TI 2022-070/III
Tinbergen Institute Discussion Paper

Dynamic Partial Correlation Models

Enzo D'Innocenzo[1]
André Lucas[1,2]

1 Vrije Universiteit Amsterdam

2 Tinbergen Institute

Dynamic Partial Correlation Models

Enzo D'Innocenzo[a,*] and Andre Lucas[a,b]

[a] Vrije Universiteit Amsterdam
[b] Tinbergen Institute

September 27, 2022

Abstract

We introduce a new, easily scalable model for dynamic conditional correlation matrices based on a recursion of dynamic bivariate partial correlation models. By exploiting the model's recursive structure and the theory of perturbed stochastic recurrence equations, we establish stationarity, ergodicity, and filter invertibility in the multivariate setting using conditions for bivariate slices of the data only. From this, we establish consistency and asymptotic normality of the maximum likelihood estimator for the model's static parameters. The new model outperforms benchmarks like the t-cDCC and the multivariate t-GAS, both in simulations and in an in-sample and out-of-sample asset pricing application to 1980–2021 US stock returns across twelve industries.

1 Introduction

Modeling multivariate covariance and correlation structures is a well-established research topic in the econometric literature given its importance for decision making under uncertainty and the typical stylized facts of many observed economic time series; see for instance the overviews of Bauwens et al. (2012) and Francq and Zakoian (2019). Since Engle (2002) and Tse and Tsui (2002), most researchers agree that empirically the most useful way to decompose the time-variation in conditional covariance matrices is via a variance and a correlation related component. A key example is the DCC model of Engle (2002), which is the standard benchmark when new models are introduced. It first accounts for time-variation in the variances of the marginal series, and then investigates whether there is any time-variation left in the correlations that cannot be attributed to time-varying variances. The second step of this approach requires one to model the dynamics of conditional correlation matrices.

*Corresponding author. Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands, e-mail: e.dinnocenzo@vu.nl

Modeling the dynamics of conditional correlation matrices is challenging given the joint restrictions that should hold for such matrices: they (i) need to be positive (semi)-definite and (ii) have ones on the diagonal. The DCC of Engle (2002) ensures this by an algorithmically simple, but theoretically hard, nonlinear matrix-transformation. The drawback of this is that it becomes hard to formulate conditions required for stationarity, ergodicity, and filter invertibility. Establishing such stochastic properties is crucial for dynamic models, since it opens the door to a rigorous econometric analysis of theoretical asymptotic properties of such models. For example, due to the complexity of the nonlinear matrix transformation in the DCC, the asymptotic properties of the quasi maximum likelihood estimator for the DCC are as yet still unknown; see also the related (heuristic) discussion in Aielli (2013).

Alternative parameterizations of correlation matrices have been proposed in the literature, like the hypersphere parameterization of the choleski decomposition of a correlation matrix as in Rapisarda et al. (2007), Creal et al. (2011), and Buccheri et al. (2021), or the log correlation matrix transformation of Archakov and Hansen (2021) as also used by Hafner and Wang (2021). For all these parameterizations, however, formulating conditions for filter invertibility remains hard. In addition, all these models are cast in matrix format, which means that contraction conditions like that of Bougerol (1993) can become increasingly strict in higher dimensions due to the use of matrix norms.[1]

In this paper, we contribute to the literature by introducing a novel class of nonlinear, heavy-tailed time-series models for dynamic conditional correlation matrices that avoid most of the above drawbacks. In particular, instead of considering a full multivariate model for the entire dynamic conditional correlation matrix at once, we define univariate nonlinear filters for conditional *partial* correlation coefficients based on bivariate slices of the data only. This also allows us to easily impose zero restrictions on particular partial correlations in case this is theoretically or empirically desirable. By stacking the different bivariate models and relying on Anderson (1958) and Joe (2006), we can easily reconstruct the full multivariate correlation matrix. The matrix constructed in this way automatically has ones on the diagonal and satisfies the restrictions of positive-definiteness.

We endow the models for the bivariate data slices with score-driven dynamics for the univariate partial correlation parameter, using a conditional Student's t distribution for the innovations. In this way we obtain a robust filter for the entire correlation matrix; see Creal et al. (2013) and Harvey (2013) for an introduction to score-driven dynamics. Given the sequence of bivariate models for the data, each model can use its own pair-specific parameters that govern the dynamics of the conditional partial correlation for that pair. This flexibility provides substantial value-added and is also empirically relevant as shown in our empirical application to US stock returns. As mentioned, all of the model's pairwise flexibility for the partial correlations does not jeopardize the positive definiteness of the

[1]For instance, for the Frobenius norm of a matrix A we have $\|A\| = \sum_{i,j} A_{i,j}^2$. If this norm has to be bounded in a Bougerol (1993) type condition, this becomes increasingly restrictive if the dimension of A grows.

implied multivariate conditional Pearson correlation matrix for the entire system in any way.

Splitting the modeling approach from a multivariate problem into a recursion of conditional models for bivariate slices of the data not only provides benefits from computational or model design perspective. We show that the approach also leads to advantages for a rigorous theoretical analysis of the model's asymptotic properties; compare Blasques et al. (2022). We consider an asymptotic setting where the sample size T goes to infinity for a fixed dimension N of the time series, and leave a setting with both N and T going to infinity to a future paper. By using the theory on perturbed stochastic recurrence equations of Straumann and Mikosch (2006), we are able to provide clear conditions for stationarity, ergodicity, and filter invertibility, as well as conditions for consistency and asymptotic normality of the maximum likelihood estimator. All these conditions only make use of univariate contraction requirements based on bivariate data slices, even if the dimension of the entire data vector is substantially larger than two. An important advantage of this approach is that the restrictions can be more relaxed than dealing with the entire multivariate system at once. In essence, we prove that the conditions for bivariate models like Blasques et al. (2018) continue to hold in slightly modified form for the fully multivariate setting. Similar rigorous results for these non-linear correlation models were not available before. We also mention that due to the use of a robust filtering method, we only require a limited ($2 + \delta$ for some small $\delta > 0$) number of moments for the observations in order for the model and filter to behave well. This stands in sharp contrast with the asymptotic theory developed for MGARCH models, like in the BEKK-GARCH models where at least 6-order moments of the observations may be required; see Comte and Lieberman (2003), Hafner and Preminger (2009), and Pedersen and Rahbek (2014).

The new model performs well in a controled simulation setting, where it outperforms typical strong benchmarks like the cDCC of Engle (2002) and Aielli (2013) based on the Student's t distribution, and the t-GAS model with hypersphere parameterization of Creal et al. (2011) and Buccheri et al. (2021). We also apply the model both in-sample and out-of-sample to study its asset pricing implications for time-series of US stock returns over the period 1980–2021 across 12 US industry portfolios as in Engle (2016), Boudt et al. (2017) and Darolles et al. (2018). The empirical application considers time-varying betas in a risk attribution model with a market (MKT - RF), size (SMB), and value (HML) risk factor and assesses performance in terms of tracking errors rather than statistical measures of fit only. The results reveal that the new model continues to outperform its benchmarks. The dynamic partial correlation model is the only model that is contained in the model confidence set (MCS) of Hansen et al. (2011) across all 12 industries, whereas the t-cDCC and t-GAS are only in there once. Also in terms of in-sample fit or out-of-sample Mincer-Zarnowitz regressions, the new model outperforms the benchmarks.

The rest of this paper is organized as follows. Section 2 introduces the model. The asymptotic properties of the model are derived in Section 3. Section 4 provides the empirical application. Section 5 concludes. All proofs and additional required technical materials are provided in the online Appendix.

2 The model

2.1 Approaches to modeling correlation matrices

Consider a real-valued N-dimensional time series $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ and a sequence of corresponding information sets $\mathcal{F}_{t-1} = \{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots\}$. We focus on modeling the dynamics of the conditional Pearson correlation matrix $\boldsymbol{R}_t$ of $\boldsymbol{y}_t$ given $\mathcal{F}_{t-1}$. More specifically, we consider the case

$$\boldsymbol{y}_t|\mathcal{F}_{t-1} \sim t\left(\mathbf{0}_N, \left(1-2\nu^{-1}\right)\cdot \boldsymbol{R}_t, \nu\right), \qquad \nu > 2, \tag{1}$$

where $t(\mu, \Omega, \nu)$ denotes an N-dimensional Student's t distribution with location μ, scale matrix Ω, and $\nu > 2$ degrees of freedom.[2] We assume $\boldsymbol{R}_t$ is a measurable function of $\mathcal{F}_{t-1}$, such that the model is observation-driven. The model can easily be extended to allow for a non-zero location and for non-unit variances as well as for other distributions. In addition, with a slight extension, our model can be extended into a dynamic Student's t copula framework. For expositional purposes, however, we focus on the current more constrained set-up in (1) to better highlight what is new in our approach.

As mentioned in the introduction, one of the challenges in models such as (1) is the parameterization of a dynamic conditional correlation matrix $\boldsymbol{R}_t$. The matrix $\boldsymbol{R}_t$ not only has to be positive definite, but also needs to have unit entries on the diagonal. So far, three main approaches to tackle this issue have been put forward in the literature. The first approach is that of Engle (2002). It models the covariance matrix directly and standardizes is by pre- and post-multiplying by the square root inverse of its diagonal to ensure the correlation matrix structure with unit entries on the diagonal. A second approach casts the correlation matrix entries into hypersphere coordinates and models the dynamic behavior of these spherical coordinates rather than of the original correlations themselves; see Rapisarda et al. (2007), Creal et al. (2011), and Buccheri et al. (2021). Finally, Archakov and Hansen (2021) introduce the possibility of modeling the strictly lower-half of the log-correlation matrix. Separate models can be used for each of these unconstrained entries. Putting the individual entries back into a matrix and taking the matrix exponential of this, one automatically recovers a proper correlation matrix. This approach is extended to a dynamic setting by Hafner and Wang

[2] Alternatively, we could use $\boldsymbol{R}_t$ as a scaling matrix and relax subsequent moment conditions even further. The current parameterization with $\nu > 2$, however, allows us to interpret $\boldsymbol{R}_t$ directly as a real Pearson correlation matrix.

(2021) using score-driven dynamics.

The non-linear re-parameterizations used to obtain a proper correlation matrix by design complicates a rigorous analysis of the asymptotic properties of the model. In addition, all of the above approaches treat the dynamics of $\boldsymbol{R}_t$ in its matrix form. This is typically accompanied by a restrictive parameterization of the matrix dynamics. A much more flexible approach would be to model each of the pairwise correlations separately. That, however, is problematic as it need not produce a positive definite correlation matrix. In this paper, we solve this by looking at pairwise patterns of *partial* correlations using the work of Anderson (1958) and Joe (2006). As a result, we need not worry about positive definiteness of the implied full correlation matrix: pairwise partial correlation coefficients can be modeled independently with the only restriction that they lie in the interval $(-1, 1)$. As long as all the pairwise partial correlations (as defined further below) lie in this interval, the implied Pearson correlation matrix will always be a proper correlation matrix. As we see later, this has important advantages, both in terms of the flexibility of the model construction, the model's computational and stability aspects, its theoretical statistical properties, and its empirical performance.

2.2 From partial correlations to correlation matrices

A conditional partial correlation $\rho_{i,j|L_{ij}\,;t}$ for a set of indices L_{ij} with $i,j \notin L_{ij}$ is defined as the correlation between $\boldsymbol{y}_{i,t}$ and $\boldsymbol{y}_{j,t}$, conditional on $\mathcal{F}_{t-1}$ and on $\boldsymbol{y}_{L_{ij},t}$, where $\boldsymbol{y}_{L_{ij},t}$ is a vector containing the values of $\boldsymbol{y}_{k,t}$ for $k \in L_{ij}$. If $L_{ij} = \emptyset$ the conditional partial correlation collapses to the standard conditional correlation coefficient (conditional on $\mathcal{F}_{t-1}$). Joe (2006) notes that every $N \times N$ correlation matrix can be parameterized in terms of $N(N-1)/2$ correlation parameters. The first $N-1$ parameters are standard pairwise Pearson conditional correlations $\rho_{i,i+1;t}$ for $i = 1, \ldots, N-1$ and $L_{ij} = \emptyset$. The remaining $(N-2)(N-1)/2$ parameters are the conditional partial correlations $\rho_{i,j|L_{ij}\,;t}$ for $L_{ij} = \{i+1, \ldots, j-1\}$ for $i = 1, \ldots, N-1$ and $j = i+1, \ldots, N$, i.e., the conditional partial correlations between $\boldsymbol{y}_{i,t}$ and $\boldsymbol{y}_{j,t}$ conditioning on all intermediate coordinates between i and j.[3]

Define $\boldsymbol{V}_{i,j\,;t} = \rho_{i,j\,;t} - \boldsymbol{R}_{i,L_{ij};t}\boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{R}_{L_{ij},j;t}$. Then the link between pairwise and partial correlations is obtained from Anderson (1958) and Joe (2006) via the recursive formula

$$\rho_{i,j|L_{ij}\,;t} = \frac{\rho_{i,j\,;t} - \boldsymbol{R}_{i,L_{ij};t}\boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{R}_{L_{ij},j;t}}{\sqrt{\left(1 - \boldsymbol{R}_{i,L_{ij};t}\boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{R}_{L_{ij},i;t}\right) \cdot \left(1 - \boldsymbol{R}_{j,L_{ij};t}\boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{R}_{L_{ij},j;t}\right)}} = \frac{\boldsymbol{V}_{i,j|L_{ij}\,;t}}{\sqrt{\boldsymbol{V}_{i,i|L_{ij}\,;t} \cdot \boldsymbol{V}_{j,j|L_{ij}\,;t}}}, \tag{2}$$

[3] There are of course many different ways to construct the full correlation matrix from a sequence of pairs. In the main text, we adhere to the original proposal of Joe (2006). In the empirical section we consider how different sequences of pairs pairs result affect the model's fit.

for $i = 1, \ldots, N-1$, $j = i+1, \ldots, N$, and $L_{ij} = \{i+1, \ldots, j-1\}$, where

$$\mathrm{corr}(\boldsymbol{y}_{i:j\,;\,t}) = \begin{bmatrix} 1 & \boldsymbol{R}_{i,L_{ij};t} & \rho_{i,j\,;\,t} \\ \boldsymbol{R}_{L_{ij},i;t} & \boldsymbol{R}_{L_{ij},L_{ij};t} & \boldsymbol{R}_{L_{ij},j;t} \\ \rho_{i,j\,;\,t} & \boldsymbol{R}_{j,L_{ij};t} & 1 \end{bmatrix}, \tag{3}$$

and $\boldsymbol{y}_{i:j\,;\,t} = (\boldsymbol{y}_{i,t}, \ldots, \boldsymbol{y}_{j,t})^\top$. Inverting (2), we easily obtain the Pearson correlation as a function of the partial correlations and Pearson correlations:

$$\rho_{i,j\,;\,t} = \boldsymbol{R}_{i,L_{ij};t}\boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{R}_{L_{ij},j;t} + \rho_{i,j|L_{ij}\,;\,t}\sqrt{\boldsymbol{V}_{i,i|L_{ij}\,;\,t} \cdot \boldsymbol{V}_{j,j|L_{ij}\,;\,t}}. \tag{4}$$

Interestingly, as Joe (2006) points out, the $N-1$ pairwise correlations and the $(N-2)(N-1)/2$ partial correlations can vary independently in the interval $(-1, 1)$. The implied Pearson correlation matrix will always be positive definite by construction. Thus, by modeling the dynamics of the partial correlations, we can use (4) to obtain a dynamic positive definite conditional correlation matrix $\boldsymbol{R}_t$ for all t.

A major advantage of parameterizing a correlation matrix in terms of its partial correlations is that we only have to consider bivariate relationships. The full multivariate nature of the problem can be deferred until we have to evaluate the full likelihood function. In addition, parameter restrictions on the dynamic partial correlations take a much simpler for than when dealing with the entire matrix $\boldsymbol{R}_t$ in one step. Finally, estimating a sequence of bivariate models can lead to computational gains compared to a fully-fledged likelihood optimization of the multivariate model, if only to obtain good starting values for the latter.

2.3 Dynamic specification of the partial correlations

To describe the dynamics of the correlation matrix $\boldsymbol{R}_t$ via its partial correlations, we use score-driven dynamics as introduced by Creal et al. (2013) and Harvey (2013). For a hypersphere and a log correlation matrix parameterization this was done Creal et al. (2011) and Hafner and Wang (2021), respectively. In our setting, however, we do not require the matrix-valued full $\boldsymbol{R}_t$, but only work with bivariate partial correlations instead.

The key step in making our approach feasible and scalable is obtained by observing that for $j > i$ the conditional distribution of $(\boldsymbol{y}_{i,t}, \boldsymbol{y}_{j,t})^\top$ in (1) conditional on $\mathcal{F}_{t-1}$ *and* $\boldsymbol{y}_{L_{ij},t} = \{\boldsymbol{y}_{k,t}\}_{k \in L_{ij}}$ is

Student's t

$$\begin{pmatrix} \boldsymbol{y}_{i,t} \\ \boldsymbol{y}_{j,t} \end{pmatrix} \mid \mathcal{F}_{t-1}, \boldsymbol{y}_{L_{ij},t} \quad \sim \quad t\Big(\boldsymbol{\mu}_{i,j|L_{ij}\,;t}\,,\; \boldsymbol{D}^{1/2}_{i,j|L_{ij}\,;t}\, \boldsymbol{R}_{i,j|L_{ij}\,;t}\, \boldsymbol{D}^{1/2}_{i,j|L_{ij}\,;t}\,,\; \nu_{i,j|L_{ij}} \Big), \tag{5}$$

where $\boldsymbol{R}_{i,j|L_{ij}\,;t}$ is the conditional partial (bivariate) correlation matrix

$$\boldsymbol{R}_{i,j|L_{ij}\,;t} = \begin{bmatrix} 1 & \rho_{i,j|L_{ij}\,;t} \\ \rho_{i,j|L_{ij}\,;t} & 1 \end{bmatrix},$$

$\nu_{i,j|L_{ij}} = \nu + \#L_{ij} = \nu + j - i - 1$ is the degrees of freedom parameter,

$$\boldsymbol{\mu}_{i,j|L_{ij}\,;t} = \begin{pmatrix} \boldsymbol{R}_{i,L_{ij};t} \\ \boldsymbol{R}_{j,L_{ij};t} \end{pmatrix} \boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\; \boldsymbol{y}_{L_{ij},t}, \tag{6}$$

is the location parameter, and

$$\boldsymbol{D}_{i,j|L_{ij}\,;t} = \frac{(\nu-2)(\nu + \boldsymbol{y}^{\top}_{L_{ij},t}\boldsymbol{R}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{y}_{L_{ij},t})}{\nu \cdot \nu_{i,j|L_{ij}}} \begin{pmatrix} \boldsymbol{V}_{i,i|L_{ij}\,;t} & 0 \\ 0 & \boldsymbol{V}_{j,j|L_{ij}\,;t} \end{pmatrix} \tag{7}$$

a diagonal matrix holding the coordinate wise scale parameters; see Roth (2013) or Ding (2016). Note that for $j = i+1$, $\#L_{ij} = 0$ and L_{ij} is the empty set, such that the location $\boldsymbol{\mu}_{i,j|L_{ij}\,;t}$ parameter collapses to zero, while $\boldsymbol{D}^{1/2}_{i,j|L_{ij}\,;t}\, \boldsymbol{R}_{i,j|L_{ij}\,;t}\, \boldsymbol{D}^{1/2}_{i,j|L_{ij}\,;t}$ collapses to the pairwise Pearson correlation matrix for $(\boldsymbol{y}_{i,t}, \boldsymbol{y}_{j,t})^{\top}$.

We can use (5) to *recursively* build the dynamic correlation matrix via univariate transition equations for the partial correlations $\rho_{i,j|L_{ij}\,;t}$ using the bivariate data slice for pair (i,j). To see this, consider a trivariate example. In a first step, we use (5) to model $(\boldsymbol{y}_{1,t}, \boldsymbol{y}_{2,t})^{\top}$. Using the score dynamics of Creal et al. (2011, 2013) this gives a transition equation for the dynamics of $\rho_{1,2,t}$. By choosing a proper re-parameterization such as $\rho_{1,2,t} = \tanh(f_{1,2,t})$, we can ensure the (partial) correlation lies in the interval $(-1,1)$ for any $f_{1,2,t} \in \mathbb{R}$. Next, we repeat this procedure for $(\boldsymbol{y}_{2,t}, \boldsymbol{y}_{3,t})^{\top}$, obtaining a model for the dynamics of $\rho_{2,3,t}$. Finally, we consider (5) for $(\boldsymbol{y}_{1,t}, \boldsymbol{y}_{3,t})^{\top}$ conditional on $\boldsymbol{y}_{2,t}$, obtaining the dynamics for (a possibly re-parameterized version of) $\rho_{1,3|2\,;t}$. To recover the Pearson correlation $\rho_{1,3,t}$ and thus the entire correlation matrix, we use $\rho_{1,3|2\,;t}$ and the correlations $\rho_{1,2,t}$ and $\rho_{2,3,t}$ obtained in the previous steps together with the inverse mapping from $\rho_{1,3|2\,;t}$ to $\rho_{1,3,t}$ in equation (4).

Because we only have to work with the bivariate conditional distributions in (5), all transition equations for the $\rho_{i,j|L_{ij}\,;t}$-s take a similar format. In particular, we have the following result.

Proposition 1 (**score recursions**)**.** *Consider a differentiable bijective parameterization* $\rho_{i,j|L_{ij}\,;t} = g(f_{i,j|L_{ij}\,;t})$ *for* $f_{i,j|L_{ij}\,;t} \in \mathbb{R}$*. Define* $\boldsymbol{y}_{i,j\,;t} = (\boldsymbol{y}_{i,t}, \boldsymbol{y}_{j,t})^\top$ *for* $j > i$ *and let* $p(\boldsymbol{y}_{i,j\,;t} \mid \boldsymbol{y}_{L_{ij},t}, \mathcal{F}_{t-1})$ *be the Student's t pdf corresponding to (5). Then we have the score expression*

$$
\begin{aligned}
s_{i,j|L_{ij}\,;t} &= \frac{\partial \log p(\boldsymbol{y}_{i,j\,;t} \mid \boldsymbol{y}_{L_{ij},t}, \mathcal{F}_{t-1})}{\partial f_{i,j|L_{ij}\,;t}} \\
&= \frac{1}{2}\boldsymbol{G}_{i,j|L_{ij}\,;t}^\top \left(\boldsymbol{R}_{i,j|L_{ij}\,;t}^{-1} \otimes \boldsymbol{R}_{i,j|L_{ij}\,;t}^{-1}\right) \times \\
&\qquad \operatorname{vec}\left(w_{i,j|L_{ij}\,;t} \cdot \boldsymbol{D}_{i,j|L_{ij}\,;t}^{-1/2} \left(\boldsymbol{y}_{i,j\,;t} - \boldsymbol{\mu}_{i,j|L_{ij}\,;t}\right) \left(\boldsymbol{y}_{i,j\,;t} - \boldsymbol{\mu}_{i,j|L_{ij}\,;t}\right)^\top \boldsymbol{D}_{i,j|L_{ij}\,;t}^{-1/2} - \boldsymbol{R}_{i,j|L_{ij}\,;t}\right),
\end{aligned}
\tag{8}
$$

for $i = 1, \ldots, N-1$*,* $j = i+1, \ldots, N$*, and* $L_{ij} = \{i+1, \ldots, j-1\}$*, with*

$$
\begin{aligned}
w_{i,j|L_{ij}\,;t} &= \frac{\nu_{i,j|L_{ij}} + 2}{\nu_{i,j|L_{ij}} + (\boldsymbol{y}_{i,j\,;t} - \boldsymbol{\mu}_{i,j|L_{ij}\,;t})^\top \boldsymbol{D}_{i,j|L_{ij}\,;t}^{-1/2} \boldsymbol{R}_{i,j|L_{ij}\,;t}^{-1} \boldsymbol{D}_{i,j|L_{ij}\,;t}^{-1/2} (\boldsymbol{y}_{i,j\,;t} - \boldsymbol{\mu}_{i,j|L_{ij}\,;t})}, \\
\boldsymbol{G}_{i,j|L_{ij}\,;t} &= \partial \operatorname{vec}(\boldsymbol{R}_{i,j|L_{ij}\,;t}) / \partial f_{i,j|L_{ij}\,;t} = \dot{g}\left(f_{i,j|L_{ij}\,;t}\right) \cdot \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}^\top.
\end{aligned}
$$

This leads to the score transition equation

$$
f_{i,j|L_{ij}\,;t+1} = \omega_{i,j|L_{ij}} + \beta_{i,j|L_{ij}}\ f_{i,j|L_{ij}\,;t} + \alpha_{i,j|L_{ij}}\ s_{i,j|L_{ij}\,;t}, \tag{9}
$$

where we use unit score scaling in the sense of Creal et al. (2013).[4]

The result in Proposition 1 has a number of key differences with earlier score-driven dynamic correlation models. We mention five of them. First, unlike the matrix equations in for instance Creal et al. (2011), Opschoor et al. (2018, 2021), and Hafner and Wang (2021), the recursions in (9) are all univariate for $i = 1, \ldots, N-1$, $j = i+1, \ldots, N$, and $L_{ij} = \{i+1, \ldots, j-1\}$. Second, as a result of this, the parameters in (9) can be estimated recursively for a given value of ν, starting with the pairs $(i, i+1)$ for $i = 1, \ldots, N-1$, followed by the pairs $(i, i+2)$ for $i = 1, \ldots, N-2$, and so on, up to the last pair $(1, N)$. Third, the dynamic parameters $\boldsymbol{\mu}_{i,j|L_{ij}\,;t}$, $\boldsymbol{D}_{i,j|L_{ij}\,;t}$, and $\boldsymbol{R}_{i,j|L_{ij}\,;t}$ all depend on the data and on values of other dynamic parameters estimated in a previous step. However, the system of equations is recursive rather than simultaneous, which has the potential to substantially simplify the estimation. Fourth, because of its bivariate nature, the current set-up of designing a dynamic correlation matrix is perfectly scalable to higher dimensions: there is no worry about the final correlation matrix $\boldsymbol{R}_t$ not being positive definite, as we have modeled the partial correlations

[4] We have also experimented with alternative forms of scaling, such as scaling the score by an additional factor $(1 - \rho_{i,j|L_{ij}}^2)$ to mitigate score step sizes near the edges of the domain. This results in somewhat smoother paths of the empirical correlations in Section 4 and modest changes in the stationarity and invertibility conditions.

directly rather than their Pearson counterparts. The scalability also allows for gains in the likelihood optimization by partially splitting it into subproblems that can potentially at least partially be run in parallel. Fifth, the approach based on partial correlations allows us to easily impose zero restrictions on some of the partial correlations if this is desirable from for instance a theoretical perspective. Imposing such restrictions in a dynamic Pearson correlation matrix parameterization, by contrast, is much harder.

2.4 Maximum Likelihood estimation

As our model is observation driven, the likelihood is known in closed form as

$$\begin{aligned}
\hat{L}_T(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta}), \qquad (10)\\
\hat{\ell}_t(\boldsymbol{\theta}) &= \Bigg\{ \log\Gamma\left(\frac{\nu+N}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\log\big((\nu-2)\pi\big) \\
&\qquad - \frac{1}{2}\log|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})| + \frac{\nu+N}{2}\log\left(1 + \frac{\boldsymbol{y}_t^\top \hat{\boldsymbol{R}}_t(\boldsymbol{\theta})^{-1}\boldsymbol{y}_t}{\nu-2}\right)\Bigg\},
\end{aligned}$$

where $\boldsymbol{\theta}$ contains ν, $\omega_{i,j|L_{ij}}$, $\alpha_{i,j|L_{ij}}$, $\beta_{i,j|L_{ij}}$, for $i = 1,\ldots,N-1$ and $j = i+1,\ldots,N$, and $\{\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})\}_{t=1}^{T}$ contains the filtered correlation matrices using the score driven recursions from Proposition 1, initialized at some $\hat{\boldsymbol{R}}_1$. In our empirical application, we set $\hat{\boldsymbol{R}}_1$ to the correlation matrix of the first 100 observations.

The likelihood in (10) can be optimized numerically using standard software to yield the maximum likelihood estimator (MLE)

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \hat{L}_T(\boldsymbol{\theta}). \qquad (11)$$

Note that the optimization could make use of the recursive structure to obtain good starting values of $\boldsymbol{\theta}$. For instance, one could recursively estimate $\nu_{i,j|L_{ij}}$, $\omega_{i,j|L_{ij}}$, $\alpha_{i,j|L_{ij}}$, and $\beta_{i,j|L_{ij}}$ in an unconstrained way using the bivariate log-likelihood only based on (5). To obtain a single starting value for ν, one could map the different $\nu_{i,j|L_{ij}}$ estimates into a single value via a moments estimator and the theoretical relation $\nu_{i,j} = \nu + j - i - 1$. This initial estimate of ν combined with the estimates of $\omega_{i,j|L_{ij}}$, $\alpha_{i,j|L_{ij}}$, and $\beta_{i,j|L_{ij}}$, can then be used to start the multivariate likelihood maximization problem in (10)–(11).

3 Asymptotic properties

In this section, we study the asymptotic properties of the model. We first study the stationarity properties of the model as a data generating process (DGP) in Section 3.1, followed by filter invertibility in Section 3.2. Finally, we study the consistency and asymptotic normality of the maximum likelihood estimator for the static parameters of the model in Section 3.3. The recursive structure of the partial correlation model will prove extremely useful here: the exponentially fast almost sure convergence of the filtered time-varying parameters $\rho_{i,j|L_{ij};t}$ allows us to use them as a plug-in estimators in subsequent recursions without loosing filter invertibility. As a result, we can obtain consistency and asymptotic normality of the static parameters $\boldsymbol{\theta}$.

3.1 Stationarity and ergodicity of the model

To establish stationarity and ergodicity of $\boldsymbol{y}_t$, we first consider the model as a DGP. Using (1), we can rewrite (9) in the stochastic recurrence equation (SRE) representation defined by Bougerol (1993) and Straumann and Mikosch (2006). In this subsection and the next, we are somewhat more meticulous regarding notation. We write $\hat{\boldsymbol{R}}_t$ as the true sequence of bivariate correlation matrices, initialized at a fixed $\hat{\boldsymbol{R}}_1$. We write $\boldsymbol{R}_t$ without a hat to indicate its uninitialized stationary and ergodic limit sequence, if it exists. Similar notation is used for the partial correlations $\rho_{i,j|L_{ij};t}$ and their transformations $f_{i,j|L_{ij};t}$. Based on the SRE representation, we formulate conditions for the convergence of the random sequences $\{\hat{f}_{i,j|L_{ij};t}\}_{t\in\mathbb{N}}$ initialized at fixed values $\hat{f}_{i,j|L_{ij};1}$ to unique strictly stationary and ergodic sequences $\{f_{i,j|L_{ij};t}\}_{t\in\mathbb{Z}}$. We make the following three assumptions.

Assumption 1. *The partial correlation coefficients are defined using the parametrization* $\rho_{i,j|L_{ij};t} = g(f_{i,j|L_{ij};t}) = \epsilon \cdot \tanh(f_{i,j|L_{ij};t})$ *for* $i = 1, \ldots, N-1$, *and* $j = i+1, \ldots, N$ *for some constant* $0 < \epsilon < 1$.

Assumption 2. *The degrees of freedom parameter* ν *of the Student's t density satisfies* $2+\delta < \nu < \infty$ *for some* $\delta > 0$.

Assumption 3. *For* $i = 1, \ldots, N-1$ *and* $j = i+1, \ldots, N$, *let*

$$\mathbb{E}\left[\log\max\left(\left|\beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} \cdot b_t\right| \;,\; \left|\beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} \cdot (1-\epsilon^2) \cdot b_t\right|\right)\right] < 0, \tag{12}$$

for $b_t = \epsilon^2 \cdot (\frac{1}{2}(\nu_{i,j|L_{ij}} + 2)\tilde{b}_t - 1)$, *and* $\tilde{b}_t$ *an i.i.d. sequence of Beta*$(2, \nu_{i,j|L_{ij}})$ *distributed random variables.*

Assumption 1 is common in the literature on dynamic conditional correlation models. It can be found in for instance Harvey (2013), or Blasques et al. (2018). In our case, it ensures that the partial correlations are never equal to ± 1, such that the correlation matrix $\boldsymbol{R}_t$ implied by the partial correlations is always (strictly) positive definite. Assumption 2 is a standard moment condition that is needed for second moments (and thus the correlation matrix) to exist. If we choose to model a scaling matrix instead, this assumption can be further relaxed to the existence of an arbitrarily small moment. Assumption 3 formulates a sufficient condition for ensuring that the recursions for $\hat{f}_{i,j|L_{ij};t}$ are contracting on average. This in turn allows us to apply Theorem 3.1 of Bougerol (1993) and conclude stationarity and ergodicity properties of the model as a DGP. The restrictions on the parameter space imposed by equation (12) can easily be checked numerically for specific values of $\alpha_{i,j|L_{ij}}$, $\beta_{i,j|L_{ij}}$, and $\nu_{i,j|L_{ij}}$.

Using the above assumptions, we can now prove the following proposition.

Proposition 2 (strict stationarity and ergodicity). *Let Assumptions 1–3 hold true. Let $\hat{\boldsymbol{R}}_1$ denote a fixed initial correlation matrix with implied partial correlations $\hat{\rho}_{i,j|L_{ij};1}$ and their transforms $\hat{f}_{i,j|L_{ij};1}$. Then, the solutions $\hat{f}_{i,j|L_{ij};t}$ of model (5)–(9) for $t \in \mathbb{N}$, initialized at $\hat{f}_{i,j|L_{ij};1}$ for $i = 1, \ldots, N-1$, $j = i+1, \ldots, N$, converge e.a.s. to unique strictly stationary and ergodic solutions $\{f_{i,j|L_{ij};t}\}_{t\in\mathbb{Z}}$. In addition, the (initialized) partial correlations $\hat{\rho}_{i,j|L_{ij};t} = g(\hat{f}_{i,j|L_{ij};t})$ and the Pearson correlations $\hat{\rho}_{i,j;t}$ converge e.a.s. to their unique stationary and ergodic limits $\{\rho_{i,j|L_{ij};t}\}_{t\in\mathbb{Z}} = \{g(f_{i,j|L_{ij};t})\}_{t\in\mathbb{Z}}$ and $\{\rho_{i,j;t}\}_{t\in\mathbb{Z}}$.*

3.2 Filter invertibility

Naturally, the true time-varying partial correlation processes $\{\rho_{i,j|L_{ij};t}\}_{t\in\mathbb{Z}} = \{g(f_{i,j|L_{ij};t})\}_{t\in\mathbb{Z}}$ are unobserved. However, due to the observation-driven nature of the model, we can easily replace them by their initialized filtered counterparts $\{\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\}_{t=1}^{T} = \{g(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}))\}_{t=1}^{T}$, where we add the argument $\boldsymbol{\theta}$ to the notation to indicate that the filter is evaluated at an arbitrary $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. These filtered partial correlations map into the filtered Pearson correlation matrices via equation (4).

To study the asymptotic properties of the MLE $\hat{\boldsymbol{\theta}}_T$, we need first to study the stochastic limit properties of the filtered processes $\{\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\}_{t=1}^{T}$, since the likelihood function depends both on the data and on these filtered processes. The appropriate convergence result for the filter is known in the literature as filter invertibility; see Straumann and Mikosch (2006), Wintenberger (2013), and Blasques et al. (2018). A complication in our setting is that *all* partial correlations are needed to construct the full correlation matrix. This is important, as unlike Wintenberger (2013) or Blasques et al. (2018, 2022) which also deal with nonlinear filtering methods for time-varying parameters, we cannot rely on standard contraction theorems such as the Bougerol's Theorem 3.1. The novelty in the result below lies in the fact that we show that the multivariate convergence follows easily from the individual univariate convergence results for the pairwise partial correlation filters based on bivariate data slices. This provides a substantial simplification of the proof. To accomplish this, we lean on the theory for perturbed stochastic recurrence equations (SREs) of Straumann and Mikosch (2006, Theorem 2.10) using a sequence of cascading SREs.

To formulate the result, we first introduce some further notation. We define the demeaned and standardized bivariate observation vectors $\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as

$$\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) = \boldsymbol{D}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})^{-1/2}\left(\boldsymbol{y}_{i,j\,;t} - \boldsymbol{\mu}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right), \tag{13}$$

with $\boldsymbol{\mu}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ and $\boldsymbol{D}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as defined in equations (6) and (7), respectively. These standardized observations make up the main input of the bivariate conditional Student's t distributions in (5). Note that $\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ not only depends on $\boldsymbol{y}_{i,t}$ and $\boldsymbol{y}_{j,t}$, but also on the pairwise correlations as gathered in $\boldsymbol{R}_{i,L_{ij};t}$ and $\boldsymbol{R}_{L_{ij},L_{ij};t}$, which have been estimated in a previous step of the cascade. We therefore also introduce the perturbed counterparts $\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ of $\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$, where we replace the elements of $\boldsymbol{R}_{i,L_{ij};t}$ and $\boldsymbol{R}_{L_{ij},L_{ij};t}$ in (13) by those of $\hat{\boldsymbol{R}}_{i,L_{ij};t}$ and $\hat{\boldsymbol{R}}_{L_{ij},L_{ij};t}$, respectively. We also distinguish three different filtered sequences: (i) the filter sequence $\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$, initialized at $\hat{f}_{i,j|L_{ij}\,;1}$ an taking $\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as inputs; (ii) the filter sequence $\hat{\hat{f}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$, initialized at the same $\hat{f}_{i,j|L_{ij}\,;1}$ but taking the stationary and ergodic $\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as inputs; and (iii) the sequence $f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$, denoting the uninitialized stationary and ergodic limiting filter that takes $\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as inputs. The first of these three is the one that is actually computed in empirical applications via the MLE procedure and is available to the user.

To formulate our proposition, we make the following assumption.[5]

Assumption 4. *The set $\boldsymbol{\Theta} \subset \mathbb{R}^d$ is a compact parameter space satisfying $\nu \geq 2 + \delta$ for some $\delta > 0$ and $\alpha_{i,j|L_{ij}} \neq 0$ for $i = 1, \ldots, N-1$ and $j = i+1, \ldots, N$, with*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \log\left|\beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot \frac{\partial s_{i,j|L_{ij}\,;t}\left(f, \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f}\right|\right] < 0. \tag{14}$$

Assumption 4 ensures that the initialized filter is contracting on average when taking the unperturbed $\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ as inputs, i.e., $\hat{\hat{f}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) \xrightarrow{e.a.s.} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$. An approach based on $\hat{\hat{f}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ is, however, infeasible: the MLE procedure can only use the perturbed $\hat{\boldsymbol{y}}^{\star}(\boldsymbol{\theta})$ based on all previously filtered pairs of (initialized) correlation estimates. Therefore, the empirical procedure produces $\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$ rather than $\hat{\hat{f}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})$. Only for $j - i = 1$ we observe $\boldsymbol{y}^{\star}(\boldsymbol{\theta})$ directly because

[5] In fact, Assumption 4 may be further relaxed by replacing the supremum over $\boldsymbol{\theta}$ in (14) by a supremum over $(\omega_{i,j|L_{ij}}, \alpha_{i,j|L_{ij}}, \beta_{i,j|L_{ij}}, \nu)$ only. In order not to overburden the (already heavy) notation further, we opt for the current simpler but more restrictive formulation.

$\boldsymbol{\mu}_{i,i+1|L_{ij};t}(\boldsymbol{\theta}) = 0$ and $\boldsymbol{D}_{i,i+1|L_{ij};t}(\boldsymbol{\theta}) = (1-2\nu^{-1})\mathbf{I}_2$. For $j-i=k>1$, however, the score recursions for the filter also use the initialized sequence $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ for $j-i=1,\ldots,k-1$. The latter are not stationary and ergodic, which prevents us from applying Bougerol (1993) as it requires stationary and ergodic inputs.

The way out of this challenge is as follows. If the filters $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ for $j-i<k$ converge exponentially fast and almost surely to their stationary and ergodic limits $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$, then we can use the results on perturbed SREs from Straumann and Mikosch (2006). In particular, under condition (14) the desired filter invertibility for $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ can then still be established for $j-i=1,\ldots,N-1$. The composite procedure boils down to the following. Starting from $j-i=1$, we recursively obtain invertibility for $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}$ for all $j-i=k=1,\ldots,N-1$. Finally, by standard continuity arguments, we conclude that filter invertibility holds for the pairwise conditional correlation coefficients $\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and the Pearson correlation matrices $\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})$. We summarize this in the following proposition.

Proposition 3 (filter invertibility). *Let Assumptions 1–4 hold true. Then, the filter processes $\{\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ initialized at fixed values $\hat{f}_{i,j|L_{ij};1}$ converge exponentially fast almost surely to the unique stationary and ergodic sequences $\{f_{i,j|L_{ij};t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ uniformly over the parameter space $\boldsymbol{\Theta}$, that is*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right| \xrightarrow{e.a.s.} 0,$$
$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right| \xrightarrow{e.a.s.} 0,$$
$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\hat{\rho}_{i,j;t}(\boldsymbol{\theta}) - \rho_{i,j;t}(\boldsymbol{\theta})\right| \xrightarrow{e.a.s.} 0,$$

as $t\to\infty$.

As a result of proposition, the impact of starting values for the filters becomes negligible asymptotically. In Appendix B we show that this result extends to the derivative processes of $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. These derivative processes play a crucial role for proving the asymptotic normality of the MLE. Filter invertibility in the multivariate model thus simplifies for the setting at hand to a sequence of univariate invertibility conditions, which are much easier to handle.

3.3 Consistency and asymptotic normality of the MLE

Our approach to establish strong consistency and asymptotic normality of the maximum likelihood estimator (MLE) for our dynamic partial correlation model relies on similar arguments as discussed in Straumann and Mikosch (2006) and Blasques et al. (2022). The idea consists in first showing that the nonstationary average log-likelihood function $T^{-1}\,\hat{L}_T(\boldsymbol{\theta})$ in (10) converges to its stationary counterpart

$T^{-1} L_T(\boldsymbol{\theta})$, which uses $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$ rather than $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$. We can then apply the uniform strong law of large numbers for stationary and ergodic processes of Rao (1962) to show that $T^{-1} L_T(\boldsymbol{\theta}) \to \mathbb{E}[\ell_t(\boldsymbol{\theta})]$ almost surely and uniformly over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The strong consistency of $\hat{\boldsymbol{\theta}}_T$ then follows by checking standard identifiability arguments. The result is stated in the following theorem.

Theorem 1. *Under Assumptions 1–4, $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ for every fixed set of starting values $\hat{f}_{i,j|L_{ij};1} \in \mathbb{R}$ for the filter for $i = 1, \dots, N-1$ and $j = i+1 \dots, N$.*

To establish the asymptotic normality of the MLE, the following two additional assumptions are needed, which are rather standard in the literature.

Assumption 5. *$\boldsymbol{\theta}_0 \in$ interior$(\boldsymbol{\Theta})$, i.e., the true parameter vector $\boldsymbol{\theta}_0$ lies in the interior of the (compact) parameter space $\boldsymbol{\Theta}$.*

Assumption 6. *For some $\delta > 0$, it holds that*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \left| \beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot \frac{\partial s_{i,j|L_{ij};t}\left(f, \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f} \right|^{2+\delta}\right] < 1. \tag{15}$$

Assumption 5 excludes situations where the true parameter lies on the boundary of the parameter space. Assumption 6 in addition requires that the score-driven filters and their derivative processes have second moments. This allows us to appeal to an appropriate central limiting result. Combining these assumptions, we obtain the following theorem, which is proved in the appendix.

Theorem 2. *Under Assumptions 1–6, and for every fixed set of starting values for the filter, $\hat{f}_{i,j|L_{ij};1} \in \mathbb{R}$ for $i = 1, \dots, N-1$ and $j = i+1 \dots, N$, we have $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \Rightarrow \mathbb{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}_0))$, where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is the Fisher information matrix evaluated at the true parameter vector $\boldsymbol{\theta}_0$.*

4 Empirical Application

4.1 Data and benchmark models

In this section we apply our model to portfolio returns of US stocks. We consider daily data from January 3, 1980 to December 31, 2021. The data are obtained from Ken French's website.[6] We consider three risk factors and 12 portfolio returns. Combining the three risk factors and the industry portfolio returns, we can test the different correlation models in economic terms from an asset pricing perspective as in Boudt et al. (2017). The three risk factors are the excess market factor Mkt-RF,

[6] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table 1: Descriptive statistics of the daily returns of the three main risk factors and the twelve US industry portfolios over the full data period, 03 January 1980 to 31 December 2021.

Series	Mean	Max	Min	Std	Skewness	Kurtosis
Mkt-RF	0.0349	15.7600	-12.0100	1.1536	0.2017	19.7633
SMB	0.0023	8.1800	-7.2700	0.6542	-0.5691	23.3088
HML	0.0157	9.0400	-6.0200	0.6796	1.0057	20.7701
NoDur	0.0341	13.9700	-9.2500	0.8607	0.2158	25.0626
Durl	0.0583	27.0600	-16.7000	1.7385	0.6772	19.3277
Manuf	0.0434	23.4000	-11.4900	1.4425	0.6376	23.1950
Enrgy	0.0420	17.2900	-9.6100	1.2444	0.4425	15.5661
Chems	0.0493	18.5000	-19.1100	1.2864	0.0229	25.6146
BusEq	0.0528	22.3900	-16.7500	1.6460	0.3410	18.1912
Telcm	0.0369	15.9800	-12.8800	0.9383	0.6995	29.5993
Utils	0.0377	17.9200	-15.2600	1.3417	0.3484	21.1585
Shops	0.0386	17.8600	-11.7900	1.1285	0.2282	21.2056
Health	0.0442	12.6200	-14.4000	1.1300	-0.1520	24.8621
Money	0.0381	19.7100	-17.2300	1.3972	0.3297	26.2926
Other	0.0331	17.5800	-11.1600	1.3642	0.3269	16.9240

the size factor SMB (Small Minus Big), and the value factor HML (High Minus Low). The 12 industry portfolio returns are for non-durables (NoDur), durables (Durbl), manufacturing (Manuf), energy (Enrgy), chemicals (Chems), business equipment (BusEq), telecom (Telcm), utilities (Utils), Shops, Health, Money, and Other. The return series are shown in Figure 1, and descriptive statistics are reported in Table 1. The results clearly reveal the standard stylized facts of high kurtosis and clear volatility clustering. The fat-tailedness warrants the use of the Student's t distribution for the analysis.

In our remaining analysis, we label the new dynamic partial correlation model as *PCorr*. Next to it, we consider two proven benchmarks: (i) the multivariate GAS$(1,1)$ model of Creal et al. (2011) (labeled t-GAS), where the correlation matrix is modeled using the hypersphere parameterization, and (ii) the cDCC model of Engle (2002) and Aielli (2013) endowed with a Student's t distribution and labeled t-cDCC. For the t-cDCC model, we use the standard targeting approach to estimate the (matrix-valued) intercept parameter of the correlation transition equation. For the partial correlation model and the matrix t-GAS model such a targeting is not available, and we estimate the intercept terms as part of the static parameter vector using standard numerical optimization. We also note that our partial correlation model has pair-specific parameters $\alpha_{i,j|L_{ij}}$ and $\beta_{i,j|L_{ij}}$, unlike the standard versions of the t-cDCC and matrix t-GAS. The latter typically only use a common scalar α and β. To put the different models on a more equal footing, we introduce the same number of pair-specific $\alpha_{i,j}$ and $\beta_{i,j}$ into the t-GAS model. This can be done without further complications due to the hypersphere parameterization in the t-GAS. For the t-cDCC, we impose a BEKK type specification with diagonal

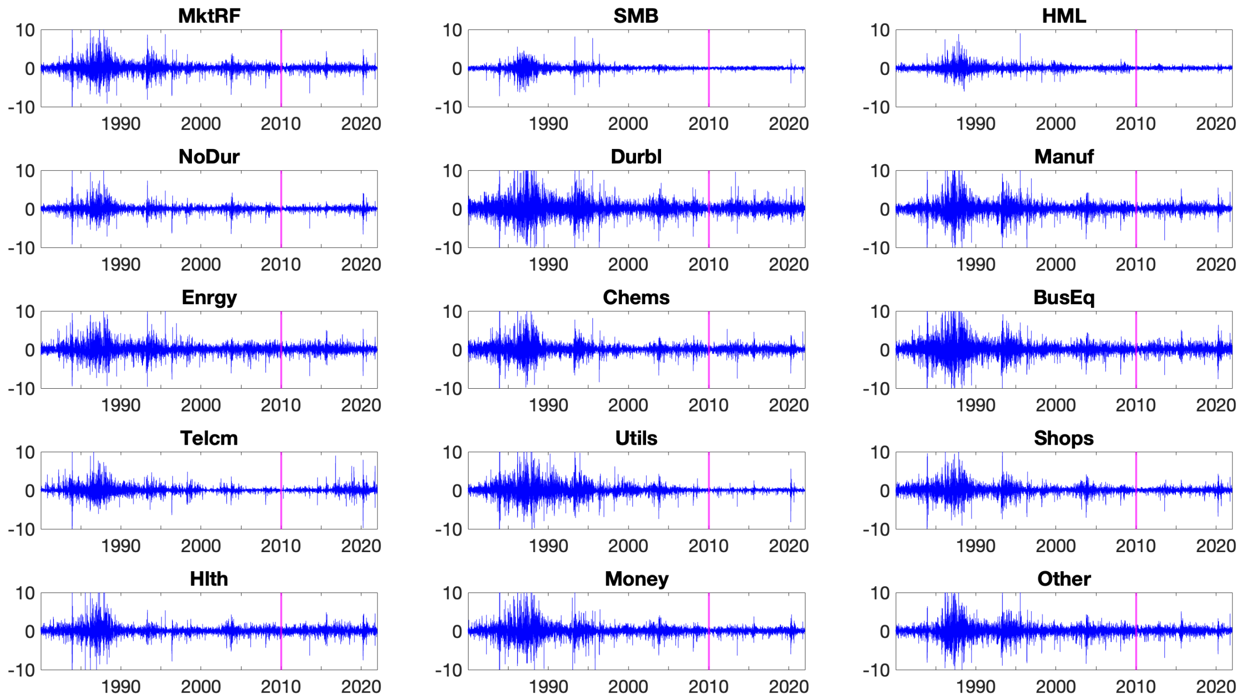


Figure 1: Daily returns on the three main risk factors and the twelve industry portfolios
Note: The period is 03 January 1980 to 31 December 2021. The vertical lines indicate the 4th of January 2010, i.e. the first trading day of 2010 and the start of the out-of-sample period.

A and B matrices holding N parameters $\alpha_{i,i}$ and $\beta_{i,i}$, respectively. This ensures positive definiteness of the correlation matrix at all times.[7]

To fully concentrate on the differences in modeling correlations, we first de-volatilize all return series using the score-driven volatility models of Creal et al. (2011, 2013) based on the Student's t distribution, also known as the Beta-t-GARCH(1,1) model of Harvey (2013). The de-volatilized series are then used as inputs for the correlation-based models. All correlation models thus work with the same input series, such that any differences cannot be attributed to differences in univariate volatility filters.

4.2 Simulation Results

To investigate the performance of the new model, we first use a controlled simulation setting. We simulate series $\{\boldsymbol{y}_t\}_{t=1}^{T}$ of $T = 1,000$ observations from a multivariate conditional Gaussian or Student's t distribution. The dimension of the time series equals $N = 4$ as in the empirical application later on. To generate time series with empirically relevant correlation dynamics, we use a 100-day rolling window to estimate time-varying empirical correlation matrices for the the four series (HML, SMB,

[7] In particular, we use $Q_t^\star = \Omega + B^{1/2} Q_{t-1}^\star B^{1/2} + A^{1/2} \boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}^\top A^{1/2}$, where A and B are diagonal matrices with parameters $\alpha_{i,i}$ and $\beta_{i,i}$, respectively.

Table 2: *MSE*, *MAE* and Frobenious norm simulation results for three different correlation models

Note: the labels PCorr, t-GAS and t-DCC indicate the new score-driven partial correlation model discussed in Section 2, the Student's t GAS model of Creal et al. (2011) with hypersphere parameterization, and the t-cDCC model of Engle (2002) with a multivariate Student's t log-likelihood, respectively. Results are based on 300 Monte Carlo experiments with sample size $T = 1000$ and $N = 4$. True correlation paths used in the data generating process are given from 100-day rolling window estimates of empirical correlation matrices of the series (HML, SMB, Mkt - RF, BusEq).

	MSE	*MAE*	*Frobenius*	*MSE*	*MAE*	*Frobenius*
	Gaussian			**Student t_7**		
PCorr	**0.0174**	**0.1036**	**0.4285**	**0.0192**	**0.1106**	**0.4543**
t-GAS	0.0264	0.1094	0.4324	0.0222	0.1204	0.4898
t-cDCC	0.0268	0.1177	0.4474	0.0273	0.1303	0.5386

Mkt-RF, BusEq), where BusEq (business equipment) denotes the return one of the sector portfolios. These rolling window estimates produce paths for $4(4-1)/2 = 6$ different pairwise correlations. We fix these paths and then generate 300 realizations of the returns $\boldsymbol{y}_t$ based on these (empirical) correlation matrices and either a Gaussian or a Student's t ($\nu = 7$) distribution, where the latter is close to the empirical estimate. For each of the simulated return series, we estimate the new time-varying partial correlation model as well as the benchmark models.

To compare the performance of the different models, we consider the mean squared error (*MSE*), the mean absolute error (*MAE*) and the Frobenius norm (*Frobenius*), where

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(\hat{\rho}_{i,j\,;\,t} - \rho_{i,j\,;\,t}\right)^2, \qquad MAE = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left|\hat{\rho}_{i,j\,;\,t} - \rho_{i,j\,;\,t}\right|,$$

and *Frobenius* $= \|\hat{\boldsymbol{R}}_t - \boldsymbol{R}_t\| = MSE^{1/2}$, where $\hat{\rho}_{i,j\,;\,t}$ and $\hat{\boldsymbol{R}}_t$ denote the filtered paths of the conditional (Pearson) correlation coefficients and the (Pearson) correlation matrix. As all three measures consider the performance of the correlation models in terms of pairwise Pearson correlations rather than partial correlations, the benchmark models (which all operate on the Pearson correlations directly) are put at an advantage compared to our new model (which operates on the partial correlations).

The results in Table 2 present a clear outcome: in both the Gaussian and Student's t case and across all three performance measures, the new partial correlation model outperforms the two benchmarks, followed by the t-GAS model and then the t-cDCC. For instance, for the empirically more relevant Student's t case, the MAE and Frobenius norm are around 10% lower for the new partial correlation model than for the t-GAS, and around 20% lower than for the t-cDCC. The improvements appear realistic: we expect all three models to do reasonably well for typical stock return series. This is confirmed by the filtered correlation paths in Figure 2. The black pattern gives the true path of the (six) correlations used in the simulations. We see that most of the time, the different correlation models follow each other quite closely. However, there are also marked differences such as for $\rho_{2,3,t}$.

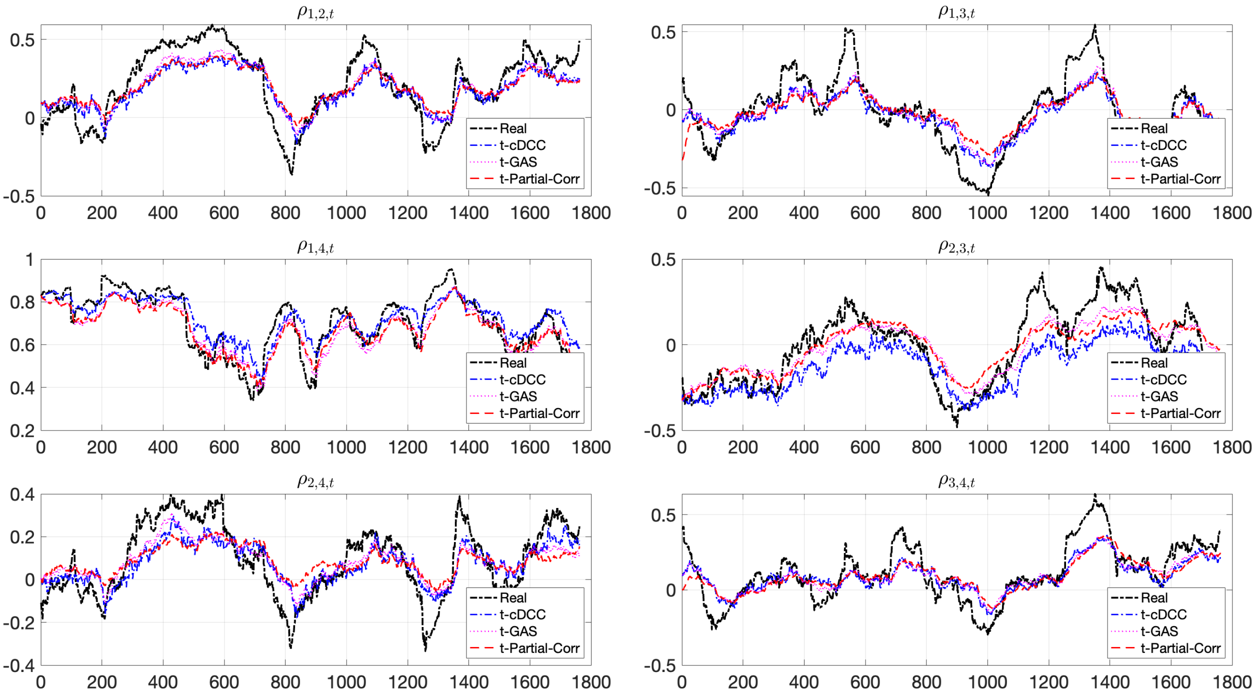


Figure 2: Comparison of the mean of the Monte Carlo simulation of the filtered conditional correlation coefficients with Student's t DGP with $\nu = 7$.

It turns out from Table 2 that on average the partial correlation model appears to do a better job in such cases where the different models produce different results. In the next subsections we continue to investigate this for the empirical rather than the simulated data.

4.3 In-sample analysis

For the empirical data, we first compare the different models based on an in-sample analysis. We estimate each model over the in-sample period 1980–2009. In the next section we then consider an out-of-sample analysis over the period 2010–2021.

Table 3 holds the differences in log-likelihood values between the different models. The left panel compares the PCorr model versus the t-GAS, whereas the right panel compares the PCorr and the t-cDCC. In all cases we take the PCorr model as the benchmark, such that positive values in the log-likelihood column signal that the new model outperforms the benchmark. The results clearly show that the PCorr model always outperforms the t-cDCC for each of the 12 industries. In comparison with the t-GAS, the new model also fits better in 8 out of the 12 industries, performs less well in only 2 cases, and at par in 2 others. Most gains are in the range of 20–50 likelihood points for PCorr versus t-GAS. Improvements are even higher at 60–110 likelihood points when comparing the PCorr versus t-cDCC model. Note that these results hold despite the fact that we made the t-GAS and

Table 3: In sample performance of the three correlation models

Note: the PCorr, t-GAS, and t-cDCC models are estimated over the sample 03 January 1980 to 31 December 2009. The log-Lik indicates the differences in log-likelihood value at the optimum. Diebold-Mariano t statistics are reported based on the MSE and MAE criterion, related to the differences in mean squared and mean absolute pricing errors, $e_t = r_{i,t} - \hat{\gamma}_{Mkt,t}(r_t^{Mkt} - r_t^F) - \hat{\gamma}_{SMB,t}SMB_t - \hat{\gamma}_{HML,t}HML_t$, where all return series are volatilty filtered, and $\hat{\gamma}_{Mkt/SMB/HML,t}$ is obtained as in (17). The MCS columns indicate whether the model is selected for the 95% model confidence set based on MSE. For the MAE criterion, the GAS and DCC model are only selected for Utils.

	PCorr versus t-GAS			PCorr versus t-cDCC			MCS		
	log-Lik	DM_{MSE}	DM_{MAE}	log-Lik	DM_{MSE}	DM_{MAE}	PCorr	GAS	DCC
NoDur	21.8 *	-6.88***	-7.69 ***	91.7 ***	-6.05***	-7.78 ***	✓		
Durl	31.9 **	-9.21***	-10.20***	84.6 ***	-7.85*	-8.71 **	✓		
Manuf	21.9 *	-9.70***	-11.44***	64.2 ***	-7.85***	-10.26***	✓		
Enrgy	-1.0 *	-4.47***	-4.84 ***	88.4 ***	-4.45***	-4.08 ***	✓		
Chems	40.6 *	-6.68*	-7.01 ***	94.1 ***	-6.15***	-7.15 ***	✓		
BusEq	-46.7***	-3.62***	-5.84 ***	64.1 ***	-2.47**	-5.91 ***	✓	✓	✓
Telcm	32.5 ***	-7.28***	-7.94 ***	110.9***	-7.81***	-7.10 ***	✓		
Utils	0.1	-1.41	-2.87 ***	87.0 ***	-1.08	-2.92 ***	✓	✓	✓
Shops	28.8 **	-7.21***	-9.00 ***	83.6 ***	-7.03***	-8.37 ***	✓		
Health	23.7 *	-5.90***	-6.23 **	91.7 ***	-5.31***	-6.68 ***	✓		
Money	-30.4*	-5.66***	-8.51 ***	76.5 ***	-5.73***	-9.99 ***	✓		
Other	50.3 ***	-8.61*	-10.01***	89.6 ***	-7.36***	-8.02 ***	✓		

t-cDCC models more flexible by endowing them with different $\alpha_{i,j}$ and $\beta_{i,j}$ parameters for different return pairs as compared to their standard scalar form from the literature.

Besides evaluating the models in terms of likelihood fit, we also compare the models in terms of their asset pricing implications as in Hansen et al. (2014), Boudt et al. (2017), and Darolles et al. (2018). For this, we consider the tracking errors

$$e_t = r_{i,t} - \gamma_{Mkt,t}(r_t^{Mkt} - r_t^F) - \gamma_{SMB,t}SMB_t - \gamma_{HML,t}HML_t, \tag{16}$$

where $r_{i,t}$ denotes the return on one of the $i = 1, \ldots, 12$ industry portfolios, and

$$\begin{pmatrix} \gamma_{Mkt,t} \\ \gamma_{SMB,t} \\ \gamma_{HML,t} \end{pmatrix} = \begin{pmatrix} \rho_{Mkt,Mkt,t} & \rho_{Mkt,SMB,t} & \rho_{Mkt,HML,t} \\ \rho_{SMB,Mkt,t} & \rho_{SMB,SMB,t} & \rho_{SMB,HML,t} \\ \rho_{HML,Mkt,t} & \rho_{HML,SMB,t} & \rho_{HML,HML,t} \end{pmatrix}^{-1} \begin{pmatrix} \rho_{Mkt,i,t} \\ \rho_{SMB,i,t} \\ \rho_{HML,i,t} \end{pmatrix}. \tag{17}$$

As all models considered are observation driven, the $\rho_{i,j\,;\,t}$ and thus also the $\beta_{i,j\,;\,t}$ are known at time $t-1$. Remember that all returns enter the tracking error equation after being de-volatilized in order to fully concentrate on the differences due to correlation modeling.

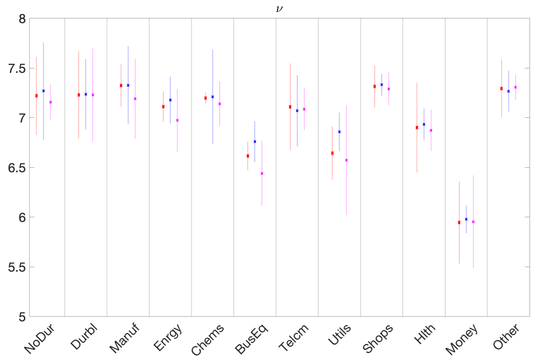
Table 3 also holds the results on the tracking error MSE and MAE by means of the Diebold-Mariano t-test statistics. Negative values indicate that the PCorr model outperforms the benchmark.

The results confirm the earlier log-likelihood analysis. We see that also in terms of the asset pricing implications of the models, the PCorr model outperforms both the t-GAS and t-cDCC. Interestingly, for a few industries the statistical log-likelihood criterion and the economic performance criteria based on tracking error MSE and MAE point to a different model ranking. Based on the economic criteria, the overall picture even seems clearer and more robust: the PCorr model outperforms the benchmarks in-sample. In terms of MAE, the outperformance is unanimous and strongly statistically significant across all industries. For MSE, it is significant for 9 (or 11) out of 12 industries at the 1% (or 10%) significance level. To account for the cross-sectional correlation of the different tests and the possibly inflated type I error due to multiple pairwise tests, we also compute the model confidence set of Hansen et al. (2011) based on the MSE criterion. We see that the PCorr model is always in the 95% model confidence set. The t-GAS and t-cDCC models, by contrast, only enter the model confidence set for two industries, and even then less often than the PCorr model over different bootstrap runs. For the model confidence set based on the MAE (not shown) the pattern is similar: whereas the PCorr model is in the model confidence set for each of the 12 industries, the t-GAS and t-cDCC model enter only for one industry (Utils).

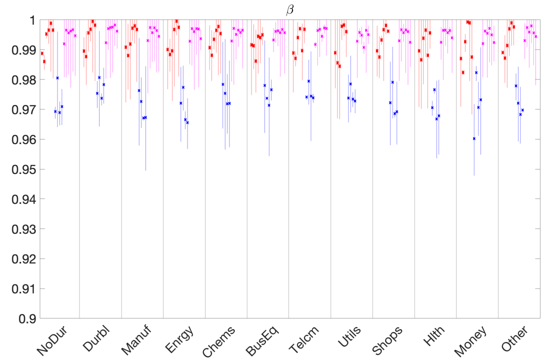
As we have 3 models with 19 parameters each, estimated across 12 industries, we have estimated almost 700 parameters in total and their standard errors. Rather than reporting all these results in a table, we visualize the estimation outcomes in Figure 3, each line indicating a point estimate and its 95% confidence interval. The figure reveals three main findings. First, as shown in Figure 3a, despite the use of several different starting values and 12 different industries, the estimates of ν are relatively stable between 6 and 7.5, indicating a realistic, moderate degree of fat-tailedness.

Second, Figure 3b indicates that the (partial) correlations in all models have a high degree of persistence: all $\beta_{i,j|L_{ij}}^{PCorr}$, $\beta_{i,i}^{DCC}$, and $\beta_{i,j}^{GAS}$ parameters are close to one across all industries and models. The estimates for the t-cDCC appear slightly lower, but one should measure persistence for the t-cDCC using a composite of $\alpha_{i,i}$ and $\beta_{i,i}$ rather than $\beta_{i,j}$ alone as in the PCorr and t-GAS model. Adding $\alpha_{i,i}$ and $\beta_{i,i}$ together, the estimates for persistence are again close across all models.

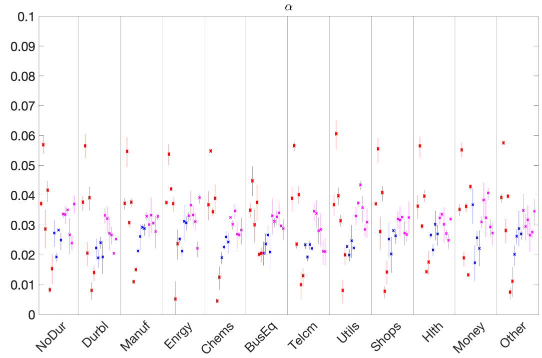
Finally, when looking at the $\alpha_{i,j|L_{ij}}^{PCorr}$, $\alpha_{i,i}^{DCC}$, and $\alpha_{i,j}^{GAS}$ parameters, we see that virtually all of these are statistically significant. It is interesting to see that the adjustment speeds for the t-GAS and t-cDCC appear much more homogeneous than those of the PCorr model. In particular the adjustment speed of SMB with the market return (MKT) is much higher in the PCorr model. The partial correlation between the industry returns and SMB given the market return $(IND, SMB \mid MKT)$, as well as that between the industry return and HML given SMB and the market return $(IND, HML \mid MKT, SMB)$ are both substantially lower. Such heterogeneity can easily be allowed for in the PCorr model. This is more complicated in t-GAS model, which scrambles this linkage between $\alpha_{i,j}$ and



(a) MLE estimates of ν



(b) MLE estimates of $\beta^{PCorr}_{i,j|L_{ij}}$, β^{DCC}_{i}, and $\beta^{GAS}_{i,j}$



(c) MLE estimates of $\alpha^{PCorr}_{i,j|L_{ij}}$, α^{DCC}_{i}, and $\alpha^{GAS}_{i,j}$

Figure 3: Parameter estimates of all correlation models across industries

Note: the top-left panel for $\alpha_{i,j}$ has 12 vertical areas, each corresponding to an industry. The red left six lines in each band provide the parameter estimates (as a point) and their confidence intervals (as a line) for the PCorr model in the order of our decomposition L_{ij}, i.e., $(i,j) = (SMB, HML), (MKT, SMB), (IND, MKT), (MKT, HML \mid SMB), (IND, SMB \mid MKT), (IND, HML \mid MKT, SMB)$ which indexes along each lower sub-diagonal of $\boldsymbol{R}_t$, starting from the first sub-diagonal. The next 4 blue lines indicate the estimates and confidence intervals for the t-cDCC model, followed by the estimates of of the t-GAS model (in the same order as for the PCorr model). The plots for $\beta_{i,j}$ and ν are structured similarly. For ν we only have one estimate per model per industry.

$\rho_{i,j}$ via the hypersphere re-parameterization. As a result, the heterogeneity in $\alpha_{i,j}$ is much less for the t-GAS model. Given the different estimation results between the different correlation models, we investigate the predictive implications of these differences within an asset pricing context in the next subsection.

4.4 Out-of-sample analysis

In our out-of-sample analysis, we fully focus on the tracking errors defined in (16), similar to Hansen et al. (2014), Boudt et al. (2017), and Darolles et al. (2018). We perform a recursive out-of-sample analysis. First, we estimate all models on the in-sample period 1980–2009. We then fix the static parameter estimates and run the filter up to 2010 to obtain the one year out-of-sample model-implied correlation matrices $\boldsymbol{R}_t$ as well as the implied coefficients $\gamma_{Mkt,t}$, $\gamma_{SMB,t}$, and $\gamma_{HML,t}$ from equations (16)–(17). These result in predicted returns $\hat{r}_{i,t}$ (conditional on the risk factors) and the corresponding tracking errors. After obtaining the tracking errors for 2010, we then add 2010 to the sample, and re-estimate the model over 1980–2010 to obtain tracking errors for 2011. We repeat this process up till the last year in the sample, giving us 2978 tracking errors.

To evaluate the forecasts, we run the regressions

$$r_{i,t} = a_0^{Mod} + a_1^{Mod}\, \hat{r}_{i,t}^{Mod} + u_{i,t}, \tag{18}$$

where $Mod \in \{\text{PCorr}, t\text{-GAS}, t\text{-cDCC}\}$. We then test the null hypothesis $H_0 : a_0^{Mod} = 0, a_1^{Mod} = 1$ using the suitable heteroskedasticity and autocorrelation consistent (HAC) estimator for the covariance matrix of the regression parameters as suggested by for instance White (1980) and MacKinnon and White (1985). In addition, we implement the Model Confidence Set (MCS) procedure developed by Hansen et al. (2011) to select the model with the smallest tracking error MSE. The results are shown in Table 4.

Also in the out-of-sample analysis, the results clearly point towards the PCorr model. The model is always contained in the model confidence set for each industry, whereas the t-GAS and t-cDCC models each only enter the model confidence set once. We also see that the null hypothesis of $a_0^{Mod} = 0$ and $a_1^{Mod} = 1$ is most rejected for the t-cDCC model with 10 out of 12 cases, followed by the t-GAS with 5 cases, and the the PCorr with only 3 out of 12 cases. We attribute this to the flexibility of the PCorr model to adapt itself to each (partial) correlation separately, with a robust propagation system due to the use of the Student's t distribution and score-driven dynamics. The t-GAS shares the robust score-driven propagation of the PCorr, but lacks the direct link to each (partial) correlation due to the complex hypersphere transformation. The t-cDCC, on the other hand, retains the direct link to the pairwise correlations, but lacks the robust propagation mechanism. From our in-sample and out-of-sample analysis, it appears that both properties of the PCorr model are useful for typical empirical data.

In Figure 4 we plot the results for the different $\gamma_{j,t}$ parameters for $j = MKT, SMB, HML$. The figure shows that though the secular movements of the three models align, there can also be substantial episodes where the time-varying parameters differ between models. The γs in this industry for the PCorr model appear somewhat smoother than for the two benchmark models. As argued in Francq and Zakoian (2019), this may help the usability of the model where γs are typically believed to be more stable over time and not change heavily from one day to another.

Table 4: Out-of-sample results

This table contains the estimates of a_1^{Mod} for $Mod \in \{PCorr, t-GAS, t-cDCC\}$ in the regression model $r_{i,t} = a_0^{Mod} + a_1^{Mod} \cdot \hat{r}_{i,t}^{Mod} + u_{i,t}$, where $\hat{r}_{i,t}^{Mod}$ is obtained (recursively) using one-year-ahead estimates of $\boldsymbol{R}_t$ and $\gamma_{MKT,t}$, $\gamma_{SMB,t}$, and $\gamma_{HML,t}$ as in (17). A $*$, $**$, or $***$ indicates rejection of $H_0 : a_0^{Mod} = 0, a_1^{Mod} = 1$, at the 10%, 5%, and 1% significance level, respectively. The MCS column indicates whether the model lies in the 95% model confidence set of Hansen et al. (2011) based on tracking error MSE. Results are similar for the 99% MCS.

	PCorr		t-GAS		t-cDCC	
	$\hat{a}_1^{PCorr}$	MCS	$\hat{a}_1^{t-GAS}$	MCS	$\hat{a}_1^{t-cDCC}$	MCS
NoDur	1.013	✓	0.987		0.966 ***	
	(0.013)		(0.013)		(0.013)	
Durbl	1.018	✓	0.956 **		0.984	
	(0.013)		(0.013)		(0.012)	
Manuf	1.012	✓	1.001		0.975 ***	
	(0.007)		(0.007)		(0.007)	
Enrgy	1.053 **	✓	1.005		0.967 ***	
	(0.023)		(0.015)		(0.013)	
Chems	1.002	✓	0.981		0.965 ***	
	(0.011)		(0.012)		(0.011)	
BusEq	0.913 ***	✓	0.886 ***		0.861 ***	
	(0.006)		(0.007)		(0.006)	
Telcm	1.000	✓	0.975		0.945 ***	
	(0.014)		(0.014)		(0.013)	
Utils	1.050	✓	0.957		0.990	✓
	(0.032)		(0.023)		(0.020)	
Shops	0.987	✓	0.983		0.946 ***	
	(0.009)		(0.009)		(0.009)	
Health	1.009	✓	0.996 *		0.958 ***	
	(0.011)		(0.012)		(0.010)	
Money	0.986 *	✓	0.982 **	✓	0.928 ***	
	(0.006)		(0.006)		(0.007)	
Other	1.011	✓	1.010 *		0.974 ***	
	(0.006)		(0.006)		(0.006)	

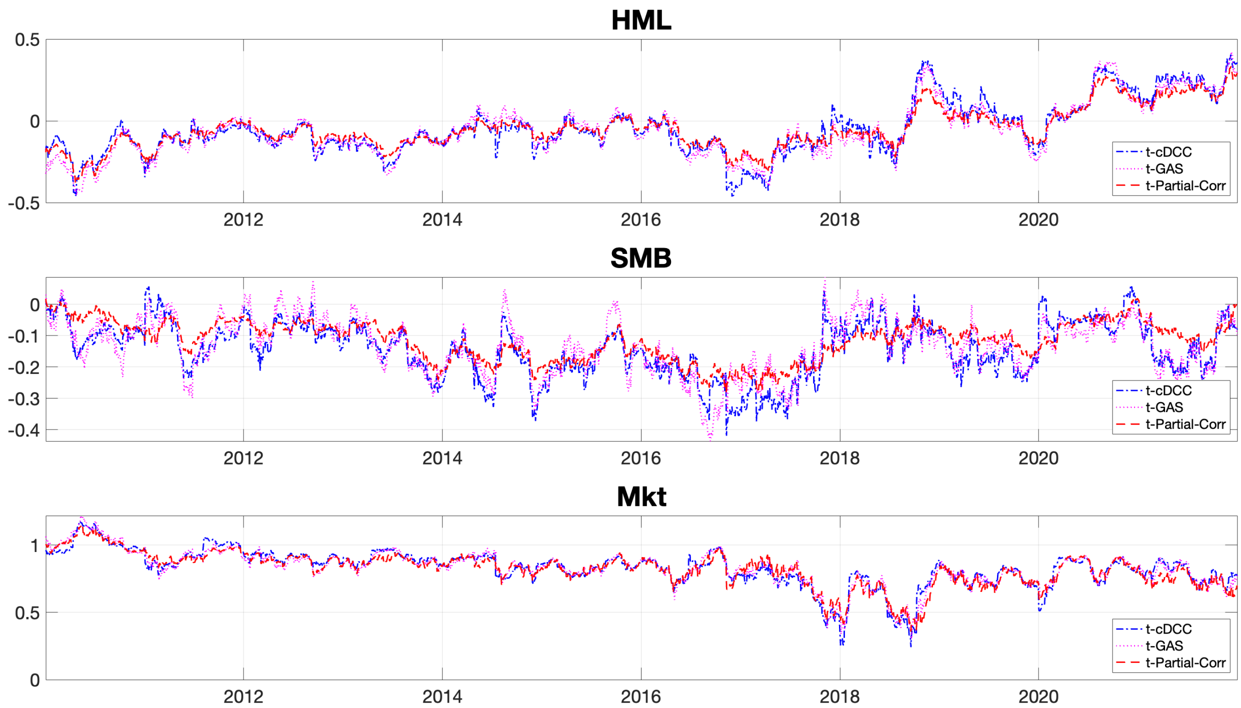


Figure 4: Recursive one-step-ahead forecasts of the conditional betas of *NoDur* (model re-estimated annually)

5 Conclusions

In this paper we introduced a recursive model for correlation matrix dynamics based on partial correlations and score-driven dynamics. The model's structure provided flexibility and interpretability, without loosing computational tractability. In addition, the recursive structure ensured stationarity and ergodicity as well as filter invertibility for any fixed dimension. The conditions needed remained of similar complexity as in the univariate time-varying correlation setting on bivariate data slices. To prove this, we used the approach of perturbed stochastic recurrence equations applied to our cascade of bivariate (conditional) models. Estimation of the full multivariate model could be carried out by straightforward maximum likelihood methods. Finally, using the stationarity and invertibility properties of the model and its filter, we were also able to prove consistency and asymptotic normality of the maximum likelihood estimator.

Both in simulations and in an in-sample and out-of-sample application to US industry stock returns, the new model was shown to outperform benchmarks such as the Student's t based cDCC and multivariate volatility GAS models.

The model also provides interesting directions for future research. For instance, though in the current paper we pursued joint estimation of all static model parameters simultaneously, the model lends itself well to a recursive (bivariate) estimation strategy that could substantially speed up computations. In addition, this recursive set-up is easily scalable to the high-dimensional setting. This, however, might impact the asymptotic properties of the model and the filter and the conditions needed to establish consistency and asymptotic normality. It would be worthwhile to investigate these issues in a future paper, both from a computational and theoretical perspective.

References

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

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis.*

Archakov, I. and P. R. Hansen (2021). A new parametrization of correlation matrices. *Econometrica 89*(4), 1699–1715.

Bauwens, L., C. Hafner, and S. Laurent (2012). *Handbook of Volatility Models and Their Applications.* Wiley Handbooks in Financial Engineering and Econometrics. John Wiley & Sons.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Blasques, F., P. Gorgi, S. J. Koopman, and O. Wintenberger (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics 12*, 1019–1052.

Blasques, F., A. Lucas, and E. Silde (2018). A stochastic recurrence equations approach for score driven correlation models. *Econometric Reviews 37*(2), 166–181.

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics 227*(2), 325–346.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews 11*(2), 143–172.

Boudt, K., S. Laurent, A. Lunde, R. Quaedvlieg, and O. Sauri (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics 196*(2), 347–367.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Buccheri, G., G. Bormetti, F. Corsi, and F. Lillo (2021). A score-driven conditional correlation model for noisy and asynchronous data: An application to high-frequency covariance dynamics. *Journal of Business & Economic Statistics 39*(4), 920–936.

Comte, F. and O. Lieberman (2003). Asymptotic theory for multivariate garch processes. *Journal of Multivariate Analysis 84*(1), 61–84.

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

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

Darolles, S., C. Francq, and S. Laurent (2018). Asymptotics of cholesky garch models and time-varying conditional betas. *Journal of Econometrics 204*(2), 223–247.

Ding, P. (2016). On the conditional distribution of the multivariate t distribution. *The American Statistician 70*(3), 293–295.

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

Engle, R. F. (2016). Dynamic conditional beta. *Journal of Financial Econometrics 14*(4), 643–667.

Francq, C. and Zakoian (2019). *GARCH Models, 2nd Edition*. John Wiley & Sons.

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

Hafner, C. M. and L. Wang (2021). A dynamic conditional score model for the log correlation matrix. *Journal of Econometrics*.

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

Hansen, P. R., A. Lunde, and V. Voev (2014). Realized beta garch: A multivariate garch model with realized measures of volatility. *Journal of Applied Econometrics 29*(5), 774–799.

Harvey, A. C. (2013). *Dynamic models for Volatility and Heavy Tails*. Cambridge University Press.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis 97*(10), 2177–2189.

MacKinnon, J. G. and H. White (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics 29*(3), 305–325.

Magnus, J. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2018). New heavy models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics 36*(4), 643–657.

Opschoor, A., A. Lucas, I. Barra, and D. Van Dijk (2021). Closed-form multi-factor copula models with observation-driven dynamic factor loadings. *Journal of Business & Economic Statistics 39*(4), 1066–1079.

Pedersen, R. S. and A. Rahbek (2014). Multivariate variance targeting in the bekk-garch model. *The Econometrics Journal 17*(1), 24–55.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics 33*(2), 659–680.

Rapisarda, F., D. Brigo, and F. Mercurio (2007). Parameterizing correlations: a geometric interpretation. *IMA Journal of Management Mathematics 18*(1), 55–73.

Roth, M. (2013). On the multivariate t distribution. Technical report from Automatic Control at Linköpings Universitet LiTH-ISY-R-3059, Department of Electrical Engineering, Linköpings Universitet.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Tse, Y. K. and A. K. C. Tsui (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics 20*(3), 351–362.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

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

Wintenberger, O. (2013). Continuous invertibility and stable qml estimation of the egarch(1,1) model. *Scandinavian Journal of Statistics 40*(4), 846–867.

Appendix to: Dynamic Partial Correlation Models

Enzo D'Innocenzo[a], André Lucas[a,b]

[a] *Vrije Universiteit Amsterdam*

[b] *Tinbergen Institute*

A Proofs

Proof of Proposition 1

Define

$$
\begin{aligned}
\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t} &= \hat{\boldsymbol{D}}^{-1/2}_{i,j|L_{ij}\,;t}\left(\boldsymbol{y}_{i,j\,;t} - \hat{\boldsymbol{\mu}}_{i,j|L_{ij}\,;t}\right), \\
\hat{\boldsymbol{D}}_{i,j|L_{ij}\,;t} &= \frac{(\nu-2)(\nu + \boldsymbol{y}^{\top}_{L_{ij},t}\hat{\boldsymbol{R}}^{-1}_{L_{ij},L_{ij};t}\boldsymbol{y}_{L_{ij},t})}{\nu \cdot \nu_{i,j|L_{ij}}}\begin{pmatrix} \hat{\boldsymbol{V}}_{i,i|L_{ij}\,;t} & 0 \\ 0 & \hat{\boldsymbol{V}}_{j,j|L_{ij}\,;t} \end{pmatrix}, \\
w_{i,j|L_{ij}\,;t} &= \frac{\nu_{i,j|L_{ij}} + 2}{\nu_{i,j|L_{ij}} + \hat{\boldsymbol{y}}^{\star\top}_{i,j|L_{ij}\,;t}\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}}.
\end{aligned}
$$

Using standard vector derivative calculus, we have

$$
\begin{aligned}
&\frac{\partial \log p(\boldsymbol{y}_{i,t}, \boldsymbol{y}_{j,t} \mid \mathcal{F}_{t-1}, \boldsymbol{y}_{L_{ij},t})}{\partial f_{i,j|L_{ij}\,;t}} = \\
&\quad \frac{\partial}{\partial f_{i,j|L_{ij}\,;t}}\left(-\tfrac{1}{2}\log|\boldsymbol{R}_{i,j|L_{ij}\,;t}| - \tfrac{1}{2}(\nu_{i,j|L_{ij}}+2)\log\left(1 + \frac{\hat{\boldsymbol{y}}^{\star\top}_{i,j\,;t}\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}\hat{\boldsymbol{y}}^{\star}_{i,j\,;t}}{\nu_{i,j|L_{ij}}}\right)\right) \\
&\quad = \tfrac{1}{2}\frac{\partial\,\mathrm{vec}(\boldsymbol{R}_{i,j|L_{ij}\,;t})^{\top}}{\partial f_{i,j|L_{ij}\,;t}} \cdot \mathrm{vec}\left(w_{i,j|L_{ij}\,;t}\cdot \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}\hat{\boldsymbol{y}}^{\star\top}_{i,j|L_{ij}\,;t}\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t} - \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}\right) \\
&\quad = \tfrac{1}{2}\frac{\partial\,\mathrm{vec}(\boldsymbol{R}_{i,j|L_{ij}\,;t})^{\top}}{\partial f_{i,j|L_{ij}\,;t}} \cdot \left(\boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}\otimes \boldsymbol{R}^{-1}_{i,j|L_{ij}\,;t}\right)\mathrm{vec}\left(w_{i,j|L_{ij}\,;t}\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}\hat{\boldsymbol{y}}^{\star\top}_{i,j|L_{ij}\,;t} - \boldsymbol{R}_{i,j|L_{ij}\,;t}\right).
\end{aligned}
$$

■

Proof of Proposition 2

Throughout the proof we take the symmetric matrix root.[8] Note that we have

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \tfrac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

$$\boldsymbol{R}^{-1/2} = \tfrac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} (1+\rho)^{-1/2} & 0 \\ 0 & (1-\rho)^{-1/2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \cdot \left(\boldsymbol{R}^{-1/2} \otimes \boldsymbol{R}^{-1/2} \right) = \frac{\begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix}}{1-\rho^2}.$$

Define

$$w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} = \frac{\nu_{i,j|L_{ij}} + 2}{\nu_{i,j|L_{ij}} + \boldsymbol{\eta}^\top_{i,j|L_{ij};t} \boldsymbol{\eta}_{i,j|L_{ij};t}},$$

and

$$\begin{aligned} s^{\boldsymbol{\eta}}_{i,j|L_{ij};t} &= \tfrac{1}{2} \dot{g}(f_{i,j|L_{ij};t}) \cdot \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \cdot \left(\boldsymbol{R}^{-1/2}_{i,j|L_{ij};t} \otimes \boldsymbol{R}^{-1/2}_{i,j|L_{ij};t} \right) \cdot \operatorname{vec}\left(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \cdot \boldsymbol{\eta}_{i,j|L_{ij};t} \boldsymbol{\eta}^\top_{i,j|L_{ij};t} - \mathbf{I}_2 \right) \\ &= \tfrac{1}{2} \epsilon \cdot \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix} \cdot \operatorname{vec}\left(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \cdot \boldsymbol{\eta}_{i,j|L_{ij};t} \boldsymbol{\eta}^\top_{i,j|L_{ij};t} - \mathbf{I}_2 \right) \\ &= \epsilon \cdot \left(w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_{1,t} \boldsymbol{\eta}_{2,t} - \tfrac{1}{2} g(f_{i,j|L_{ij};t})\ w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t + g(f_{i,j|L_{ij};t}) \right) \end{aligned}$$

where for the second equality we used Assumption 1 and thus $\dot{g}(f) = \epsilon \cdot (1 - g(f)^2) = \epsilon \cdot (1-\rho^2)$.

By Assumption 2, using the model as a Data Generating Process (DGP), that is, $\boldsymbol{y}_{i,j|L_{ij};t} = \boldsymbol{R}^{1/2}_{i,j|L_{ij};t} \boldsymbol{\eta}_t$, we can rewrite the score-driven transition equation of Proposition 1 under the DGP as

$$f_{i,j|L_{ij};t+1} = \omega_{i,j|L_{ij}} + \beta_{i,j|L_{ij}} f_{i,j|L_{ij};t} + \alpha_{i,j|L_{ij}} s^{\boldsymbol{\eta}}_{i,j|L_{ij};t}.$$

Note that for given $f_{i,j|L_{ij};t}$ all moments of $w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top$ and thus of the rewritten score $s^{\boldsymbol{\eta}}_{i,j|L_{ij};t}$ exist due to its uniform boundedness in $\boldsymbol{\eta}_t$. As a result, for a fixed initialization $f_{i,j|L_{ij};1}$ we directly obtain $\mathbb{E}[\log^+ |f_{i,j|L_{ij};t+1}|\ ] < \infty$. To use Theorem 3.1 of Bougerol (1993), we therefore only need to prove

[8] Other roots can be taken as well, but typically indicate smaller stationarity regions; compare Blasques et al. (2018). Note that the Bougerol (1993) condition only provides a sufficient condition, such that we are free to take the matrix root that results in the widest region.

that the recursion is contracting on average. To do this, we note

$$\frac{\partial f_{i,j|L_{ij};t+1}}{\partial f_{i,j|L_{ij};t}} = \beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} \cdot \epsilon^2 \cdot \left(1 - g(f_{i,j|L_{ij};t})^2\right) \cdot \left(\tfrac{1}{2} w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t - 1\right)$$

such that we require

$$\begin{aligned} \mathbb{E}\left[\log \sup_f \left| \frac{\partial f_{i,j|L_{ij};t+1}}{\partial f_{i,j|L_{ij};t}} \Big|_{f_{i,j|L_{ij};t}=f} \right|\right] &= \\ \mathbb{E}\left[\log \sup_f \left| \beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} \cdot \epsilon^2 \cdot \left(1 - g(f)^2\right) \cdot \left(\tfrac{1}{2} w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t - 1\right)\right|\right] &< 0. \end{aligned}$$

Define $\tilde{b_t} = \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t / (\nu_{i,j|L_{ij}} + \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t)$ and $b_t = \epsilon^2 \cdot (\frac{1}{2}(\nu_{i,j|L_{ij}} + 2)\tilde{b_t} - 1)$, and note that $\tilde{b_t}$ has a $\mathrm{Beta}(2, \nu_{i,j|L_{ij}})$ distribution. Given the specification of $g(f) = \epsilon \cdot \arctan(f)$, the supremum over f inside the expectation is reached at either $g(f) = 0$ or $g(f) = \epsilon$. Inserting all this inside the above expectation, the required condition simplifies to

$$\begin{aligned} \mathbb{E}\left[\log \sup_f \left| \beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} \cdot \epsilon^2 \cdot \left(1 - g(f)^2\right) \cdot \left(\tfrac{1}{2} w^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \boldsymbol{\eta}_t^\top \boldsymbol{\eta}_t - 1\right)\right|\right] &= \\ \mathbb{E}\left[\log \sup_f \left| \beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} \cdot \left(1 - g(f)^2\right) \cdot b_t \right|\right] &= \\ \mathbb{E}\left[\log \max\left(\left|\beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}} b_t\right| \ , \ \left|\beta_{i,j|L_{ij}} - \alpha_{i,j|L_{ij}}(1-\epsilon^2) b_t\right|\right)\right] &< 0. \end{aligned}$$

This is clearly satisfied through Assumption 3. Theorem 3.1 of Bougerol (1993) now implies that each initialized $f_{i,j|L_{ij};t}$ converges (e.a.s.) to a unique stationary and ergodic limit sequence.

As the mappings $\rho_{i,j|L_{ij};t} = g(f_{i,j|L_{ij};t})$ are all continuously differentiable with $\sup_f \dot{g}(f) = \epsilon$, we obtain $|\hat{\rho}_{i,j|L_{ij};t} - \rho_{i,j|L_{ij};t}| = |g(\hat{f}_{i,j|L_{ij};t}) - g(f_{i,j|L_{ij};t})| \leq \epsilon \cdot |\hat{f}_{i,j|L_{ij};t} - f_{i,j|L_{ij};t}|$, such that the e.a.s. convergence of $\hat{\rho}_{i,j|L_{ij};t}$ follows directly from that of $\hat{f}_{i,j|L_{ij};t}$.

The e.a.s. convergence of $\hat{\boldsymbol{R}}_t$ follows similarly by combining the e.a.s. convergence of $\hat{\rho}_{i,j|L_{ij};t}$, the properties of the mapping from $\rho_{i,j|L_{ij};t}$ into $\boldsymbol{R}_t$, and the fact that under Assumption 1 the correlation matrices $\boldsymbol{R}_t$ and their filtered equivalents $\hat{\boldsymbol{R}}_t$ are never singular. ∎

Proof of Proposition 3

Under the maintained assumptions, we can apply Proposition 2 to conclude that $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic. Using Assumption 1 and thus $g(f_{i,j|L_{ij};t}) = \epsilon \cdot \arctan(f_{i,j|L_{ij};t})$, we have

$$\begin{aligned}&\tfrac{1}{2}\hat{\boldsymbol{G}}_{i,j|L_{ij};t}\boldsymbol{\mathcal{D}}_2^\top\left(\hat{\boldsymbol{R}}_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\hat{\boldsymbol{R}}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^{-1} = \\ &\qquad \tfrac{1}{2}\,\frac{\epsilon}{1-\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2}\cdot\left(-2\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ 1+\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\ ,\ 1+\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\ ,\ -2\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right).\end{aligned}$$

We define $\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})$ as

$$\begin{aligned}\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \hat{\boldsymbol{D}}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{-1/2}\left(\boldsymbol{y}_{i,j;t}-\hat{\boldsymbol{\mu}}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right),\\ \hat{\boldsymbol{D}}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{(\nu-2)(\nu+\boldsymbol{y}_{L_{ij},t}^\top\hat{\boldsymbol{R}}^{-1}(\boldsymbol{\theta})_{L_{ij},L_{ij};t}\boldsymbol{y}_{L_{ij},t})}{\nu\cdot\nu_{i,j|L_{ij}}}\begin{pmatrix}\hat{\boldsymbol{V}}_{i,i|L_{ij};t}(\boldsymbol{\theta}) & 0\\ 0 & \hat{\boldsymbol{V}}_{j,j|L_{ij};t}(\boldsymbol{\theta})\end{pmatrix},\end{aligned}$$

as a perturbed bivariate data sequence, with $\hat{\boldsymbol{V}}_{i,i;t}(\boldsymbol{\theta}) = 1-\hat{\boldsymbol{R}}_{i,L_{ij};t}(\boldsymbol{\theta})\cdot\hat{\boldsymbol{R}}^{-1}_{L_{ij},L_{ij};t}(\boldsymbol{\theta})\cdot\hat{\boldsymbol{R}}_{L_{ij},i;t}(\boldsymbol{\theta})$. The perturbation is due to the initialization of the filter sequences. We can write the initialized filter recursions

$$\hat{f}_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) = \phi\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}),\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) = \omega_{i,j|L_{ij}}+\beta_{i,j|L_{ij}}\cdot\hat{f}_{i,j|L_{ij};t+1}(\boldsymbol{\theta})+\alpha_{i,j|L_{ij}}\cdot s_{i,j|L_{ij};t}(\boldsymbol{\theta}), \tag{A.1}$$

$$\begin{aligned}s_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}),\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta});\boldsymbol{\theta}\right) \qquad \text{(A.2)}\\ &= \frac{\epsilon}{1-\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2}\cdot\Bigg(\\ &\qquad\left(1+\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(\hat{w}_{i,j|L_{ij};t}\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})-\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\\ &\qquad-\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(\hat{w}_{i,j|L_{ij};t}\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})-2\right)\Bigg),\\ \hat{w}_{i,j|L_{ij};t} &= \hat{w}_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}),\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta});\boldsymbol{\theta}\right)\\ &= \frac{\nu_{i,j|L_{ij}}+2}{\nu_{i,j|L_{ij}}+\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\hat{\boldsymbol{R}}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\hat{\boldsymbol{y}}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}.\end{aligned}$$

Also, we note that the contraction condition in equation (14) of Assumption 4 entails the following derivative

$$\begin{aligned}s^f_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}),\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta});\boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})} = \frac{2\epsilon\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})}{1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2}\cdot\Bigg(\qquad \text{(A.3)}\\ \left(1+\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(w_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})-\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\end{aligned}$$

$$
\begin{aligned}
&- \rho_{i,j|L_{ij};t}(\boldsymbol{\theta}) \left(\hat{w}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2\right) \Bigg)\\
&+ \epsilon \cdot \Bigg(\left(2\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\right)\left(w_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\\
&+ \left(1+\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(w^{f}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\right)\\
&- \left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\right)\left(w_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2\right)\\
&- \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2\left(w^{f}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
w^{f}_{i,j|L_{ij};t} &= \frac{\partial w_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})} \qquad \text{(A.4)}\\
&= \frac{\partial \operatorname{vec}(\boldsymbol{R}_{i,j|L_{ij};t})^{\top}}{\partial f_{i,j|L_{ij};t}} \cdot \frac{\nu_{i,j|L_{ij}}+2}{\left(\nu_{i,j|L_{ij}} + \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^2}\\
&\qquad \left(\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \otimes \boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\left(\boldsymbol{y}^{\star}_{i,j|L_{ij};t} \otimes \boldsymbol{y}^{\star}_{i,j|L_{ij};t}\right),
\end{aligned}
$$

evaluated at some fixed point $f_{i,j|L_{ij};t}(\boldsymbol{\theta}) = f$.

We first *assume* that $\sup_{\theta\in\Theta} |\hat{\boldsymbol{y}}^{\star}_{L_{ij},L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{L_{ij},L_{ij};t}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$, i.e., that $\hat{\boldsymbol{y}}^{\star}_{L_{ij},L_{ij};t}(\boldsymbol{\theta})$ converges uniformly e.a.s. to a unique stationary and ergodic limit $\boldsymbol{y}^{\star}_{L_{ij},L_{ij};t}(\boldsymbol{\theta})$, and then prove the e.a.s. convergence of $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ to $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and the existence of a log moment. The complete result then follows by induction after starting the recursion at $i-j=1$ and noting that

$$
\begin{aligned}
&\hat{\boldsymbol{y}}^{\star}_{i,i+1|L_{ij};t}(\boldsymbol{\theta}) = \boldsymbol{y}^{\star}_{i,i+1|L_{ij};t}(\boldsymbol{\theta}) = \hat{\boldsymbol{D}}_{i,i+1|L_{ij};t}(\boldsymbol{\theta})^{-1/2}\boldsymbol{y}_{i,i+1;t} = \sqrt{\frac{(\nu-2)}{\nu}} \cdot \boldsymbol{y}_{i,i+1;t},\\
&\hat{\boldsymbol{D}}_{i,i+1|L_{ij};t}(\boldsymbol{\theta}) = \boldsymbol{D}_{i,i+1|L_{ij};t}(\boldsymbol{\theta}) = \frac{(\nu-2)}{\nu} \cdot \mathbf{I}_2,
\end{aligned}
$$

where $\hat{\boldsymbol{y}}^{\star}_{i,i+1;t}(\boldsymbol{\theta})$ is obviously stationary and ergodic due to the stationarity of $\boldsymbol{y}_{i,i+1;t}$.

For the remainder of the proof, we thus assume $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{\boldsymbol{y}}^{\star}_{L_{ij},L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{L_{ij},L_{ij};t}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$. If we consider the filter recursion in (A.1) using the uninitialized stationary and ergodic $\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ rather than the perturbed $\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, we can easily see that a log moment exists for a fixed $\hat{f}_{i,j|L_{ij};1}$:

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log^{+}\left|\phi\left(\hat{f}_{i,j|L_{ij};1}, \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) - \hat{f}_{i,j|L_{ij};1}\right|\right] \leq\\
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log^{+}\left(\left|\omega_{i,j|L_{ij}}\right| + \left|\beta_{i,j|L_{ij}} - 1\right| \cdot \left|\hat{f}_{i,j|L_{ij};1}\right| + \left|\alpha_{i,j|L_{ij}}\right| \cdot \frac{K_1\epsilon}{1-\epsilon^2},\right) \leq K_2 < \infty,
\end{aligned}
$$

where the K_i denote finite positive constants, and where we have used the uniform boundedness of the filtered $|\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})| \leq \epsilon$ via Assumption 1, as well as the uniform boundedness of the score expression $s_{i,j|L_{ij};t}(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta})$ in $\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ due to the analytical form of the filtered weights

$\hat{w}_{i,j|L_{ij};t}(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta})$.

Additionally, by similar arguments, we have that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f} \log^{+}\left|\frac{\partial s_{i,j|L_{ij};t}\left(f, \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f}\right|\right] \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log^{+}\left(\frac{K_1 2\epsilon^2}{1-\epsilon^2} + K_2\epsilon\right) \leq K_3 < \infty.$$

The e.a.s. convergence of the filter that takes $\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ as input to a unique stationary and ergodic limit then follows by Theorem 3.1 of Bougerol (1993) if we can prove that the filtering equation is contracting on average. This, however, follows immediately from Assumption 4.

The last part of the proof consists in showing that the perturbed filter recursions converge to the same limits as their unperturbed counterparts. Following Theorem 2.10 of Straumann and Mikosch (2006), this follows by showing

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left| s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};1}\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right) - s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};1}\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right| \xrightarrow{e.a.s.} 0, \quad \text{(A.5)}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{f}\left|\frac{\partial s_{i,j|L_{ij};t}\left(f\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f} - \frac{\partial s_{i,j|L_{ij};t}\left(f\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f}\right| \xrightarrow{e.a.s.} 0, \quad \text{(A.6)}$$

as $t \to \infty$.

To prove (A.5), note that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \hat{\hat{f}}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right| \leq K \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|,$$

where $K < \infty$ according to technical Lemma 3. Since $K < \infty$ and thus $\mathbb{E}[\log^{+} K] < \infty$, the desired convergence on the left hand side in (A.5) follows as an application of Lemma 2.1 of Straumann and Mikosch (2006) and the assumed e.a.s. convergence of $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})|$.

To prove (A.6), we note that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial s_{i,j|L_{ij};t}\left(f\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f} - \frac{\partial s_{i,j|L_{ij};t}\left(f\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f}\right| \leq$$
$$K \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|,$$

where $K < \infty$ according to technical Lemma 3. Similar as for (A.5), the result then follows as an application of Lemma 2.1 of Straumann and Mikosch (2006).

We can now conclude that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - f_{i,j|L_{ij};t}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ for all $i = 2, \ldots, N$ and

$j = 1, \ldots, i-1$.

To conclude the e.a.s. convergence of $\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ to its limiting process, note that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right| &= \epsilon \cdot \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\tanh(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})) - \tanh(f_{i,j|L_{ij};t}(\boldsymbol{\theta}))\right| \\ &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|,\end{aligned}$$

where the inequality follows by taking a first order Taylor series expansion.

To conclude the e.a.s. convergence of $\hat{\rho}_{i,j;t}(\boldsymbol{\theta})$ to its limiting process, note that for $i = j+1$ we have $\hat{\rho}_{i,j;t}(\boldsymbol{\theta}) = \hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, such that the result follows directly from the e.a.s. convergence of the partial correlation. For $i > j+1$, the result then follows by induction. Note that from (4) we have

$$\hat{\rho}_{i,j;t} = \hat{\boldsymbol{R}}_{i,L_{ij};t}\hat{\boldsymbol{R}}^{-1}_{L_{ij},L_{ij};t}\hat{\boldsymbol{R}}_{L_{ij},j;t} + \hat{\rho}_{i,j|L_{ij};t}\sqrt{\hat{\boldsymbol{V}}_{i,i|L_{ij};t} \cdot \hat{\boldsymbol{V}}_{j,j|L_{ij};t}}, \tag{A.7}$$

where $\hat{\boldsymbol{R}}_{L_{ij},L_{ij};t}$ is never singular due to Assumption 1 and $\epsilon < 1$. This mapping is a series of products and sums of elements of $\boldsymbol{R}_{L_{ij},L_{ij};t}$ and $\boldsymbol{R}_{i,L_{ij};t}$, each term of which converges e.a.s. to its limiting process by a direct application of Lemma 2.1 of Straumann and Mikosch (2006) and Lemma TA.16 of Blasques et al. (2022). for $i > j+1$. ∎

Proof of Theorem 1

By the triangle inequality, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})]\right| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta})\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}L_T(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})]\right|. \tag{A.8}$$

To show that the first term converges almost surely to zero, we write

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta})\right| \leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta})\right|, \tag{A.9}$$

and then note that by the Cesàro mean, the first term on the right hand side of inequality (A.8) converges to zero almost surely if $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$.

We have

$$\begin{aligned}&2 \cdot (\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta})) \\ &= \log|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})| - \log|\boldsymbol{R}_t(\boldsymbol{\theta})| + (\nu+N)\left[\log\left(1 + \frac{\boldsymbol{y}_t^\top\hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right) - \log\left(1 + \frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right]\end{aligned}$$

$$\leq tr\Big(\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\Big) + \frac{\nu+N}{\nu-2}\boldsymbol{y}_t^\top \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})\Big(\boldsymbol{R}_t(\boldsymbol{\theta}) - \hat{\boldsymbol{R}}_t(\boldsymbol{\theta})\Big)\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t,$$

where the inequality follows from Theorem 11.27 in Magnus and Neudecker (2019), Lemma A.1 of Bollerslev and Wooldridge (1992) and the standard log-inequality $\log(1+x) \leq x \ \forall x > -1$. Due to Assumption 1 with $\epsilon < 1$ and the mapping between partial and Pearson correlations, we automatically have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\right\| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \lambda_1^{-1}(\boldsymbol{R}_t(\boldsymbol{\theta})) < K, \tag{A.10}$$

for some $0 < K < \infty$

For any $N \times N$ matrix $\boldsymbol{A}$ it holds that $\operatorname{tr}\boldsymbol{A} \leq N \cdot \|\boldsymbol{A}\|$. We also have $\nu > 2$ by Assumption 2, while from Proposition 3, we obtain

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta})\right| &\leq N\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\right\| + K_1\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})\right\|\|\boldsymbol{y}_t\|^2 \\
&\leq N\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\gamma_{i,j}^{-t}\right) + K_1\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\gamma_{i,j}^{-t}\right)\|\boldsymbol{y}_t\|^2,
\end{aligned}$$

for some $K, c > 0$ by following similar arguments as in Hafner and Preminger (2009). Since $\gamma_{i,j} > 1$ and $\mathbb{E}[\|\boldsymbol{y}_t\|^2] < \infty$, we obtain the desired almost sure convergence

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{1}{T}\hat{L}_T(\boldsymbol{\theta}) - \frac{1}{T}L_T(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0,$$

as $T \to \infty$, by a straightforward application of the Markov's inequality and the Borel-Cantelli Lemma.

To prove the almost sure convergence of the second term on the right hand side of inequality (A.8), we only need to show that $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta})|] < \infty$ such that we can apply the uniform law of large numbers for stationary and ergodic processes of Rao (1962). Using the expression for the log-likelihood function from equation (10), we have

$$\begin{aligned}
\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\ell_t(\boldsymbol{\theta})\right|\right] \leq &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\log\Gamma\left(\frac{\nu+N}{2}\right)\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\log\frac{\nu}{2}\right| + \frac{N}{2}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\log((\nu-2)\pi)\right| \\
&+ \frac{1}{2}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\log|\boldsymbol{R}_t(\boldsymbol{\theta})|\right] + \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\nu+N}{2}\log\left(1 + \frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|\right] < \infty,
\end{aligned} \tag{A.11}$$

where the last inequality follows as a consequence of Assumptions 2 - 4, the uniform boundedness of $\boldsymbol{R}_t(\boldsymbol{\theta})$ (being a correlation matrix), the uniform lower bound from equation A.10, and the existence of second moments of $\boldsymbol{y}_t$. As a result, we obtain

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{T} L_T(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})] \right| \xrightarrow{a.s.} 0,$$

as $T \to \infty$.

To conclude the proof, we establish identifiability: $\mathbb{E}[\ell_t(\boldsymbol{\theta}_0)] > \mathbb{E}[\ell_t(\boldsymbol{\theta})]\ \forall \boldsymbol{\theta}_0 \neq \boldsymbol{\theta}$. The proof is by contradiction. Assume there is a $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ with $\mathbb{E}[\ell_t(\boldsymbol{\theta})] = \mathbb{E}[\ell_t(\boldsymbol{\theta}_0)]$, where $\mathbb{E}[\ell_t(\boldsymbol{\theta}_0)] < \infty$ by equation (A.11). By Gibb's inequality, this implies that $\nu = \nu_0$ and $\boldsymbol{R}_t(\boldsymbol{\theta}) = \boldsymbol{R}_t(\boldsymbol{\theta}_0)$ almost surely for this specific $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. This, however, leads to a contradiction. We note that there is a one-to-one relationship between the components of the lower (or upper) triangular part of the conditional correlation matrix $\boldsymbol{R}_t(\boldsymbol{\theta})$, and the partial conditional correlations coefficients $\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})$ for $i = 2, \dots, N$, $j = 1, \dots, i-1$. Therefore $\boldsymbol{R}_t(\boldsymbol{\theta}) = \boldsymbol{R}_t(\boldsymbol{\theta}_0)$ (a.s.) implies $\rho_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \rho_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)$ (a.s.). This, however, cannot hold for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, because the equality $\boldsymbol{R}_t(\boldsymbol{\theta}) = \boldsymbol{R}_t(\boldsymbol{\theta}_0)$ entails that

$$\begin{aligned} 0 = f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}_0) - f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) =& \omega_{0,i,j|L_{ij}} - \omega_{i,j|L_{ij}} + (\alpha_{0,i,j|L_{ij}} - \alpha_{i,j|L_{ij}}) s^{\boldsymbol{\eta}}_{i,j|L_{ij};t} \\ & + (\beta_{0,i,j|L_{ij}} - \beta_{i,j|L_{ij}}) f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0), \end{aligned}$$

almost surely. We thus have

$$(\alpha_{0,i,j|L_{ij}} - \alpha_{i,j|L_{ij}}) s^{\boldsymbol{\eta}}_{i,j|L_{ij};t} = \upsilon_{i,j|L_{ij};t}$$

where $\upsilon_{i,j|L_{ij};t}$ is an $\mathcal{F}_t$-measurable random variable. It follows that since the conditional distribution of $\upsilon_{i,j|L_{ij};t}|\mathcal{F}_t$ is not degenerate, it must be that $\alpha_{0,i,j|L_{ij}} = \alpha_{i,j|L_{ij}}$, which yields

$$0 = \omega_{0,i,j|L_{ij}} - \omega_{i,j|L_{ij}} + (\beta_{0,i,j|L_{ij}} - \beta_{i,j|L_{ij}}) f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0).$$

This in turn implies $\omega_{0,i,j|L_{ij}} = \omega_{i,j|L_{ij}}$ and $\beta_{0,i,j|L_{ij}} = \beta_{i,j|L_{ij}}$ by the fact that $f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)$ has a non-degenerate distribution given the non-degenerate distribution of $s^{\boldsymbol{\eta}}_{i,j|L_{ij};t}$ and Assumptions 3 and 4. This contradicts the initial premise $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and thus proves the theorem.

The strong consistency of the the MLE $\hat{\boldsymbol{\theta}}_T$ is then guaranteed by the compactness of the parameter space $\boldsymbol{\Theta}$ and noting that all the conditions of Theorem 3.4 in White (1994) are satisfied. ∎

Proof of Theorem 2

By the strong consistency established in Theorem 1 combined with Assumption 5, we have that the MLE $\hat{\boldsymbol{\theta}}_T$ lies inside an arbitrarily small neighbourhood of $\boldsymbol{\theta}_0$ for sufficiently large T. Using the first order condition for the MLE from (11) and Lemma 7, we obtain

$$\mathbf{0} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \hat{\ell}_t(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_T) + o_p(1),$$

where we note the difference between the log-likelihood functions $\hat{\ell}_t(\hat{\boldsymbol{\theta}}_T)$ and $\ell_t(\hat{\boldsymbol{\theta}}_T)$, the former using the initialized filter, and the latter using its stationary and ergodic limit.

Taking a Taylor expansion, we get

$$o_p(1) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) + \frac{1}{T} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_T^{\star}) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_T^{\star}$ lies between $\hat{\boldsymbol{\theta}}_T$ and the true $\boldsymbol{\theta}_0$. For sufficiently large T we then obtain that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) + o_p(1) = -\frac{1}{T} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_T^{\star}) \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0). \tag{A.12}$$

In Lemma 6, we prove that $T^{-1/2} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)$ obeys the central limit theorem for martingales of Billingsley (1961) and satisfies the Fisher's information matrix equality. Moreover, Lemma 8 ensures that the average $-T^{-1} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_T^{\star})$ converges to the positive definite Fisher's information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$, almost surely. Hence, as $T \to \infty$, by solving equation (A.12), we obtain by the Slutsky's Theorem (see Vaart (1998)) that

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

■

B Technical Lemmas

Define the operators $\nabla^{\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}$ and $\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$, where $\boldsymbol{\theta}$ contains ν, $\omega_{i,j|L_{ij}}$, $\alpha_{i,j|L_{ij}}$, $\beta_{i,j|L_{ij}}$, for $i = 1, \ldots, N-1$ and $j = i+1, \ldots, N$. To avoid ambiguous notations, we also define $\nabla^{\nu} = \frac{\partial}{\partial \nu}$, $\nabla^{\nu\nu} = \frac{\partial^2}{\partial \nu^2}$. We use $\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x)$ to denote the usual digamma function.

Lemma 3. *Consider the score expression and its derivative with respect to $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$*

$$s_{i,j|L_{ij};t}(\boldsymbol{\theta}) = s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ , \ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ ; \ \boldsymbol{\theta}\right),$$

$$s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ , \ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ ; \ \boldsymbol{\theta}\right) \ ; \ \boldsymbol{\theta})}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})},$$

as defined in equations (A.2) *and* (A.3)*, respectively.*

Under Assumption 1, we have that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ , \ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ ; \ \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})} \right| < \infty, \tag{B.13}$$

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ , \ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ ; \ \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \partial \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})} \right| < \infty. \tag{B.14}$$

Proof. By straightforward algebra we get that

$$\frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ , \ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ ; \ \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})} = \frac{\epsilon}{1 - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2} \cdot \Bigg($$

$$\left(1 - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(1 - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\right)\left(w^{\boldsymbol{y}^{\star}}_{i,j|L_{ij};t} \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + 2 w_{i,j|L_{ij};t} \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Bigg)$$

$$w^{\boldsymbol{y}^{\star}}_{i,j|L_{ij};t} = \frac{\partial w_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})}$$

$$= \frac{-2\left(\nu_{i,j|L_{ij}} + 2\right)}{\left(\nu_{i,j|L_{ij}} + \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top \boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^2} \cdot \boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}),$$

and

$$\frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ , \ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \ ; \ \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \partial \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})} = \frac{2\epsilon \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})}{1 - \rho_{i,j|L_{ij};t}(\boldsymbol{\theta})^2} \cdot \Bigg($$

$$
\begin{aligned}
&\Big(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Big)\left(w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})+2w_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Big)\\
&+\epsilon\cdot\Bigg(\Big(\left(2\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})-1\right)\left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Big)\\
&\qquad\qquad\left(w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})+2w_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\\
&+\left(w^{f\boldsymbol{y}}_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})+2w_{i,j|L_{ij};t}\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
w^{f\boldsymbol{y}^\star}_{i,j|L_{ij};t}&=\frac{\partial^2 w_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}),\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta});\boldsymbol{\theta}\right)}{\partial f\partial \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}\\
&=\frac{\partial\operatorname{vec}(\boldsymbol{R}_{i,j|L_{ij};t}(\boldsymbol{\theta}))^\top}{\partial f_{i,j|L_{ij};t}}\cdot\frac{-4(\nu_{i,j|L_{ij}}+2)}{\left(\nu_{i,j|L_{ij}}+\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^2}\\
&\qquad\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\left(\boldsymbol{y}^\star_{i,j|L_{ij};t}\otimes\boldsymbol{y}^\star_{i,j|L_{ij};t}\right)\\
&\quad+\frac{\nu_{i,j|L_{ij}}+2}{\left(\nu_{i,j|L_{ij}}+\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})^\top\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)^2}\\
&\qquad\left(\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\left(\left(\mathbf{I}_2\otimes\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)+\left(\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\otimes\mathbf{I}_2\right)\right).
\end{aligned}
$$

By exploiting the analytical forms of the weights $w_{i,j|L_{ij};t}$, $w^{\boldsymbol{y}^\star}_{i,j|L_{ij};t}$ and $w^{f\boldsymbol{y}^\star}_{i,j|L_{ij};t}$, and the parameterization given in Assumption 1, we can show that the uniform bounds in equations (B.13) and (B.14) are easily satisfied.

In fact, one only needs to note that there exists general positive constants K_1 and K_2, such that

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}\right|\leq\frac{1}{1-\epsilon^2}\Big(\left(1-\epsilon\left(1-\epsilon\right)\right)\left(K_1+2K_2\right)\Big)<\infty,
$$

and also that

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})\partial\boldsymbol{y}^\star_{i,j|L_{ij};t}(\boldsymbol{\theta})}\right|\leq\\
&\frac{2\epsilon^2}{1-\epsilon^2}\Big(\left(1-\epsilon\left(1-\epsilon\right)\right)\left(K_1+2K_2\right)\Big)+\epsilon\cdot\Big(\Big(1+\left(\left(2\epsilon-1\right)\left(1-\epsilon\right)\right)\Big)\left(K_1+2K_2\right)\Big)<\infty,
\end{aligned}
$$

■

Lemma 4. *Under the Assumptions 1-4:*

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})-\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\xrightarrow{e.a.s.}0, \tag{B.15}
$$

as $t \to \infty$.

Furthermore, under the additional Assumption 6, we have

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|^m\right] < \infty, \tag{B.16}$$

for any integer $m \geq 2$.

Proof. As in the proof of Proposition 3, to prove the uniform exponentially fast convergence in (B.15), we can show that the conditions S.1-S.3 of Theorem 2.10 in Straumann and Mikosch (2006) hold true for the first derivative processes

$$\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) = \begin{pmatrix} \nabla^{\omega} f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) \\ \nabla^{\alpha} f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) \\ \nabla^{\beta} f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) \\ \nabla^{\nu} f_{i,j|L_{ij};t+1}(\boldsymbol{\theta}) \end{pmatrix} = \boldsymbol{w}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + X_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \tag{B.17}$$

where

$$X_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \beta_{i,j|L_{ij}} + \alpha_{i,j|L_{ij}} \cdot \frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})}, \tag{B.18}$$

$$\boldsymbol{w}_{i,j|L_{ij};t}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}} \\ \frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}} \\ \frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}}} \\ \frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \nu} \end{pmatrix},$$

such that

$$\frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}} = 1, \qquad \frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}} = s_{i,j|L_{ij};t}(\boldsymbol{\theta}),$$

$$\frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}}} = f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \qquad \frac{\partial f_{i,j|L_{ij};t+1}(\boldsymbol{\theta})}{\partial \nu_{i,j|L_{ij}}} = \alpha_{i,j|L_{ij}} s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}),$$

where the term $s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) = s^{\nu}_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)$ is given by

$$
\begin{aligned}
s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \nu} \qquad \text{(B.19)}\\
&= \frac{\epsilon}{1-\hat{\rho}_{i,j|L_{ij};t}(\boldsymbol{\theta})^2} \cdot \Bigg(\\
&\qquad \left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\left(1-\rho_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\right)\left(w^{\nu}_{i,j|L_{ij};t}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\Bigg).\\
w^{\nu}_{i,j|L_{ij};t} &= \frac{\partial w_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \nu} \qquad \text{(B.20)}\\
&= \frac{\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2}{\left(\nu_{i,j|L_{ij}} + \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})^{\top}\boldsymbol{R}^{-1}_{i,j|L_{ij};t}(\boldsymbol{\theta})\boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - 2\right)^2}.
\end{aligned}
$$

We start by verify conditions S.1 and S.2 in Theorem 2.10 of Straumann and Mikosch (2006), which are directly implied if the following uniform bounds

$$
\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|X_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|\right] < \infty, \qquad \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{w}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right] < \infty.
$$

However, we first note that, by Proposition 3, it holds that $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|X_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|\right] < K_3 < \infty$, and furthermore

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{w}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right] \leq & 1 + \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|s_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|\right] + \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|\right]\\
& + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\alpha_{i,j|L_{ij}}\right|\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right|\right] < \infty,
\end{aligned}
$$

which is again implied by Proposition 3, the compactness of the parameter space, and the fact that

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|s^{\nu}_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)\right| \leq \left(1-\epsilon\left(1-\epsilon\right)\right)\frac{K_1\epsilon}{1-\epsilon^2} \leq K_2 < \infty.
$$

Then, conditions S.1 and S.2 in Theorem 2.10 of Straumann and Mikosch (2006) are directly satisfied.

Now, in the present case, proving that condition S.3 in Theorem 2.10 of Straumann and Mikosch (2006) is equivalent of proving that

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\hat{\boldsymbol{w}}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - \boldsymbol{w}_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\| \xrightarrow{e.a.s.} 0, \qquad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\hat{X}_{i,j|L_{ij};t}(\boldsymbol{\theta}) - X_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\| \xrightarrow{e.a.s.} 0.
$$

By Proposition 3, Lemma 3 and invoking again the mean value theorem, it is immediate to infer that, for a sufficiently large t, we obtain that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\hat{\boldsymbol{w}}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})-\boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|\leq K\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})-\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right|,\xrightarrow{e.a.s.}0,$$

where $K<\infty$, by an application with Lemma 2.1 of Straumann and Mikosch (2006). Analogously

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\hat{X}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})-X_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|\leq K\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})-\boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right|,\xrightarrow{e.a.s.}0,$$

We then conclude that S.3 is satisfied and (B.15) holds true.

Finally, we prove the existence of the integer $m\geq 1$ in (B.16), i.e., the arbitrary large number of bounded moments of the derivative processes. We remark again that we give details for the derivatives in (i). The fact that $\{\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and are stationary and ergodic implies that they admit the following almost sure representations

$$\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})=\boldsymbol{w}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})+\sum_{p=1}^{\infty}\left(\prod_{q=1}^{p}X_{i,j|L_{ij}\,;t-q}\right)\boldsymbol{w}_{i,j|L_{ij}\,;t-p}(\boldsymbol{\theta}),$$

Now, by Assumption 4, the compactness of the parameter space $\boldsymbol{\Theta}$, and the uniformly boundedness of the score function (and its derivative), it holds that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t+1}(\boldsymbol{\theta})\right\|\leq K_w+\sum_{p=1}^{\infty}\gamma_{i,j|L_{ij}}^{-p}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\boldsymbol{w}_{i,j|L_{ij}\,;t-p}(\boldsymbol{\theta})\right\|,$$

where $K_w\geq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{w}_{i,j\,;t}(\boldsymbol{\theta})\|$.

Thus, the result in (B.16) can be established by repeated applications of the Minkowski and Hölder inequalities. This result follows because $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}|f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})|^m]<\infty$ with $m\geq 1$ as implied by Proposition 3, together with Proposition TA.3 of Blasques et al. (2022) to the unperturbed derivative processes $\{\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$. In fact, we only need to note that their conditions (iii) and (iv) are directly implied by the uniform bound of the score equations together with Assumption 4. ■

Lemma 5. *Under the Assumptions 1-6:*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta\theta}}\hat{f}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})-\nabla^{\boldsymbol{\theta\theta}} f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|\xrightarrow{e.a.s.}0,\tag{B.21}$$

as $t\to\infty$.

Furthermore, we also have

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta\theta}} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\right\|^m\right] < \infty, \tag{B.22}$$

for any integer $m \geq 1$.

Proof. To prove this Lemma, we can show again that the conditions S.1-S.3 in Theorem 2.10 of Straumann and Mikosch (2006) for the second derivative processes hold true for the second derivative processes

$$\begin{aligned}
\nabla^{\boldsymbol{\theta\theta}} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) = &\begin{pmatrix}
\nabla^{\omega\omega} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) & \nabla^{\omega\alpha} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) & \nabla^{\omega\beta} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) & \nabla^{\omega\nu} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) \\
\star & \nabla^{\alpha\alpha} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) & \nabla^{\alpha\beta} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) & \nabla^{\alpha\nu} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) \\
\star & \star & \nabla^{\beta\beta} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) & \nabla^{\beta\nu} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta}) \\
\star & \star & \star & \nabla^{\nu\nu} f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta})
\end{pmatrix} \\
=&\boldsymbol{W}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) + X_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta\theta}} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}),
\end{aligned}$$

with $X_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})$ as defined in equation (B.18) and

$$\boldsymbol{W}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) = \begin{pmatrix}
\frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}^2} & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}\partial \alpha_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}\partial \beta_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}\partial \nu_{i,j|L_{ij}}} \\
\star & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}^2} & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}\partial \beta_{i,j|L_{ij}}} & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}\partial \nu_{i,j|L_{ij}}} \\
\star & \star & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}}^2} & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}}\partial \nu_{i,j|L_{ij}}} \\
\star & \star & \star & \frac{\partial^2 f_{i,j|L_{ij}\,;\,t+1}(\boldsymbol{\theta})}{\partial \nu^2}
\end{pmatrix},$$

such that

$$\frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}^2} = \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}\partial \alpha_{i,j|L_{ij}}} = \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}\partial \beta_{i,j|L_{ij}}} = \frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \omega_{i,j|L_{ij}}\partial \nu_{i,j|L_{ij}}} = 0,$$

$$\begin{aligned}
\frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}}^2} =& s^{f}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\nabla^{\alpha} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) + \alpha_{i,j|L_{ij}} s^{ff}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\nabla^{\alpha} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^2 \\
\frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}}^2} =& 2\nabla^{\beta} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) + \alpha_{i,j|L_{ij}} s^{ff}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\nabla^{\beta} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^2 \\
\frac{\partial^2 f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})}{\partial \nu_{i,j|L_{ij}}^2} =& \alpha_{i,j|L_{ij}}\left(s^{\nu\nu}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta}) + s^{ff}_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})\nabla^{\nu} f_{i,j|L_{ij}\,;\,t}(\boldsymbol{\theta})^2\right),
\end{aligned}$$

and moreover, we have

$$
\begin{aligned}
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}} \partial \beta_{i,j|L_{ij}}} &= \left(1 + \alpha_{i,j|L_{ij}} s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\beta} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right) \nabla^{\alpha} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \\
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \alpha_{i,j|L_{ij}} \partial \nu_{i,j|L_{ij}}} &= s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + \alpha_{i,j|L_{ij}} s^{f\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\alpha} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \\
\frac{\partial^2 f_{i,j|L_{ij};t}(\boldsymbol{\theta})}{\partial \beta_{i,j|L_{ij}} \partial \nu_{i,j|L_{ij}}} &= \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) + \alpha_{i,j} s^{f\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\nu} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \nabla^{\beta} f_{i,j|L_{ij};t}(\boldsymbol{\theta}),
\end{aligned}
$$

where

$$
\begin{aligned}
s^{ff}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta})^2}, \\
s^{\nu\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial \nu^2_{i,j|L_{ij};t}}, \\
s^{f\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta}) &= \frac{\partial^2 s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta}); \boldsymbol{\theta}\right)}{\partial f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \partial \nu}.
\end{aligned}
$$

From this formulas and Proposition 3.4 of Blasques et al. (2022) it is obvious that the same arguments discussed in Lemma 4 apply sequentially, yielding the desired results in (B.21) and (B.22). ■

Lemma 6. *Under Assumption 1-6, the process* $\{\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$ *is a square integrable martingle difference, that is,* $\mathbb{E}[\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) | \mathcal{F}_{t-1}] = \mathbf{0}$ *and* $\mathbb{E}[(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))^{\top}] < \infty$.

Moreover, we have that

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0) \Rightarrow \mathcal{N}\left(\mathbf{0}, \mathbb{E}[(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0))^{\top}]\right).
$$

Proof. To show the zero mean property of the score vector, we take term-wise derivatives of the log-likelihood function $\ell_t(\boldsymbol{\theta})$ in (10) for each couple of indices (i, j), in order to obtain the following score vector:

$$
\nabla^{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) = \begin{pmatrix} \nabla^{\nu} \ell_t(\boldsymbol{\theta}) \\ \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) s_{i,j|L_{ij};t}(\boldsymbol{\theta}) \end{pmatrix},
$$

where

$$
\nabla^{\nu} \ell_t(\boldsymbol{\theta}) = \frac{1}{2}\left[\psi\left(\frac{\nu + N}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{N}{\nu - 2} - \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{(\nu - 2)^2} w_t - \log\left(1 + \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu - 2}\right)\right],
$$

$$w_t = w_t\left(\boldsymbol{R}_t(\boldsymbol{\theta})\,,\,\boldsymbol{y}_t\,;\,\boldsymbol{\theta}\right) = \frac{\nu+N}{\nu-2+\boldsymbol{y}_t^\top \boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t},$$

and with $s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right) = s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)$ as defined in equation (A.2), respectively.

Now, by a straightforward application of the conditional expectation we obtain

$$\mathbb{E}\left[\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\left(\begin{matrix}\nabla^{\nu}\ell_t(\boldsymbol{\theta}_0)\\ \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}_0\right)\end{matrix}\right)\,\middle|\,\mathcal{F}_{t-1}\right] = \left(\begin{matrix}0\\0\end{matrix}\right),$$

where the last equality follow because the derivatives $\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)$ are $\mathcal{F}_{t-1}$-measurable, whereas the conditional expectations of $\nabla^{\nu}\ell_t(\boldsymbol{\theta}_0)$, and $s_{i,j|L_{ij}\,;t}(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}_0)\ ;\ \boldsymbol{\theta}_0)$ are obviously zero almost surely, since, by Assumption 2, they are the terms of the conditional score vector of the multivariate Student t density function evaluated at the true parameter vector $\boldsymbol{\theta}_0$. On the other hand, to show that $\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)$ is square integrable, it suffices to prove that the derivatives of the log-likelihood have a uniformly bounded second moment, that is

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|^2\right] < \infty. \tag{B.23}$$

An application of the Cauchy-Schwartz inequality, we can show that

$$\begin{aligned}
\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|^2\right] &\leq \mathbb{E}\left[\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right| + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right)\right\|\right)^2\right]\\
&\leq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right|^2\right] + 2\left(\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right|^2\right]\cdot\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|^2\cdot\left(\frac{K_1\epsilon}{1-\epsilon^2}\right)^2\right]\right)^{1/2}\\
&\quad+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\right\|^2\cdot\left(\frac{K_1\epsilon}{1-\epsilon^2}\right)^2\right],
\end{aligned}$$

where the last inequality follows by the arguments discussed in Proposition 3 since the uniform boundedness of the score $s_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}) = s_{i,j|L_{ij}\,;t}(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}), \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta}); \boldsymbol{\theta})$ implies the existence of an arbitrary large number of bounded moments. Hence

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|s_{i,j|L_{ij}\,;t}\left(\boldsymbol{\theta}\right)\right\|^2\right] = \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|s_{i,j|L_{ij}\,;t}\left(f_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij}\,;t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right\|^2\right] \leq \left(\frac{K_1\epsilon}{1-\epsilon^2}\right)^2 < \infty.$$

Moreover, from Lemma 4 it also holds that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left\|\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|^2\right] < \infty.$$

Now, by the compactness of the parameter space $\boldsymbol{\Theta}$, we can also show that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right|^2\right] \\
\leq &\frac{1}{4}\mathbb{E}\left[\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\psi\left(\frac{\nu+N}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{N}{\nu-2}-\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2}w_t\right|+\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|\right)^2\right] \\
\leq &\frac{1}{4}\mathbb{E}\left[|K_\nu-K_1|^2\right]+\frac{1}{2}\left(|K_\nu-K_1|\cdot\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|\right]\right) \\
&\quad+\frac{1}{4}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|^2\right],
\end{aligned}$$

where the second inequality holds because from the compactness of the parameter space $\boldsymbol{\Theta}$ with $2<\nu<\infty$, $\exists K_\nu \geq \psi\left(\frac{\nu+N}{2}\right)+\psi\left(\frac{\nu}{2}\right)+\frac{N}{\nu-2}$, together with the the analytical form of the weights w_t, which implies that $\exists K_1>0$ such that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^2}w_t\right|^2\right]\leq K_1<\infty.$$

Moreover, it is obvious that from the second moment bound $\mathbb{E}[\|\boldsymbol{y}_t\|^2]<\infty$ and the lower bound in (A.10) we also have that $\exists K_2>0$ such that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}}\left|\log\left(1+\frac{\boldsymbol{y}_t^\top\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}\right)\right|^2\right]\leq K_2<\infty,$$

by virtue of the inequality $\log(1+x)\leq x \ \forall x\geq -1$. By collecting all the results obtained above, we conclude that (B.23) holds true.

Finally, we simply note that the Fisher's information equality $\mathbb{E}[(\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0))(\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0))^\top]=\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ follows by standard arguments since by Assumption 2 the $\ell_t(\boldsymbol{\theta}_0)$ is the true conditional log-density of the Student's t distribution. This concludes the proof.

$\blacksquare$

Lemma 7. *Under Assumptions 1-4,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{T}\nabla^{\boldsymbol{\theta}}\hat{L}_T(\boldsymbol{\theta})-\frac{1}{T}\nabla^{\boldsymbol{\theta}}L_T(\boldsymbol{\theta})\right\|\xrightarrow{a.s.}0, \tag{B.24}$$

as $T\to\infty$.

Proof. An application of the triangle inequality yields

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{1}{T}\nabla^{\boldsymbol{\theta}}\hat{L}_T(\boldsymbol{\theta})-\frac{1}{T}\nabla^{\boldsymbol{\theta}}L_T(\boldsymbol{\theta})\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}}\hat{\ell}_t(\boldsymbol{\theta})-\nabla^{\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\Big(I+II\Big),$$

with

$$\begin{aligned}I:=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|&\nabla^{\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\\&-\ \nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\Big\|,\end{aligned}$$

and

$$II:=\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\nu}\hat{\ell}_t(\boldsymbol{\theta})-\nabla^{\nu}\ell_t(\boldsymbol{\theta})\right\|.$$

As a first step, we focus on I. We recognize that each term of

$$\nabla^{\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)$$

is a continuous function of $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and its derivatives. In contrast, the terms in

$$\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)$$

are continuous functions of the stationary counterparts, i.e. $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and its derivatives. Therefore, by means of elementary decomposition, we can write

$$\begin{aligned}I\leq&\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ -\ \nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right)\\&\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|s_{i,j|L_{ij};t}\left(\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \hat{\boldsymbol{y}}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)-s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right|\\&+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}}\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ -\ \nabla^{\boldsymbol{\theta}}f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|s_{i,j|L_{ij};t}\left(f_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ,\ \boldsymbol{y}^{\star}_{i,j|L_{ij};t}(\boldsymbol{\theta})\ ;\ \boldsymbol{\theta}\right)\right|.\end{aligned} \tag{B.25}$$

Now, we note that in view of Proposition 3 and Lemma 4, we can easily show that both the first and the second addends of the inequality (B.25) vanish *e.a.s.* as $t \to \infty$, as implied by Lemma 2.1 in Straumann and Mikosch (2006).

Therefore, there exists some finite constant $K_I > 0$ such that

$$I \leq K_I \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \gamma_{i,j}^{-t}$$

and since for $\gamma_{i,j}^{-t} < 1 \; \forall t \in \mathbb{N}$, we obtain that $I \xrightarrow{e.a.s.} 0$, as $t \to \infty$.

As concerns II, we have that

$$\begin{aligned}\nabla^{\nu} \hat{\ell}_t(\boldsymbol{\theta}) - \nabla^{\nu} \ell_t(\boldsymbol{\theta}) = \frac{1}{2} \Bigg[& \frac{\boldsymbol{y}_t^{\top} \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{(\nu-2)^2} w_t - \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{(\nu-2)^2} \hat{w}_t \\ & + \log\left(1 + \frac{\boldsymbol{y}_t^{\top} \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu-2}\right) - \log\left(1 + \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu-2}\right) \Bigg]. \end{aligned} \tag{B.26}$$

We can then combine the facts that: (i) $0 < \nu < \infty$ by Assumption 2, (ii) the lower bound obtained in (A.10) and (iii) the uniform bound $\sup_{\boldsymbol{\theta} \in \boldsymbol{\theta}} |w_t| \leq 1$, in order to see that for the first added in squared brackets of the right hand side of equation (B.26), it holds that

$$\begin{aligned}\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\boldsymbol{y}_t^{\top} \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{(\nu-2)^2} \hat{w}_t - \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{(\nu-2)^2} w_t \right| &\leq c_{\nu} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| tr\Big(\boldsymbol{y}_t \boldsymbol{y}_t^{\top} (\boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) - \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta})) \Big) \right| \\ &= c_{\nu} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| tr\Big(\boldsymbol{y}_t \boldsymbol{y}_t^{\top} \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) (\hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})) \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \Big) \right| \\ &\leq c_{\nu} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) \right\| \left\| \hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta}) \right\| \left\| \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \right\| \| \boldsymbol{y}_t \boldsymbol{y}_t^{\top} \| \\ &\leq c_{\nu} K \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta}) \right\| \| \boldsymbol{y}_t \|^2. \end{aligned}$$

Moreover, since $\log x \leq x - 1 \; \forall x \geq 1$, the same result holds for the second added in squared brackets of the right hand side of equation (B.26), in fact

$$\begin{aligned}\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \log\left(1 + \frac{\boldsymbol{y}_t^{\top} \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu-2}\right) - \log\left(1 + \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu-2}\right) \right| &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\boldsymbol{y}_t^{\top} \hat{\boldsymbol{R}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu-2} - \frac{\boldsymbol{y}_t^{\top} \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{y}_t}{\nu-2} \right| \\ &\leq c_{\nu} K \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \hat{\boldsymbol{R}}_t(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta}) \right\| \| \boldsymbol{y}_t \|^2, \end{aligned}$$

for some $K, c_{\nu} > 0$. We can now recall that the conditional correlation matrix $\hat{\boldsymbol{R}}_t(\boldsymbol{\theta})$ is a continuous function of each $\hat{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, whereas $\boldsymbol{R}_t(\boldsymbol{\theta})$ is a continuous function of each of the stationary counterpart $f_{i,j|L_{ij};t}(\boldsymbol{\theta})$. Therefore, by Proposition 3 it holds that

$$II \leq 2c_\nu K \|\boldsymbol{y}_t\|^2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \gamma_{i,j}^{-t},$$

and since $\gamma_{i,j}^{-t} < 1$, $\forall t \in \mathbb{N}$ and $\mathbb{E}[\|\boldsymbol{y}_t\|^2] < \infty$ we obtain that $II \xrightarrow{e.a.s.} 0$, as $t \to \infty$.

In conclusion, the uniform convergence in (B.24) holds true. ∎

Lemma 8. *Under Assumptions 1-6,*

$$-\frac{1}{T}\sum_{t=1}^{T} \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_T) \xrightarrow{a.s.} -\mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}_0)] = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0), \tag{B.27}$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is positive definite.

Proof. First, we establish the almost sure convergence in (B.27), by proving that the second derivatives of the log-likelihood function has a uniformly bounded moment, that is

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta})\right\|\right] < \infty. \tag{B.28}$$

Then, analogously to the Proof of Thorem 1, we apply again the uniform law of large numbers for stationary and ergodic processes of Rao (1962).

Taking term-wise second derivatives of the log-likelihood function $\ell_t(\boldsymbol{\theta})$ in (10) for each couple of indices (i,j), we obtain the following Hessian matrix:

$$\begin{aligned}&\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta})\\ &= \begin{pmatrix} \nabla^{\nu\nu}\ell_t(\boldsymbol{\theta}) & \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}) \\ \star & \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f^{\top}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + s_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right)\end{pmatrix},\end{aligned}$$

where

$$\begin{aligned}\nabla^{\nu\nu}\ell_t(\boldsymbol{\theta}) = \frac{1}{4}\Bigg[&\psi\prime\left(\frac{\nu+N}{2}\right) - \psi\prime\left(\frac{\nu}{2}\right) + \frac{2N}{(\nu-2)^2} + \frac{4\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^3}w_t \\ &+ \frac{2N\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{(\nu-2)^3(\nu+N)^2}w_t^3 - \frac{2(\nu+N)\boldsymbol{y}_t^{\top}\boldsymbol{R}_t^{-1}(\boldsymbol{\theta})\boldsymbol{y}_t}{\nu-2}w_t\Bigg],\end{aligned}$$

with w_t as defined in Lemma 6, $s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, $s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})$ and $s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})$, are as defined in equations

(A.2), (A.3) and (B.19), respectively. Now, by elementary calculations, we note that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\right\|\right] \leq \mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla^{\nu\nu}\ell_t(\boldsymbol{\theta})\right\|\right] + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|s^{\nu}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right]$$
$$+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|s^{f}_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f^{\top}_{i,j|L_{ij};t}(\boldsymbol{\theta}) + s_{i,j|L_{ij};t}(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta})\right\|\right].$$

In view of the the uniform boundedness properties of the score expression and the results obtained in Proposition 3, and Lemmas 4 and 5 it is clear that all the elements of the Hessian matrix are still polynomials of uniformly bounded random variables, with an arbitrary large number of finite moments. Thus, repeated applications of the Cauchy-Schwartz inequality and the Minkowsky inequality, and some straightforward calculations, allow us to conclude that also the Hessian matrix is uniformly bounded, and hence (B.28) holds true. Therefore, a straightforward application of the uniform law of large number for stationary and ergodic sequences of Rao (1962) give us the desired almost sure convergence in (B.27).

Second, we show that $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is positive definite. To do so, we note that the strong consistency of the MLE $\hat{\boldsymbol{\theta}}_T$ established in Theorem 1, implies that as $T\to\infty$, $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ and hence $\hat{\boldsymbol{\theta}}_T \in V(\boldsymbol{\theta}_0)$ almost surely, where $V(\boldsymbol{\theta}_0)$ denotes a neighbourhood of $\boldsymbol{\theta}_0$.

We thus have that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)]\right\| \leq \left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0) - \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)]\right\|$$
$$+\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}) - \frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)\right\|.$$

However, since $\{\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is stationary and ergodic, it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0) \xrightarrow{a.s.} \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)] = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0),$$

and therefore, by the uniform law of large numbers of Rao (1962), $\exists\delta > 0$ such that

$$\lim_{T\to\infty}\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\hat{\boldsymbol{\theta}}_T) - \mathbb{E}[\nabla^{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0)]\right\| \leq \delta.$$

As the constant $\delta > 0$ can be chosen as small as we want, we conclude that the almost sure convergence in (B.27) holds true. In conclusion, it remains to be shown that $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is invertible. As argued by Darolles et al. (2018) in their proof of Theorem 4.3, if $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ were not invertible, than there would exists some vector $\boldsymbol{\lambda} \in \mathbb{R}^d$, where d denotes the dimension of the compact parameter space $\boldsymbol{\theta}$, such that $\boldsymbol{\lambda}^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)\boldsymbol{\lambda} = 0$ with $\boldsymbol{\lambda} \neq \mathbf{0}$. However, from the lower bound derived in (A.10), it is clear that also the matrix $\boldsymbol{R}_t^{-1}(\boldsymbol{\theta}_0) \otimes \boldsymbol{R}_t^{-1}(\boldsymbol{\theta}_0)$ is almost surely positive definite.

Thus, if $\exists \boldsymbol{\lambda} \neq \mathbf{0}$ such that $\boldsymbol{\lambda}^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)\boldsymbol{\lambda} = 0$, then it must also be that

$$\boldsymbol{\lambda}^\top \nabla^{\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0) = 0. \qquad \boldsymbol{\lambda}^\top \nabla^{\boldsymbol{\theta}\boldsymbol{\theta}} f_{i,j|L_{ij};t}(\boldsymbol{\theta}_0)\boldsymbol{\lambda} = 0,$$

almost surely.

By drawing attention to the expression of the first and second derivative processes given in Lemmas 4 and 5, it is straightforward to see that each of these processes are linearly independent, because they can be rewritten in terms of *IID* random vectors by following the same arguments already discussed in the Proof of Theorem 2. Therefore, we conclude that $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ is positive definite, thus completing the proof.

■