).

In conclusion, Assumptions 1 and 2, ensure that the sequence composed by $\{(f_t^2, (\boldsymbol{\sigma}_t^2)^\top, \boldsymbol{\epsilon}_t^\top)\}$ converges to the unique stationary ergodic solution $\{(f_t^2, (\boldsymbol{\sigma}_t^2)^\top, \boldsymbol{\epsilon}_t^\top)\}_{t \in \mathbb{Z}}$. Thus, the same old true for $\{f_t^2 \sigma_{it}^2\}_{t \in \mathbb{Z}}$ and $\{f_t \sigma_{it} \epsilon_{it}\}_{t \in \mathbb{Z}}$ for $i = 1, \dots, N$ by Proposition 3.36 of White (2001), implying that the series $\{x_{it}\}_{t \in \mathbb{Z}}$ is second-order stationary. \square

A.2 Proof of Proposition 2

Proof. By similar arguments as those in the Proof of Proposition 1, if $\mathbb{E}[\epsilon_{it}^j] < \infty$, we obtain necessary and sufficient conditions for the existence of the moments of the *Beta-t-GARCH* and the common factor process, that is

$$\mathbb{E}[c(\epsilon_{it})^{j/2}] < 1 \quad \text{for } j = 2, 4, \dots,$$

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 (22) in the Proof of Proposition 1 and recall that $\{\epsilon_{it}\}_{t \in \mathbb{Z}}$ forms an *IID* sequence of positive random

variables. Then, by virtue of the Minkowsky's inequality, we have

$$\begin{aligned}\mathbb{E}[\sigma_{i,t+1}^{2d}] &\leq g(\epsilon_{it})^d + g(\epsilon_{it})^d \mathbb{E} \left[\sum_{k=1}^{\infty} \prod_{i=0}^k c(\epsilon_{i,t-i})^d \right] \\ &\leq g(\epsilon_{it})^d + g(\epsilon_{it})^d \sum_{k=1}^{\infty} \left(\mathbb{E}[c(\epsilon_{i,t-i})^d] \right)^{k/d},\end{aligned}$$

thus, if $\mathbb{E}[c(\epsilon_{it})^{j/2}] < 1$ for $j = 2, 4, \dots$, 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 common factor process $\{f_{t+1}^2\}_{t \in \mathbb{Z}}$, since it admits the same representation, displayed in equations (23). 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 (6) 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

$$\|x_{it}\|_m = \|f_t \sigma_{it} \epsilon_{it}\|_m \leq \|f_t\|_{n_f} \|\sigma_{it}\|_{n_\sigma} \|\epsilon_{it}\|_{n_\epsilon},$$

where, for any scalar random variable x , we define the norm $\|x\|_n = (\mathbb{E}[|x|^n])^{1/n}$.

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 $i = 1, \dots, N$. In particular, this means that

$$f_t \in L^{n_f}, \quad \sigma_{it} \in L^{n_\sigma}, \quad \epsilon_{it} \in L^{n_\epsilon},$$

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

$$m = \frac{n_f n_\sigma n_\epsilon}{n_\sigma n_\epsilon + n_f n_\epsilon + n_f n_\sigma}.$$

□

A.3 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

$$\hat{f}_{t+1}^2 = \frac{\omega}{1 - (\beta - \alpha)} + \alpha \frac{1}{N} \sum_{i=1}^N \sum_{j=0}^{t-1} (\beta - \alpha)^j x_{i,t-j}^2 + (\beta - \alpha)^{t-1} \hat{f}_1^2, \quad (24)$$

where \hat{f}_1^2 is some initial value. It is clear, then, that the condition $|\beta - \alpha| < 1$ ensures that asymptotically, as $t \rightarrow \infty$, the impact of this initial value vanishes and hence the filter is invertible. In particular, \hat{f}_t converges to the limit

$$f_{t+1}^2 = \frac{\omega}{1 - (\beta - \alpha)} + \alpha \frac{1}{N} \sum_{i=1}^N \sum_{j=0}^{\infty} (\beta - \alpha)^j x_{i,t-j}^2. \quad (25)$$

We now consider the perturbed stochastic recurrence equation

$$\hat{\sigma}_{i,t+1}^2 = \hat{\Phi}_t(\hat{\sigma}_{it}^2),$$

which is a perturbed version of the filtered $\hat{\sigma}_{i,t+1}^2 = \Phi_t(\hat{\sigma}_{it}^2)$ since $\hat{\Phi}_t$ depends on the nonstationary filtered factor \hat{f}_t^2 whereas Φ_t depends on the limit stationary filtered factor f_t^2 . The unperturbed version takes the limit $\{f_t^2\}_{t \in \mathbb{Z}}$ and generate $\hat{\sigma}_{i,t+1}^2 = \Phi_t(\hat{\sigma}_{it}^2)$. We verify the condition of Theorem 2.10 in Straumann and Mikosch (2006) and proceed step-by-step.

First, the conditions of the Bougerol's Theorem must hold for the unperturbed recurrence equation. To start, we note that $\{x_{it}\}_{t \in \mathbb{Z}}$ and $\{f_t^2\}_{t \in \mathbb{Z}}$ are stationary and ergodic and the initial log-moment

condition is satisfied since

$$\begin{aligned} & \mathbb{E} \left[\log^+ \left| \delta_i \left[(\phi_i - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^2/f_t^2)}{(\nu_i - 2)\bar{\sigma}_i^2 + (x_{it}^2/f_t^2)} \right] \bar{\sigma}_i^2 \right| \right] \\ & \leq \log^+ |\delta_i| + \log^+ |\phi_i - \kappa_i| + \log^+ |\kappa_i| \\ & \quad + \mathbb{E} \left[\log^+ \left| \frac{(\nu_i + 1)(x_{it}^2/f_t^2)}{(\nu_i - 2)\bar{\sigma}_i^2 + (x_{it}^2/f_t^2)} \bar{\sigma}_i^2 \right| \right] + \log^+ |\bar{\sigma}_i^2|, \end{aligned}$$

and we can see that

$$\mathbb{E} \left[\log^+ \left| \frac{(\nu_i + 1)(x_{it}^2/f_t^2)}{(\nu_i - 2)\bar{\sigma}_i^2 + (x_{it}^2/f_t^2)} \bar{\sigma}_i^2 \right| \right] < \infty$$

since it is uniformly bounded in $\bar{\sigma}_i^2 \geq \delta_i = (1 - \phi_i) > 0$ implied by $|\phi_i| < 1$ from Assumptions 2 and 3, and uniformly bounded $\forall x_{it} \in \mathbb{R}$. We also require that the unperturbed recurrence equation satisfies the contraction condition

$$\mathbb{E} \left[\log \sup_{\sigma_i^2} \left\| \frac{\partial \Phi(\sigma_{it}^2)}{\partial \sigma_{it}^2} \right\| \right] < 0,$$

which is verified from the fact that

$$A_{it}(\psi_i) = \frac{\partial \Phi(\sigma_{it}^2)}{\partial \sigma_{it}^2} = (\phi_i - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^4/f_t^4)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2}, \quad (26)$$

and therefore

$$\sup_{\sigma_i^{2*}} \left\| (\phi_i - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^4/f_t^4)}{[(\nu_i - 2)\sigma_i^{2*} + (x_{it}^2/f_t^2)]^2} \right\| \leq \max_{i=1, \dots, N} \left\{ |\phi_i - \kappa_i|; |\phi_i + \kappa_i \nu_i| \right\} < 1,$$

because, as before $\sigma_i^{2*} \geq \delta_i = (1 - \phi_i) > 0$ by Assumption 3 and $f_t \geq \omega > 0$. Moreover, is uniformly bounded $\forall x_{it} \in \mathbb{R}$ and then, the result follows from Assumption 4. In practice, the contraction condition will ensure that

$$\sup_{\theta \in \Theta} \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \xrightarrow{\text{e.a.s.}} 0 \quad \text{as} \quad t \rightarrow \infty.$$

Second, we verify the logarithmic moment for the stationary solution $\{\sigma_{it}^2\}_{t \in \mathbb{Z}}$, which is clearly implied

by Proposition 2, in fact

$$\mathbb{E}[\sigma_{i,t+1}^j] = \mathbb{E}[\Phi(\sigma_{it}^2)^{j/2}] < \infty$$

for $j = 2, 4, \dots$

Third, the perturbed stochastic recurrence equation $\widehat{\widehat{\sigma}}_{i,t+1}^2 = \widehat{\Phi}_t(\widehat{\widehat{\sigma}}_{it}^2)$, must converges to the unperturbed counterpart, that is verified as an application of the mean value theorem. We have

$$\sup_{\theta \in \Theta} \|\widehat{\Phi}_t(\widehat{\widehat{\sigma}}_{it}^2) - \Phi_t(\widehat{\sigma}_{it}^2)\| \leq \sup_{\theta \in \Theta} \left\| \frac{\partial \Phi_t^*(\widehat{\sigma}_{it}^2)}{\partial f_t^{2*}} \right\| \sup_{\theta \in \Theta} \|\widehat{f}_t^2 - f_t^2\|,$$

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 \widehat{b}_t , that is the nonstationary counterpart of

$$b_t = \frac{(x_{it}^2 / f_t^2)}{(\nu_i - 2) \sigma_{it}^2 + (x_{it}^2 / f_t^2)}, \quad (27)$$

which takes $\widehat{\sigma}_{it}^2$ and \widehat{f}_t^2 instead of σ_{it}^2 and f_t^2 . It can be check that \widehat{b}_t is bounded between 0 and 1. Then,

we can write

$$\begin{aligned} \sup_{\theta \in \Theta} \left\| \frac{\partial \Phi_t^*(\widehat{\sigma}_{it}^2)}{\partial f_t^{2*}} \right\| &\leq \sup_{\theta \in \Theta} \sup_f \left\| \frac{1}{f_t^{2*}} \right\| \sup_{\theta \in \Theta} \|\kappa_i \widehat{\sigma}_{it}^2 (\nu_i + 1) (\widehat{b}_{it}^2 - \widehat{b}_{it})\| \\ &\leq \sup_{\theta \in \Theta} \left\| \frac{1}{\omega} \right\| \sup_{\theta \in \Theta} \|\kappa_i \widehat{\sigma}_{it}^2 (\nu_i + 1) (\widehat{b}_{it}^2 - \widehat{b}_{it})\| < \infty, \end{aligned}$$

where the last inequality holds because $\omega > 0$, as imposed by Assumption 1, and $\widehat{\sigma}_{it}^2$ is uniformly bounded.

We obtain

$$\sup_{\theta \in \Theta} \|\widehat{\Phi}_t(\widehat{\widehat{\sigma}}_{it}^2) - \Phi_t(\widehat{\sigma}_{it}^2)\| \xrightarrow{\text{c.a.s.}} 0 \quad \text{as} \quad t \rightarrow \infty.$$

as a straightforward application of Lemma 2.1 in Straumann and Mikosch (2006), since as $t \rightarrow \infty$ 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 constant of the unperturbed map, that can be verified again by the mean value theorem. We have

$$\sup_{\theta \in \Theta} \|\hat{\Phi}_t(\hat{\sigma}_{it}^2) - \Phi_t(\hat{\sigma}_{it}^2)\| \leq \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \Phi_t^*(\sigma_{it}^2)}{\partial \sigma_{it}^{2*} \partial f_t^{2*}} \right\| \sup_{\theta \in \Theta} \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\|,$$

where

$$\begin{aligned} \frac{\partial^2 \Phi_t(\sigma_{it}^2)}{\partial \sigma_{it}^2 \partial f_t^2} &= \frac{\kappa_i(\nu_i + 1)(x_{it}^4/f_t^6)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2} - \frac{(2[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)](\nu_i - 2))\kappa_i(\nu_i + 1)(x_{it}^4/f_t^6)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]^2} \\ &\quad - \frac{\kappa_i(\nu_i + 1)(x_{it}^2/f_t^4)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]} + \frac{\kappa_i\sigma_{it}^2(\nu_i - 2)(\nu_i + 1)(x_{it}^2/f_t^4)}{[(\nu_i - 2)\sigma_{it}^2 + (x_{it}^2/f_t^2)]} \end{aligned}$$

Therefore, by the same argument as before, we can rewrite the above equation in terms of the random variables b_t , in fact

$$\begin{aligned} \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \Phi_t^*(\sigma_{it}^2)}{\partial \sigma_{it}^{2*} \partial f_t^{2*}} \right\| &\leq \sup_{\theta \in \Theta} \sup_f \left\| \frac{\kappa_i(\nu_i + 1)\hat{b}_t}{f_t^{2*}} \right\| \\ &\quad \times \left[1 + \|\hat{b}_t\| + \left(\|2[(\nu_i - 2)\sigma_{it}^{2*} + (x_{it}^2/f_t^{2*})](\nu_i - 2)\| \right) + \|\sigma_{it}^{2*}(\nu_i - 2)\| \right] \end{aligned}$$

so that, it is possible to verify that

$$\sup_{\theta \in \Theta} \left\| \frac{\partial^2 \Phi_t^*(\sigma_{it}^2)}{\partial \sigma_{it}^{2*} \partial f_t^{2*}} \right\| < \infty,$$

and since $\sup_{\theta \in \Theta} \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \xrightarrow{\text{e.a.s.}} 0$ as $t \rightarrow \infty$ 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 (4), it remains to show that the filtered $\{\hat{f}_t^2 \hat{\sigma}_{it}^2\}_{t \in \mathbb{N}}$ is still invertible and hence will converge to the unique stationary and ergodic solution $\{f_t^2 \sigma_{it}^2\}_{t \in \mathbb{N}}$ for $i = 1, \dots, N$ and for any fixed starting points \hat{f}_1^2 and $\hat{\sigma}_1^2$.

We easily achieve this result by using the following elementary decomposition

$$\begin{aligned} \|(\hat{f}_t^2 \hat{\sigma}_{it}^2) - (f_t^2 \sigma_{it}^2)\| &= \|(\hat{f}_t^2 - f_t^2) \sigma_{it}^2 + f_t^2 (\hat{\sigma}_{it}^2 - \sigma_{it}^2) + (\hat{f}_t^2 - f_t^2) (\hat{\sigma}_{it}^2 - \sigma_{it}^2)\| \\ &\leq \|\hat{f}_t^2 - f_t^2\| \|\sigma_{it}^2\| + \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\| \|f_t^2\| + \|\hat{f}_t^2 - f_t^2\| \|\hat{\sigma}_{it}^2 - \sigma_{it}^2\|, \end{aligned} \quad (28)$$

for $i = 1, \dots, N$ 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 $\mathbb{E}[\sup_{\theta \in \Theta} \|\sigma_{it}^2\|] < \infty$ for $i = 1, \dots, N$. 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

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

which completes the proof. \square

A.4 Proof of Lemma 4.1

Proof. We begin our proof with the first uniform convergence in (13), where $1 \leq N < \infty$ is fixed. The first step is to show that the ratio $x_{it}^2 / [(\nu_i - 2)(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))]$ has at least two bounded moments. As a matter of fact, this condition will be easily satisfied as soon as its denominator is bounded away from zero. To this end, we first note that by definition and assumption 1 $f_t^2(\boldsymbol{\lambda}) \geq \omega > 0$. Moreover, by Assumption 3 $\sigma_{it}^2(\boldsymbol{\psi}_i) \geq \delta = (1 - \phi_i) > 0$ for $i = 1, \dots, N$. Therefore, the existence of the first two moments of $x_{it}^2 / [(\nu_i - 2)(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))]$ is easily satisfied from the distributional assumptions, since the necessary condition $\min_{i=1, \dots, N} \nu_i > 2$ is always ensured by the definition of the model.

Now we verify that $\mathbb{E}[|\log(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))|] < \infty$ and then we have

$$\mathbb{E}[|\log(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))|] \leq \mathbb{E}[|\log f_t^2(\boldsymbol{\lambda})|] + \mathbb{E}[|\log \sigma_{it}^2(\boldsymbol{\psi}_i)|].$$

It is possible to show that $\mathbb{E}[|\log f_t^2(\boldsymbol{\lambda})|] < \infty$, since under Assumption 4, Proposition 2 entails

$\mathbb{E}[|\log f_t^2(\boldsymbol{\lambda})|] < \infty$. We note that, $\mathbb{E}[\log f_t^{2\varepsilon}(\boldsymbol{\lambda})] < \infty$ for some $\varepsilon > 0$ due to the compactness of Θ . Additionally, $f_t^2(\boldsymbol{\lambda})$ is bounded away from zero almost surely because $\omega > 0$ and hence $\mathbb{E}[\log f_t^2(\boldsymbol{\lambda})] < \infty$ which implies $\mathbb{E}[|\log f_t^2(\boldsymbol{\lambda})|] < \infty$.

Using similar arguments it is possible to show that $\mathbb{E}[|\log \sigma_{it}^2(\boldsymbol{\psi}_i)|] < \infty$ since under Assumption 2 $\min_{i=1, \dots, N} \nu_i > 2$ and $\phi < 1$. Note also that the multiplicative recursion $\{(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}$ is bounded away from zero for all $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}_i$ with $i = 1, \dots, N$ almost surely and therefore we obtain that $\mathbb{E}[\log(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))] < \infty$ which further implies $\mathbb{E}[|\log(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))|] < \infty$.

The first result in (13) is then obtained by appealing 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 $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} |\ell_{it}(\boldsymbol{\theta})|] < \infty$, which is clearly satisfied by virtue of the discussion above and standard continuity arguments. We obtain this result as an application of Theorem 3.5.8 in Stout (1974) on the likelihood function over the stationary and ergodic sequence $\{(f_t^2, (\boldsymbol{\sigma}_t^2)^\top, \boldsymbol{\epsilon}_t^\top)\}_{t \in \mathbb{Z}}$.

□

A.5 Proof of Lemma 4.2

Proof. Consider the conditional density (8) and the likelihood (9). 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

$$\begin{aligned} 0 &= \ell_{it}(\boldsymbol{\theta}) - \ell_{it}(\boldsymbol{\theta}_0) \\ &= \frac{1}{2} \log \frac{(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))}{(f_t^2(\boldsymbol{\lambda}_0)\sigma_{it}^2(\boldsymbol{\psi}_{i0}))} \\ &\quad - \frac{(\nu_{i0} + 1)}{2} \log \left[\left(1 + \frac{x_{it}^2}{(\nu_{i0} - 2)(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))}\right) / \left(1 + \frac{x_{it}^2}{(\nu_{i0} - 2)(f_t^2(\boldsymbol{\lambda}_0)\sigma_{it}^2(\boldsymbol{\psi}_{i0}))}\right) \right], \end{aligned}$$

where the equation holds if and only if $(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)) = (f_t^2(\boldsymbol{\lambda}_0)\sigma_{it}^2(\boldsymbol{\psi}_{i0}))$ for all t and $i = 1, \dots, N$, 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 $\{(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{Z}}$ and $\{(f_t^2(\boldsymbol{\lambda}_0)\sigma_{it}^2(\boldsymbol{\psi}_{i0}))\}_{t \in \mathbb{Z}}$ 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

$$\eta_{it} = \left(\frac{1}{N} \sum_{i=1}^N \sigma_{it}^2 \epsilon_{it}^2 - 1 \right) \quad u_{it} = \left(\frac{(\nu_{i0} + 1) \epsilon_{it}^2}{(\nu_{i0} - 2) + \epsilon_{it}^2} - 1 \right).$$

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

$$\begin{aligned} & (f_{t+1}^2(\boldsymbol{\lambda}) \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)) - (f_{t+1}^2(\boldsymbol{\lambda}_0) \sigma_{i,t+1}^2(\boldsymbol{\psi}_{i0})) \\ &= (\omega - \omega_0) - (\omega \phi_i - \omega_0 \phi_{i0}) + \xi_{it} f_t^2(\boldsymbol{\lambda}_0) + \zeta_{it} \sigma_{it}^2(\boldsymbol{\psi}_{i0}) + \varsigma_{it} (f_t^2(\boldsymbol{\lambda}_0) \sigma_{it}^2(\boldsymbol{\psi}_{i0})), \end{aligned} \quad (29)$$

where

$$\varsigma_{it} = (\beta \phi_i - \beta_0 \phi_{i0}) + \{(\alpha \kappa_i - \alpha_0 \kappa_{i0})\} \eta_{it} u_{it} + \{(\alpha \phi_i - \alpha_0 \phi_{i0})\} \eta_{it} + \{(\beta \kappa_i - \beta_0 \kappa_{i0})\} u_{it}, \quad (30)$$

$$\xi_{it} = (\beta - \beta_0) - (\beta \phi_i - \beta_0 \phi_{i0}) + \{(\alpha - \alpha_0) - (\alpha \phi_i - \alpha_0 \phi_{i0})\} \eta_{it}, \quad (31)$$

$$\zeta_{it} = (\omega \phi_i - \omega_0 \phi_{i0}) + \{(\omega \kappa_i - \omega_0 \kappa_{i0})\} u_{it}. \quad (32)$$

The discussion may start by considering the terms $(\omega - \omega_0)$ and $(\omega \phi_i - \omega_0 \phi_{i0})$. It is clear that every random variable involved in equations (29) 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 $(f_{t+1}^2(\boldsymbol{\lambda}) \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)) - (f_{t+1}^2(\boldsymbol{\lambda}_0) \sigma_{i,t+1}^2(\boldsymbol{\psi}_{i0})) = 0$ we should have the equivalence

$$(\omega - \omega_0) - (\omega \phi_i - \omega_0 \phi_{i0}) = \xi_{it} f_t^2(\boldsymbol{\lambda}_0) + \zeta_{it} \sigma_{it}^2(\boldsymbol{\psi}_{i0}) + \varsigma_{it} (f_t^2(\boldsymbol{\lambda}_0) \sigma_{it}^2(\boldsymbol{\psi}_{i0})),$$

but clearly, under the conditions $\alpha > 0$ and $\kappa_i > 0$ for $i = 1, \dots, N$, this is not possible, since the distributions of $f_{t+1}^2(\boldsymbol{\lambda}_0)$, $\sigma_{i,t+1}^2(\boldsymbol{\psi}_{i0})$ and then $(f_{t+1}^2(\boldsymbol{\lambda}_0) \sigma_{i,t+1}^2(\boldsymbol{\psi}_{i0}))$ are ensured to be nondegenerate almost surely for all t . As a consequence, one must have

$$\omega = \omega_0 \quad \phi_i = \phi_{i0}.$$

Now consider the equation (32). 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 (31) give us

$$(\beta - \beta_0) - (\beta\phi_i - \beta_0\phi_{i0}) = \{(\alpha - \alpha_0) - (\alpha\phi_i - \alpha_0\phi_{i0})\}\eta_{it}.$$

However, we note again that $\beta = \beta_0$ and $\alpha = \alpha_0$ because η_{it} is random. Then, the discussion extends trivially for (30). Therefore, the proof is complete. \square

A.6 Proof of Theorem 4.3

Proof. First, we split the likelihood function as follows

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_{NT}(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\mathcal{L}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\|. \quad (33)$$

Let us consider the difference of the first term in the right hand side. Applying the mean value expansion around $(\hat{f}_t^{2*}(\boldsymbol{\lambda})\hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))$ for $i = 1, \dots, N$ yields

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_{NT}(\boldsymbol{\theta})\| \\ & \leq \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))} \right\| \sup_{\boldsymbol{\theta} \in \Theta} \|(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)) - (f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\|. \end{aligned} \quad (34)$$

where $(f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))$ lies between $(\hat{f}_t^2(\boldsymbol{\lambda})\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))$ and $(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))$. The partial derivative can be expressed as

$$\frac{\partial \mathcal{L}_{NT}(\boldsymbol{\theta})}{\partial (f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{1}{(2f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))} [(\nu_i + 1)b_{it}^*(\boldsymbol{\theta}) - 1], \quad (35)$$

where $b_{it}^*(\boldsymbol{\theta})$ is defined as

$$b_{it}^*(\boldsymbol{\theta}) = \frac{x_{it}^2 / (f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))}{\nu_i - 2 + x_{it}^2 / (f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))},$$

and it can be easily noted that this is a random variable bounded in the interval $[0, 1)$.

Now, considering the fact that both $(f_t^{2*}(\boldsymbol{\lambda})\sigma_{it}^{2*}(\boldsymbol{\psi}_i))$ and $(\hat{f}_t^{2*}(\boldsymbol{\lambda})\hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))$ lie in $[c, +\infty)$ and

$[g, +\infty)$, where

$$c = \inf_{\boldsymbol{\theta} \in \Theta} \left(\frac{\omega}{1 - (\beta - \alpha)} \right), \quad g = \inf_{\boldsymbol{\theta} \in \Theta} \left(\frac{\delta_i}{1 - (\phi_i - \kappa_i)} \right),$$

such that $c > 0$ and also $g > 0$, we obtain that

$$\sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial \mathcal{L}_{NT}(\boldsymbol{\theta})}{\partial (f_t^{2*}(\boldsymbol{\lambda}) \sigma_{it}^{2*}(\boldsymbol{\psi}_i))} \right\| \leq \frac{1}{2cg} \left(\frac{\max_{\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N \in \Psi} (\nu_i + 1)}{cg} + 1 \right) = \bar{M},$$

and therefore

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_{NT}(\boldsymbol{\theta})\| \\ \leq \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^{\infty} \bar{M} \times \sup_{\boldsymbol{\theta} \in \Theta} \|(f_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)) - (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))\|. \end{aligned} \quad (36)$$

Clearly, the first term on the right hand side is finite and $\sup_{\boldsymbol{\theta} \in \Theta} \|(f_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)) - (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))\| \xrightarrow{\text{e.a.s.}}$

0 by virtue of Proposition 3. Hence, we satisfy the conditions of Lemma 2.1 in Straumann and Mikosch

(2006) and we obtain $\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_{NT}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0$, as $T \rightarrow \infty$.

Moreover, we can show that $\sup_{\boldsymbol{\theta} \in \Theta} \|\mathcal{L}_{NT}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0$, as an application of Lemma 4.1.

In Lemma 4.2 we have established the identifiability of the unique maximizer $\boldsymbol{\theta}_0 \in \Theta$ where, we note that Θ is compact and the continuity of $\mathcal{L}_{NT}(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$, $\forall t \in \mathbb{N}$ implies that the limit $\mathcal{L}_N(\boldsymbol{\theta})$ is also a continuous function in $\boldsymbol{\theta}$. It follows that $\boldsymbol{\theta}_0 \in \Theta$ is also unique, see White (1994). The strong consistency thus follows for N fixed and $T \rightarrow \infty$. \square

A.7 Proof of Proposition 4

Proof. First, note that

$$\nabla (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)) = (\nabla f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)) + f_t^2(\boldsymbol{\lambda}) (\nabla \sigma_{it}^2(\boldsymbol{\psi}_i)), \quad (37)$$

with $\nabla f_t^2(\boldsymbol{\lambda})$ composed by

$$\begin{aligned}\frac{\partial f_{t+1}^2(\boldsymbol{\lambda})}{\partial \omega} &= (\beta - \alpha) \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial \omega} + 1, \\ \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial \alpha} &= (\beta - \alpha) \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial \alpha} + \frac{1}{N} \sum_{i=1}^N x_{it}^2 - f_t^2(\boldsymbol{\lambda}), \\ \frac{\partial f_{t+1}^2(\boldsymbol{\lambda})}{\partial \beta} &= (\beta - \alpha) \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial \beta} + f_t^2(\boldsymbol{\lambda}),\end{aligned}\tag{38}$$

while for $i = 1, \dots, N$, $\nabla \sigma_{it}^2(\boldsymbol{\psi}_i)$ is composed by

$$\begin{aligned}\frac{\partial \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial \phi_i} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial \sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial \phi_i} + \sigma_{it}^2(\boldsymbol{\psi}_i) - 1, \\ \frac{\partial \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial \kappa_i} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial \sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial \kappa_i} + \sigma_{it}^2(\boldsymbol{\psi}_i) u_{it}(\boldsymbol{\psi}_i), \\ \frac{\partial \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial \nu_i} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial \sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial \nu_i} + \kappa_i \sigma_{it}^2(\boldsymbol{\psi}_i) a_{it}(\boldsymbol{\psi}_i),\end{aligned}\tag{39}$$

where $A_{it}(\boldsymbol{\psi}_i)$ is as in (26) and

$$u_{it}(\boldsymbol{\psi}_i) = \left(\frac{(\nu_i + 1)(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))} - 1 \right),\tag{40}$$

$$a_{it}(\boldsymbol{\psi}_i) = \left(\frac{(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))} - \frac{\sigma_{it}^2(\boldsymbol{\psi}_i)(\nu_i + 1)(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^2} \right).\tag{41}$$

In the same vein of Proposition 3, we shall prove the e.a.s. convergence by verifying the conditions of Theorem 2.10 in Straumann and Mikosch (2006).

First, the conditions of the Bougerol's Theorem must hold for the unperturbed recurrence equations in (38) and (39).

As concern the recursions in (38), it is easy to see that

$$\begin{aligned}\mathbb{E} \left[\log^+ \left| \frac{\partial f_{t+1}^2(\boldsymbol{\lambda})}{\partial \omega} \right| \right] &\leq \log^+ |\beta - \alpha| + \log^+ \left| \frac{\partial \bar{f}^2}{\partial \omega} \right| < \infty, \\ \mathbb{E} \left[\log^+ \left| \frac{\partial f_{t+1}^2(\boldsymbol{\lambda})}{\partial \alpha} \right| \right] &\leq \log^+ |\beta - \alpha| + \log^+ \left| \frac{\partial \bar{f}^2}{\partial \alpha} \right| + \mathbb{E} \left[\log^+ \left| \frac{1}{N} \sum_{i=1}^N x_{it}^2 \right| \right] + \log^+ |\bar{f}^2| < \infty, \\ \mathbb{E} \left[\log^+ \left| \frac{\partial f_{t+1}^2(\boldsymbol{\lambda})}{\partial \beta} \right| \right] &\leq \log^+ |\beta - \alpha| + \log^+ \left| \frac{\partial \bar{f}^2}{\partial \beta} \right| + \mathbb{E} \left[\log^+ \left| \frac{\partial \bar{f}^2}{\partial \beta} \right| \right] + \log^+ |\bar{f}^2| < \infty,\end{aligned}$$

where the first and the third inequalities follow trivially, while for the second inequality we obtain the

log-moment from Assumption 2 since $\min_{i=1,\dots,N} \nu_i > 2$ implies $\mathbb{E}[x_{it}^2] < \infty$ for $i = 1, \dots, N$.

Moreover, for the recursions in (39) we first note that

$$\mathbb{E} \left[\log^+ |A_{it}(\boldsymbol{\psi}_i)| \right] = \mathbb{E} \left[\log^+ \left| (\phi_i - \kappa_i) + \kappa_i \frac{(\nu_i + 1)(x_{it}^4/f_t^4(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^2} \right| \right] < \infty,$$

since it is uniformly bounded in $\sigma_{it}^2 > \delta_i = (1 - \phi_i) > 0$ by Assumption 3, and uniformly bounded $\forall x_{it} \in \mathbb{R}$. Then, we need to show that equations (40) and (41), have finite log-moments. However, by similar arguments as in the proof of Proposition 3, it is straightforward to show that all the three random variables are uniformly bounded, and therefore

$$\mathbb{E}[\log^+ |u_{it}(\boldsymbol{\psi}_i)|] < \infty, \quad \text{and} \quad \mathbb{E}[\log^+ |a_{it}(\boldsymbol{\psi}_i)|] < \infty.$$

It follows that all the recursions in (39) have finite log-moments. In addition, we also note that the contraction conditions stated in Proposition 3 hold for the unperturbed recurrence equations, and thus we fulfil the requirements of Bougerol's Theorem for the unperturbed recurrence equations.

Second, we need to show that the perturbed derivatives in $\nabla \hat{f}_t^2(\boldsymbol{\lambda})$ and $\nabla \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)$ converge to the unperturbed counterparts. In that case, the aforementioned perturbed recurrence equations depend on the nonstationary sequences $\{\hat{f}_t^2(\boldsymbol{\lambda})\}_{t \in \mathbb{N}}$ and $\{\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{N}}$ respectively.

As regards the derivatives in $\nabla f_t^2(\boldsymbol{\lambda})$, it is straightforward to see that the contraction condition $|\beta - \alpha| < 1$ stated in Assumption 4 will ensure that the perturbed derivatives converge e.a.s. to the unperturbed counterparts, that is

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \hat{f}_t^2(\boldsymbol{\lambda}) - \nabla f_t^2(\boldsymbol{\lambda})\| \xrightarrow{\text{e.a.s.}} 0, \quad \text{as} \quad t \rightarrow \infty. \quad (42)$$

On the other hand, for $\nabla \sigma_{it}^2(\boldsymbol{\psi}_i)$, we obtain the desired convergence by showing that

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{A}_{it}(\boldsymbol{\psi}_i) - A_{it}(\boldsymbol{\psi}_i)\| \xrightarrow{\text{e.a.s.}} 0 \quad \text{as} \quad t \rightarrow \infty, \quad (43)$$

and

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{u}_{it}(\boldsymbol{\psi}_i) - u_{it}(\boldsymbol{\psi}_i)\| \xrightarrow{\text{e.a.s.}} 0, \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta} \|\hat{a}_{it}(\boldsymbol{\psi}_i) - a_{it}(\boldsymbol{\psi}_i)\| \xrightarrow{\text{e.a.s.}} 0. \quad (44)$$

An application of the mean value theorem yields

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{A}_{it}(\boldsymbol{\psi}_i) - A_{it}(\boldsymbol{\psi}_i)\| \leq \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial \hat{A}_{it}(\boldsymbol{\psi}_i)}{\partial \sigma_{it}^{2*}} \right\| \sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i) - \sigma_{it}^2(\boldsymbol{\psi}_i)\|, \quad (45)$$

where σ_{it}^{2*} is a point between $\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)$ and $\sigma_{it}^2(\boldsymbol{\psi}_i)$, so that

$$\frac{\partial A_{it}(\boldsymbol{\psi}_i)}{\partial \sigma_{it}^{2*}} = - \frac{2\kappa_i(\nu_i + 1)(\nu_i - 2)(x_{it}^4/f_t^4(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^{2*} + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^3},$$

which is uniformly bounded, in fact

$$\sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial \hat{A}_{it}(\boldsymbol{\psi}_i)}{\partial \sigma_{it}^{2*}} \right\| \leq \sup_{\boldsymbol{\theta} \in \Theta} 2\kappa_i(\nu_i + 1)(\nu_i - 2) \|\hat{b}_t^2(1 - \hat{b}_t)\| < \infty,$$

where \hat{b}_t is as in (27), and hence, bounded between 0 and 1. We then obtain the desired e.a.s. convergence in (43) by an application of Lemma 2.1 in Straumann and Mikosch (2006), since as $t \rightarrow \infty$ the norm $\sup_{\boldsymbol{\theta} \in \Theta} \|\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i) - \sigma_{it}^2(\boldsymbol{\psi}_i)\|$ vanishes by Proposition 3. With the same arguments it is possible to show that the claimed convergence in (44) holds true. Thus, we get

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i) - \nabla \sigma_{it}^2(\boldsymbol{\psi}_i)\| \xrightarrow{\text{e.a.s.}} 0, \quad \text{as} \quad t \rightarrow \infty, \quad (46)$$

and therefore, by combining the result obtained in (42) and (46), and recalling (37) the e.a.s. convergence of the derivative processes follows. □

A.8 Proof of Proposition 5

Proof. We note that

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\| \leq \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla f_t^2(\boldsymbol{\lambda})\| \sup_{\boldsymbol{\theta} \in \Theta} \|\sigma_{it}^2(\boldsymbol{\psi}_i)\| + \sup_{\boldsymbol{\theta} \in \Theta} \|f_t^2(\boldsymbol{\lambda})\| \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \sigma_{it}^2(\boldsymbol{\psi}_i)\|. \quad (47)$$

As the recursion for $\sigma_{it}^2(\boldsymbol{\psi}_i)$ in equation (3), the derivatives in (39) together with the components (40) and (41) are uniformly bounded and hence, we obtain arbitrary large number of bounded moments n'_σ by following closely the arguments of Proposition 2.

Specifically, let us consider the recursion

$$\nabla \sigma_{i,t+1}^2(\boldsymbol{\psi}_i) = A_{it}(\boldsymbol{\psi}_i) \nabla \sigma_{it}^2(\boldsymbol{\psi}_i) + \sigma_{it}^2(\boldsymbol{\psi}_i) \mathbf{S}_{it}(\boldsymbol{\psi}_i), \quad (48)$$

where

$$\mathbf{S}_{it}(\boldsymbol{\psi}_i) = \begin{bmatrix} 1 - 1/\sigma_{it}^2(\boldsymbol{\psi}_i) \\ u_{it}(\boldsymbol{\psi}_i) \\ \kappa_{it} a_{it}(\boldsymbol{\psi}_i) \end{bmatrix}.$$

Therefore, by recursive arguments and given the results in Propositions 3 and 4, we obtain

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)\| &\leq \sup_{\boldsymbol{\theta} \in \Theta} \|A_{it}(\boldsymbol{\psi}_i)\| \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \sigma_{it}^2(\boldsymbol{\psi}_i)\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\sigma_{it}^2(\boldsymbol{\psi}_i) \mathbf{S}_{it}(\boldsymbol{\psi}_i)\| \\ &\leq c_A^t \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \sigma_{i1}^2(\boldsymbol{\psi}_i)\| + \sum_{j=1}^{t-1} c_A^{t-j} \sup_{\boldsymbol{\theta} \in \Theta} \|\sigma_{i,t-j}^2(\boldsymbol{\psi}_i) \mathbf{S}_{it}(\boldsymbol{\psi}_i)\| \\ &\leq \sum_{j=1}^{\infty} c_A^{t-j} \sup_{\boldsymbol{\theta} \in \Theta} \|\sigma_{i,t-j}^2(\boldsymbol{\psi}_i) \mathbf{S}_{it}(\boldsymbol{\psi}_i)\|, \end{aligned}$$

where $c_A < 1$ by the contraction condition in Assumption 4. Thus, the desired result now follows by noting that both $\sigma_{it}^2(\boldsymbol{\psi}_i)$ and $\mathbf{S}_{it}(\boldsymbol{\psi}_i)$ for $i = 1, \dots, N$ are uniformly bounded and then $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \sigma_{i,t+1}^2(\boldsymbol{\psi}_i)\|^{n'_\sigma}] < \infty$ for $n'_\sigma \rightarrow \infty$.

Analogously, we define

$$\nabla f_{t+1}^2(\boldsymbol{\lambda}) = (\beta - \alpha) \nabla f_t^2(\boldsymbol{\lambda}) + \mathbf{r}_t(\boldsymbol{\lambda}), \quad (49)$$

where

$$\mathbf{r}_t(\boldsymbol{\lambda}) = \begin{bmatrix} 1 \\ \frac{1}{N} \sum_{i=1}^N x_{it}^2 - f_t^2(\boldsymbol{\lambda}) \\ f_t^2(\boldsymbol{\lambda}) \end{bmatrix}$$

and hence, we obtain

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla f_{t+1}^2(\boldsymbol{\lambda})\| &\leq c_b \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla f_t^2(\boldsymbol{\lambda})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{r}_t(\boldsymbol{\lambda})\| \\
&\leq c_b^t \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla f_1^2(\boldsymbol{\lambda})\| + \sum_{j=1}^{t-1} c_b^{t-j} \sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{r}_{t-j}(\boldsymbol{\lambda})\| \\
&\leq \sum_{j=1}^{\infty} c_b^{t-j} \sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{r}_{t-j}(\boldsymbol{\lambda})\|,
\end{aligned}$$

with $c_b < 1$ by the contraction condition in Assumption 4, giving us the arbitrary large n'_f as a corollary of Proposition 2.

In conclusion, taking into account (47) we get

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\|_{m'} &\leq \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla f_t^2(\boldsymbol{\lambda})\|_{n'_f} \sup_{\boldsymbol{\theta} \in \Theta} \|\sigma_{it}^2(\boldsymbol{\lambda})\|_{n_\sigma} \\
&\quad + \sup_{\boldsymbol{\theta} \in \Theta} \|f_t^2(\boldsymbol{\lambda})\|_{n_f} \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \sigma_{it}^2(\boldsymbol{\lambda})\|_{n'_\sigma},
\end{aligned}$$

and by applying the generalized Hölder's inequality yields the claimed number of bounded moments

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

where $m' > 2$. □

A.9 Proof of Lemma 4.4

Proof. First, we note that for $i = 1, \dots, N$ and $t = 1, \dots, T$,

$$\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) = \frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial (\boldsymbol{\lambda}, \boldsymbol{\psi}_i)} + \frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial (f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))} \nabla (f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)). \quad (50)$$

The first partial derivatives in equation (50) is a continuous function of strictly stationary and ergodic processes and hence, by an application of the continuous mapping theorem for $\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)$ we infer that $\{\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)\}_{t \in \mathbb{Z}}$ is also stationary and ergodic.

Moreover, because $\{f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{Z}}$ and the derivative processes $\{\nabla (f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))\}_{t \in \mathbb{Z}}$ are

adapted to \mathcal{F}_{t-1} we have that

$$\mathbb{E} \left[\nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \middle| \mathcal{F}_{t-1} \right] = \mathbb{E} \left[\frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)} \middle| \mathcal{F}_{t-1} \right] + \mathbb{E} \left[\frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))} \middle| \mathcal{F}_{t-1} \right] \nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)).$$

The first term in the right hand side of (50) is

$$\frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial \nu_i} = \frac{1}{2} \left[\psi \left(\frac{\nu_i + 1}{2} \right) - \psi \left(\frac{\nu_i}{2} \right) - \frac{1}{\nu_i - 2} + \frac{\nu_i + 1}{\nu_i - 2} b_{it}(\boldsymbol{\theta}) - \log(1 - b_{it}(\boldsymbol{\theta})) \right], \quad (51)$$

while the second term is

$$\frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))} = \frac{1}{2(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))} [(\nu_i + 1)b_{it}(\boldsymbol{\theta}) - 1]. \quad (52)$$

Hence, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we just need to show that, by using the properties of the beta distributed random variables

$$\mathbb{E}[b_{it}(\boldsymbol{\theta}_0) | \mathcal{F}_{t-1}] = 1/(\nu_{i0} + 1) \quad \text{and}$$

$$\mathbb{E}[\log(1 - b_{it}(\boldsymbol{\theta}_0)) | \mathcal{F}_{t-1}] = \psi(\nu_{i0}/2) - \psi((\nu_{i0} + 1)/2),$$

thus, it is straightforward to see that for $i = 1, \dots, N$

$$\mathbb{E}[\nabla \ell_{it}(\boldsymbol{\lambda}_0, \boldsymbol{\psi}_{i0}) | \mathcal{F}_{t-1}] = 0.$$

Note also that this property entails the uniformly boundedness of (51) because $b_{it}(\boldsymbol{\theta})$ is bounded between 0 and 1 for any $\boldsymbol{\theta} \in \Theta$.

Therefore, we obtain the claimed second moment bound by noting that a combination of the Minkowski's, Hölder's and the Jensen's inequality yields

$$\begin{aligned} \mathbb{E} \left[\left| \nabla \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \right|^2 \right] &\leq \left\{ \left(\mathbb{E} \left[\left| \frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)} \right|^2 \right] \right)^{1/2} \right. \\ &\quad \left. + \left(\mathbb{E} \left[\left| \frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i))} \right|^4 \right] \right)^{1/4} \left(\mathbb{E} \left[\left| \nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)) \right|^4 \right] \right)^{1/4} \right\}^2 < \infty, \end{aligned}$$

since, as noted above, the first term in the right hand side is in (51) and the second term is in (52)

and both are uniformly bounded $\forall x_{it} \in \mathbb{R}$, and moreover from Assumptions 1, 2 and 3, we have that $f_t^2(\boldsymbol{\lambda}) \geq \omega > 0$ and $\sigma_{it}^2(\boldsymbol{\psi}_i) \geq \delta_i = (1 - \phi_i) > 0$. The moment bounds of the third term are obtained as an application of Proposition 5 by setting $m' = 4$. \square

A.10 Proof of Lemma 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,

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \nabla \mathcal{L}_{NT}(\boldsymbol{\theta})\| \\ & \leq \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\frac{\partial \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))}}{\frac{\partial^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))^2}} \right\| \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{(\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)) - (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))}{\nabla (\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)) - \nabla (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} \right\|, \end{aligned} \quad (53)$$

where $(\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))$ lies between $(\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))$ and $(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))$. We have already shown in Theorem 4.3 that the first partial derivative in the first term of right hand side of the inequality in (53) is uniformly bounded. Moreover, the second partial derivative is

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_{NT}(\boldsymbol{\theta})}{\partial (f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))^2} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[-\frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))^2} [(\nu_i + 1)b_{it}(\boldsymbol{\theta}) - 1] \right. \\ & \quad \left. - \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} [(\nu_i + 1)b_{it}(\boldsymbol{\theta})(1 - b_{it}(\boldsymbol{\theta}))] \right], \end{aligned}$$

which is uniformly bounded, while both the second term in the right hand side of the inequality in (53) above converge almost surely exponentially fast to zero as shown in Propositions 3 and 4 respectively. Hence, we satisfy the conditions of Lemma 2.1 in Straumann and Mikosch (2006) and we obtain $\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \nabla \mathcal{L}_{NT}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0$ with N fixed and $T \rightarrow \infty$.

Therefore, the score function obeys the central limit theorem for martingales of Billingsley (1961), since the existence of the covariance matrix \mathbf{V} is entailed by Proposition 5. The claimed convergence in distribution is achieved by appealing to Theorem 18.10 (*iv*) of van der Vaart (1998). \square

A.11 Proof of Proposition 6

Proof. We have

$$\nabla^2(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)) = (\nabla^2 f_t^2(\boldsymbol{\lambda}))\sigma_{it}^2(\boldsymbol{\psi}_i) + 2\nabla(f_t^2(\boldsymbol{\lambda})\sigma_{it}^2(\boldsymbol{\psi}_i)) + f_t^2(\boldsymbol{\lambda})(\nabla^2\sigma_{it}^2(\boldsymbol{\psi}_i)), \quad (54)$$

with $\nabla^2 f_t^2(\boldsymbol{\lambda})$ composed by

$$\begin{aligned} \frac{\partial^2 f_{t+1}^2(\boldsymbol{\lambda})}{\partial\omega^2} &= (\beta - \alpha) \frac{\partial^2 f_t^2(\boldsymbol{\lambda})}{\partial\omega^2}, \\ \frac{\partial^2 f_{t+1}^2(\boldsymbol{\lambda})}{\partial\alpha^2} &= (\beta - \alpha) \frac{\partial^2 f_t^2(\boldsymbol{\lambda})}{\partial\alpha^2} - 2 \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial\alpha}, \\ \frac{\partial^2 f_{t+1}^2(\boldsymbol{\lambda})}{\partial\beta^2} &= (\beta - \alpha) \frac{\partial^2 f_t^2(\boldsymbol{\lambda})}{\partial\beta^2} + 2 \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial\beta}, \\ \frac{\partial^2 f_{t+1}^2(\boldsymbol{\lambda})}{\partial\alpha\partial\beta} &= (\beta - \alpha) \frac{\partial^2 f_t^2(\boldsymbol{\lambda})}{\partial\alpha\partial\beta} + \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial^2\alpha} - \frac{\partial f_t^2(\boldsymbol{\lambda})}{\partial\beta}, \end{aligned} \quad (55)$$

while for $i = 1, \dots, N$, $\nabla^2\sigma_{it}^2(\boldsymbol{\psi}_i)$ is composed by

$$\begin{aligned} \frac{\partial^2\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\phi_i^2} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial^2\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\phi_i^2} + \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\phi_i} + 1, \\ \frac{\partial^2\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i^2} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial^2\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i^2} + u_{it}(\boldsymbol{\psi}_i) \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i} + c_{it}(\boldsymbol{\psi}_i), \\ \frac{\partial^2\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\nu_i^2} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial^2\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\nu_i^2} + (\kappa_i a_{it}(\boldsymbol{\psi}_i) + \kappa_i e_{it}(\boldsymbol{\psi}_i)) \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\nu_i} + \kappa_i \sigma_{it}^2(\boldsymbol{\psi}_i) d_{it}(\boldsymbol{\psi}_i), \\ \frac{\partial^2\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\phi_i\partial\kappa_i} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial^2\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\phi_i\partial\kappa_i} + \frac{\partial\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i} + c_{it}(\boldsymbol{\psi}_i), \\ \frac{\partial^2\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\phi_i\partial\nu_i} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial^2\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\phi_i\partial\nu_i} + \kappa_i e_{it}(\boldsymbol{\psi}_i) \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\phi_i} + \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\nu_i}, \\ \frac{\partial^2\sigma_{i,t+1}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i\partial\nu_i} &= A_{it}(\boldsymbol{\psi}_i) \frac{\partial^2\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i\partial\nu_i} + \kappa_i e_{it}(\boldsymbol{\psi}_i) \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\nu_i} + u_{it}(\boldsymbol{\psi}_i) \frac{\partial\sigma_{it}^2(\boldsymbol{\psi}_i)}{\partial\kappa_i} + \sigma_{it}^2(\boldsymbol{\psi}_i) q_{it}(\boldsymbol{\psi}_i), \end{aligned} \quad (56)$$

where

$$\begin{aligned}
c_{it}(\boldsymbol{\psi}_i) &= \left(\frac{(\nu_i + 1)(x_{it}^4/f_t^4(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^2} - 1 \right), \\
d_{it}(\boldsymbol{\psi}_i) &= \left(\frac{2\sigma_{it}^2(\boldsymbol{\psi}_i)(\nu_i + 1)(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^4(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^3} - \frac{2\sigma_{it}^2(\boldsymbol{\psi}_i)(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^2} \right), \\
e_{it}(\boldsymbol{\psi}_i) &= \left(\frac{(x_{it}^4/f_t^4(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^2} - \frac{2\sigma_{it}^2(\boldsymbol{\psi}_i)(\nu_i + 1)(x_{it}^4/f_t^4(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^3} \right), \\
q_{it}(\boldsymbol{\psi}_i) &= \left(\frac{(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))} - \frac{\sigma_{it}^2(\boldsymbol{\psi}_i)(\nu_i + 1)(x_{it}^2/f_t^2(\boldsymbol{\lambda}))}{[(\nu_i - 2)\sigma_{it}^2(\boldsymbol{\psi}_i) + (x_{it}^2/f_t^2(\boldsymbol{\lambda}))]^2} \right).
\end{aligned} \tag{57}$$

The expressions in (55) and (56) show that the contraction conditions $|\beta - \alpha| < 1$ and $\mathbb{E}[\log \sup_{\sigma_i^2} \|A_{it}(\boldsymbol{\psi}_i)\|] < 0$ for the recursion filters $\{\hat{f}_t(\boldsymbol{\lambda})\}_{t \in \mathbb{N}}$ and $\{\hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{N}}$ for $i = 1, \dots, N$ stated in Proposition 3 are the same for the perturbed second derivative processes $\{\nabla^2 \hat{f}_t^2(\boldsymbol{\lambda})\}_{t \in \mathbb{N}}$ and $\{\nabla^2 \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{N}}$. Therefore, we note that the proof of Propositions 3 and 6 can be easily adapted since all the terms in (57) are uniformly bounded with finite log-moments and therefore, there exist unique stationary and ergodic solutions $\{\nabla^2 f_t^2(\boldsymbol{\lambda})\}_{t \in \mathbb{Z}}$ and $\{\nabla^2 \sigma_{it}^2(\boldsymbol{\psi}_i)\}_{t \in \mathbb{Z}}$ such that

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \hat{f}_t^2(\boldsymbol{\lambda}) - \nabla^2 f_t^2(\boldsymbol{\lambda})\| \xrightarrow{\text{e.a.s.}} 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i) - \nabla^2 \sigma_{it}^2(\boldsymbol{\psi}_i)\| \xrightarrow{\text{e.a.s.}} 0.$$

Thus, by combining the latter result, the e.a.s. convergence obtained in Propositions 3 and 6, and recalling (54), we conclude the proof. \square

A.12 Proof of Lemma 4.6

Proof. The first part of the proof follows the same arguments given in the proof of the consistency Theorem 4.3. Then, we split the second derivatives of likelihood function as follows

$$\begin{aligned}
& \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \nabla^2 \mathcal{L}_N(\boldsymbol{\theta})\| \\
& \leq \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta}) - \nabla^2 \mathcal{L}_N(\boldsymbol{\theta})\|. \tag{58}
\end{aligned}$$

To prove that the first term in the right hand side of (58) we note that

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta})\| \\ & \leq \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\frac{\partial \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))}}{\frac{\partial^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))^2}} \right\| \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{(\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}(\boldsymbol{\psi}_i)) - (f_t^2(\boldsymbol{\lambda}) \sigma_{it}(\boldsymbol{\psi}_i))}{\nabla^2(\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i)) - \nabla^2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} \right\|. \end{aligned} \quad (59)$$

where $(\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))$ lies between $(\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))$ and $(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))$. We have already shown in Theorem 4.3 and Lemma 4.5 that the first to partial derivatives in the first term of the right hand side of equation (59) are uniformly bounded, with finite log-moments. As concern the third partial derivative, we have

$$\begin{aligned} \frac{\partial^3 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta})}{\partial (\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))^3} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left[\frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))^4} [(\nu_i + 1) b_{it}(\boldsymbol{\theta}) - 1] \right. \\ & \quad + \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} [(\nu_i + 1) b_{it}(\boldsymbol{\theta}) (1 - b_{it}(\boldsymbol{\theta}))] \\ & \quad + \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))^2} [(\nu_i + 1) b_{it}(\boldsymbol{\theta}) (1 - b_{it}(\boldsymbol{\theta}))] \\ & \quad + \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} [(\nu_i + 1) b_{it}(\boldsymbol{\theta}) (1 - b_{it}(\boldsymbol{\theta})^2)] \\ & \quad \left. - \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} [(\nu_i + 1) b_{it}^2(\boldsymbol{\theta}) (1 - b_{it}(\boldsymbol{\theta}))] \right], \end{aligned}$$

which is uniformly bounded, while all the terms in the right hand side of the inequality in (59) above converge almost surely exponentially fast to zero as shown in Propositions 3, 4 and 6 respectively.

Hence, we satisfy the conditions of Lemma 2.1 in Straumann and Mikosch (2006) and we obtain

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}) - \nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta})\| \xrightarrow{\text{a.s.}} 0 \text{ with } N \text{ fixed and } T \rightarrow \infty.$$

The second term in (58) will also vanish if the moment bound $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)\|] < \infty$ is satisfied since $\{\nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ is stationary and ergodic.

Now, for a single observation we note that

$$\begin{aligned}\nabla^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) &= \frac{\partial^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(\boldsymbol{\lambda}, \boldsymbol{\psi}_i) \partial(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)^\top} + \frac{\partial \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} \nabla^2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)) \\ &\quad + \nabla(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)) (\nabla(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)))^\top.\end{aligned}\quad (60)$$

The components of the first term in (60) are

$$\begin{aligned}\frac{\partial^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial \nu_i^2} &= \frac{1}{4} \left[\psi' \left(\frac{\nu_i + 1}{2} \right) - \psi' \left(\frac{\nu_i}{2} \right) \right] \\ &\quad + \frac{1}{2} \left[\frac{1}{(\nu_i - 2)^2} - \frac{\nu_i + 1}{(\nu_i - 2)^2} b_{it}(\boldsymbol{\theta})(1 - b_{it}(\boldsymbol{\theta})) - \frac{3}{(\nu_i - 2)} b_{it}(\boldsymbol{\theta}) \right],\end{aligned}\quad (61)$$

$$\frac{\partial^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)) \partial \nu_i} = \frac{1}{2} \left[b_{it}(\boldsymbol{\theta}) - (\nu_i + 1) b_{it}(\boldsymbol{\theta})(1 - b_{it}(\boldsymbol{\theta})) \right],\quad (62)$$

and finally

$$\begin{aligned}\frac{\partial^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)}{\partial(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))^2} &= - \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))^2} [(\nu_i + 1) b_{it}(\boldsymbol{\theta}) - 1] \\ &\quad - \frac{1}{2(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))} [(\nu_i + 1) b_{it}(\boldsymbol{\theta})(1 - b_{it}(\boldsymbol{\theta}))].\end{aligned}\quad (63)$$

In line with previous arguments, we can easily check the uniformly boundedness of (61), (62) and (63).

Moreover, from Theorem 4.3 and Proposition 6 we already know that the components of the components of the second term in (60) are uniformly bounded. In conclusion, the existence of the moment bounds for the outer product $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)) (\nabla(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i)))^\top\|]$ is entailed in Proposition 5, by requiring that $m' = 4$.

Therefore, under maintained assumptions, the moment bound $\mathbb{E}[\sup_{\boldsymbol{\theta} \in \Theta} \|\nabla^2 \ell_{it}(\boldsymbol{\lambda}, \boldsymbol{\psi}_i)\|] < \infty$ is always satisfied. \square

A.13 Proof of Theorem 4.7

Proof. We discuss the proof for $T \rightarrow \infty$. Standard arguments for the asymptotic normality proof and the Taylor's theorem, lead to the expansion of the conditional likelihood's score around a neighbourhood

of $\boldsymbol{\theta}_0$, which yields

$$\begin{aligned}
\mathbf{0} &= \sqrt{NT} \nabla \mathcal{L}_{NT}(\hat{\boldsymbol{\theta}}_T) \\
&= \sqrt{NT} \left[\nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}_0) - \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}_0) \right] + \sqrt{NT} \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}_0) \\
&\quad + \left[(\nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta}_0) - \nabla^2 \mathcal{L}_N(\boldsymbol{\theta}_0)) + (\nabla^2 \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}^*) - \nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta}_0)) + \nabla^2 \mathcal{L}_N(\boldsymbol{\theta}_0) \right] \\
&\quad \times \left[\sqrt{NT}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \right], \tag{64}
\end{aligned}$$

where $\boldsymbol{\theta}^*$ is on the cord between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$, componentwise.

From the first line of (64), the convergence of the first difference in square brackets is obtained as follows. Let us consider the expansion

$$\begin{aligned}
&\sqrt{NT} \|\nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}_0) - \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}_0)\| \\
&\leq \sqrt{NT} \left\| \frac{\partial \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}_0)}{\partial (\hat{f}_t^{2*}(\boldsymbol{\lambda}_0) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_{i0}))} \right\| \left\| \frac{(\hat{f}_t^2(\boldsymbol{\lambda}_0) \hat{\sigma}_{it}(\boldsymbol{\psi}_{i0})) - (f_t^2(\boldsymbol{\lambda}_0) \sigma_{it}(\boldsymbol{\psi}_{i0}))}{\nabla (\hat{f}_t^2(\boldsymbol{\lambda}_0) \hat{\sigma}_{it}(\boldsymbol{\psi}_{i0})) - \nabla (f_t^2(\boldsymbol{\lambda}_0) \sigma_{it}(\boldsymbol{\psi}_{i0}))} \right\|, \tag{65}
\end{aligned}$$

where $(\hat{f}_t^{2*}(\boldsymbol{\lambda}) \hat{\sigma}_{it}^{2*}(\boldsymbol{\psi}_i))$ lies between $(\hat{f}_t^2(\boldsymbol{\lambda}) \hat{\sigma}_{it}^2(\boldsymbol{\psi}_i))$ and $(f_t^2(\boldsymbol{\lambda}) \sigma_{it}^2(\boldsymbol{\psi}_i))$. Now we note that under maintained assumption, Propositions (3) and (4) imply that the terms in the second component of the right hand side of (65) will vanish as $T \rightarrow \infty$. As concern the first component of the right hand side of (65), we have already shown in Theorem 4.3 and Lemma 4.5 that the considered partial derivatives are uniformly bounded and hence, we can write

$$\|\nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}_0) - \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}_0)\| \leq O_p(1) o_{e.a.s.}(1) = o_{e.a.s.}(1),$$

which implies that $\sqrt{NT} \|\nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}_0) - \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}_0)\| \xrightarrow{\text{a.s.}} 0$ as $T \rightarrow \infty$, since N is fixed. Therefore, since Lemma 4.5 entails the fact that $\sqrt{NT} \nabla \mathcal{L}_{NT}(\boldsymbol{\theta}_0)$ obeys the CLT for martingales, we obtain that $\sqrt{NT} \nabla \hat{\mathcal{L}}_{NT}(\boldsymbol{\theta}_0)$ has the same asymptotic distribution by the asymptotic equivalence lemma, see Lemma 4.7 White (2001).

Now consider the second line of (64). Lemma 4.6 demonstrates that the initial 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 law of large numbers guarantee that $\|\nabla^2 \mathcal{L}_{NT}(\boldsymbol{\theta}_0) - \nabla^2 \mathcal{L}_N(\boldsymbol{\theta}_0)\| \xrightarrow{\text{a.s.}} 0$ as $T \rightarrow \infty$, since N is fixed, see Lemma 4.6. Thus, the nonsingularity of the limit $\nabla^2 \mathcal{L}_N(\boldsymbol{\theta})$ is implied by the uniqueness of $\boldsymbol{\theta}_0 \in \Theta$ as maximizer of $\mathcal{L}_N(\boldsymbol{\theta})$, see Lemma 4.2 and then, we can solve the equation above. Finally, we apply the Slutsky's Lemma (see Lemma 2.8 (iii) of van der Vaart (1998)) and complete the proof. \square

References

- Alexander, C. (2001). Orthogonal GARCH. *Mastering Risk*, Financial Times(2):21–38.
- Amado, C. and Teräsvirta, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics*, 175(2):142 – 153.
- Barigozzi, M., Brownlees, C., Gallo, G. M., and Veredas, D. (2014). Disentangling systematic and idiosyncratic dynamics in panels of volatility measures. *Journal of Econometrics*, 182(2):364 – 384.
- Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.
- Billingsley, P. (1961). The Lindeberg-Lévy theorem for Martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.
- Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12:1019–1052.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327.
- Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.
- Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization*, 31(4):942–959.
- Cox, D. R. (1981). Statistical analysis of time series: Some recent developments (with discussion). *Scandinavian Journal of Statistics*, 8(2):93–115.

- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.
- Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164:188–205.
- Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. and Rangel, J. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21(3):1187–1222.
- Engle, R. F. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3):339–350.
- Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.
- Engle, R. F. and Lee, G. (1999). A long-run and short-run component model of stock return volatility. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, pages 475–497.
- Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a Factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1):213 – 237.
- Engle, R. F. and Watson, M. W. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76(376):774–781.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

- Gonzalo, J. and Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13(1):27–35.
- Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics*, 159(1):55 – 73.
- Harvey, A. C. (2013). *Dynamic models for Volatility and Heavy Tails*. Cambridge University Press.
- Harvey, A. C. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, 109(507):1112–1122.
- Ling, S. and McAleer, M. (2002). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics*, 106(1):109 – 117.
- Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.
- Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30:481–493.
- Stout, W. (1974). *Almost sure convergence*. Academic Press.
- Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics*, 34(5):2449–2495.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley, New York, 2nd edition.
- van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.
- van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564.

Wang, F. and Ghysels, E. (2015). Econometric analysis of volatility component models. *Econometric Theory*, 31(2):362–393.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians*. Emerald Group Publishing.