A robust score-driven filter for multivariate time series

Published Version:
Enzo D’Innocenzo, Alessandra Luati, Mario Mazzocchi (2023). A robust score-driven filter for multivariate time series. ECONOMETRIC REVIEWS, 42(5), 441-470 [10.1080/07474938.2023.2198930].

Availability:
This version is available at: https://hdl.handle.net/11585/936633 since: 2023-07-26

Published:
DOI: http://doi.org/10.1080/07474938.2023.2198930

A robust score-driven filter for multivariate time series

Enzo D'Innocenzo*
Department of Economics, University of Bologna
Piazza Antonio Scaravilli 2, 40126 Bologna, Italy
e-mail: enzo.dinnocenzo2@unibo.it

Alessandra Luati
Department of Mathematics, Imperial College London
180 Queen's Gate, SW7 2AZ, London
e-mail: a.luati@imperial.ac.uk
and

Mario Mazzocchi
Department of Statistical Sciences, University of Bologna,
Via delle Belle Arti 41, 40126 Bologna, Italy,
e-mail: m.mazzocchi@unibo.it

March 17, 2023

*Corresponding author.

Abstract

A multivariate score-driven filter is developed to extract signals from noisy vector processes. By assuming that the conditional location vector from a multivariate Student's t distribution changes over time, we construct a robust filter which is able to overcome several issues that naturally arise when modeling heavy-tailed phenomena and, more in general, vectors of dependent non-Gaussian time series. We derive conditions for stationarity and invertibility and estimate the unknown parameters by maximum likelihood. Strong consistency and asymptotic normality of the estimator are derived. Analytical formulae are derived which consent to develop estimation procedures based on a fast and reliable Fisher scoring method. An extensive Monte-Carlo study is designed to assess the finite samples properties of the estimator, the impact of initial conditions on the filtered sequence, the performance when some of the underlying assumptions are violated, such as symmetry of the underlying distribution and homogeneity of the degrees of freedom parameter across marginals. The theory is supported by a novel empirical illustration that shows how the model can be effectively applied to estimate consumer prices from home scanner data.

Keywords: Robust filtering, Multivariate models, Score-driven models, Homescan data.

1 Introduction

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

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

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

Within the class of observation-driven models, score-driven models are a valid option for modeling time series that do not fall in the category of linear Gaussian processes. Examples have been proposed in the context of volatility estimation and originally referred to as generalised autoregressive score (GAS) models, Creal et al. (2013), and as dynamic conditional score (DCS) models, Harvey (2013). The key feature of these models is that the dynamics of time-varying parameters are driven by the score of the conditional likelihood, which needs not necessarily to be Gaussian but can be heavy tailed, as earlier discussed by Masreliez (1975) and Masreliez and Martin (1977). For example, it may follow a Student's t distribution as in Harvey and Luati (2014) and Linton and Wu (2020), an exponential generalized beta distribution, as in Caivano et al. (2016), a binomial distribution as in the vaccine example by Hansen and Schmidtblaicher (2019), or represented by a mixture, see Lucas et al. (2019). The optimality of the score as a driving force for time varying parameters in observation-driven models is discussed in Blasques et al. (2015). According to which conditional distribution is adopted, specific situations may be conveniently handled due to the properties of the score. As an example for the univariate case, if a heavy-tailed distribution is specified, namely Student's t, the resulting score-driven model yields a simple and natural model-based signal extraction filter which is robust to extreme observations, without any external interventions or diagnostics, like dummy variables or outlier detection, see Harvey and Luati (2014).

In score-driven models, as well as in all observation-driven models, the time varying parameters are updated by filtering procedures, i.e. weighted sums of functions of past observations, given some initial conditions that can be fixed or estimated along with the static parameters. A robust filtering procedure should assign less weight to extreme observations in order to prevent biased inference of the signal and the parameters. In particular, the work of Calvet et al. (2015) provides a remarkable application of robust methods when dealing with contaminated observations. The authors show that a substantial efficiency gain can be achieved by huberizing the derivative of the log-observation density. As we show in the present study, the same holds if one considers an alternative robustification method, based on the specification of a conditional multivariate Student's t distribution. A similar approach can be found in Prucha and Kelejian (1984) and Fiorentini et al. (2003), where the multivariate Student's t distribution provides a valid alternative to relax the normality assumption. In the context of score-driven models, Creal et al. (2014) mention the relevance of modeling high-frequency data with outliers and heavy tails by means of the multivariate Student's t distribution.

In this paper, we develop a score-driven filter for the time-varying location of a multivariate Student's t distribution and derive its stochastic and asymptotic properties. The specification is similar to the multivariate model for the location addressed in Harvey (2013) and has some traits in common with the quasi-vector autoregressive model by Blazsek et al. (2017). Both these contributions extend the univariate model by Harvey and Luati (2014) to the case of $N > 1$ time series, but neither the probabilistic nor the full asymptotic theory is derived for the multivariate specification. As a matter of fact, some aspects of the probabilistic and asymptotic theory in multivariate non-linear models are non-trivial. Notable examples are the proof of asymptotic irrelevance of initial conditions for the filtering recursions and the calculation of higher order derivatives required to characterize asymptotic normality. These aspects are covered by the present paper, so that a comprehensive theory for the class of multivariate score driven models for the location parameter is available and easy to apply. We envisage three main contributions to the existing literature.

The first contribution of the paper is the derivation of the probabilistic theory behind the multivariate dynamic score-driven filter for conditional Student's t distributions, including the conditions of stationarity, ergodicity and invertibility, that we face in a similar spirit of Comte and Lieberman (2003) and Hafner and Preminger (2009) for the multivariate conditional variance models by Baba et al. (1990) and Engle and Kroner (1995). Invertibility is a particularly delicate issue in multivariate non linear models (see Blasques et al. 2018) and its proof typically require high level assumptions, often difficult to verify in practice. In the paper, we derive conditions of invertibility of the model, following Straumann and Mikosch (2006). In addition, we derive a sufficient condition for invertibility and a consequent non degenerate region of the parameter space where convergence is valid for the given model.

As in score driven models the conditional likelihood is available in closed form, we estimate the static parameters with the method of maximum likelihood and prove strong consistency and asymptotic normality of the estimators. A relevant contribution is the derivation of a closed-form expression for the conditional expected Fisher information matrix, that approximates the inverse variance-covariance matrix of the estimator, usually unavailable analytically in non linear non Gaussian models, see the discussion in Fiorentini et al. (2003). It is noteworthy to remark that when the degrees of freedom of the Student's t distribution tend to infinity, we recover a linear Gaussian model. These theoretical results provide the basis for further generalisations, such as, for instance, the spatial model by Gasperoni et al. (2021). As the asymptotic results are derived under the assumption of a correct specification, potential model misspecification is investigated through a large scale simulation study. There, the multivariate Student's t specification is compared with alternative distributions that show features such as asymmetry and heterogeneity in the degrees of freedom, in addition to heavy tails. Finite sample properties, such as impact of initial conditions on relatively small samples and the capability of recognizing Gaussian data, are also analysed through simulations.

The second contribution is the development of an estimation scheme grounded on Fisher's scoring method, based on the closed-form analytic expression of the conditional Fisher information matrix discussed so far, which can be directly implemented into any statistical or matrix-friendly software. The computational gain of the scoring algorithm based on analytical formulae, compared to the usual ones based on numerical derivatives, allow a fast implementation of the method, even in case of large temporal or cross-section dimension.

The third contribution of the paper is an innovative application, dealing with estimation of regional consumer prices based on home scanner data. The use of scanner data to compute official consumer price indices (CPIs) is gaining popularity, because of their timeliness and a high level of product and geographical detail Shapiro and Feenstra (2003). However, they also suffer from a variety of shortcomings, which make time series of scanner data prices (SDPs) potentially very noisy, especially when they are estimated for population sub-groups, or at the regional level Silver (1995). There is extensive research and a lively debate on the issues related to the computation and use of scanner data based CPIs. In a dedicated session of the 2019 meeting of the the Ottawa Group on Price Indices, it has been suggested[1] to adopt model-based filtering techniques to extract the signal from scanner-based time series of price data. These filtered estimates lose the classical price index formula interpretation, but are expected to deliver the same information content with a better signal-to-noise ratio. We show that our robust multivariate model, applied to SDPs, provides information on the dynamics of the time series and on their interrelations without being affected from outlying observations, which are naturally downweighted in the updating mechanism. This holds also in comparison with a fitted linear Gaussian model and with estimated univariate models by Harvey and Luati (2014).

The paper is organised as follows. In Section 2 the filter is specified. Section 3 deals with the stochastic properties of the filter, while in Section 4, likelihood inference is discussed. Section 5 describes the Monte Carlo experiments. The empirical analysis is reported in section 6. Some concluding remarks are drawn in Section 7. The proofs of the results stated in the paper are collected in Appendix A. Online supplementary materials contain further details of the Monte Carlo study, the relevant

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

quantities for deriving the conditional Fisher information matrix and implementing the Fisher scoring algorithm, as well as the proofs of some auxiliary Lemmata.

2 The Multivariate Student's t Location Filter

Let us consider a $\mathbb{R}^N$-vector of stochastic processes $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$, $N \geq 1$, and let $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \boldsymbol{y}_{t-3}, \dots\}$ be its filtration at time $t-1$. The following stochastic representation of $\boldsymbol{y}_t$ is considered,

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t, \tag{1}$$

where $\boldsymbol{\mu}_t$ is a time varying location vector of $\mathbb{R}^N$, $\boldsymbol{\Omega}$ is a $N \times N$ scale matrix that we assume to be static and $\boldsymbol{\epsilon}_t \sim \boldsymbol{t}_\nu(\boldsymbol{0}_N, \boldsymbol{I}_N)$ is an independent identically distributed (*IID*) multivariate standard t-variate. With $\boldsymbol{0}_N$ we denote the null vector of $\mathbb{R}^N$ and with $\boldsymbol{I}_N$ the $N \times N$ identity matrix.

Our interest is in recovering $\boldsymbol{\mu}_t$ based on a set of observed time series from $\boldsymbol{y}_t$, for $t = 1, \dots, T$, where $T \in \mathbb{N}$. With no distributional assumptions on the dynamics of $\boldsymbol{\mu}_t$, a filter can be specified,

$$\boldsymbol{\mu}_{t+1|t} = \phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}), \tag{2}$$

that is a stochastic recurrence equation (SRE), where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is a vector of unknown static parameters, $\boldsymbol{\mu}_{t|t-1}$ is a $\mathbb{R}^N$-random vector that takes values in $\boldsymbol{\mathcal{M}} \subset \mathbb{R}^N$ and $\phi : \boldsymbol{\mathcal{M}} \times \mathbb{R}^N \times \boldsymbol{\Theta} \mapsto \boldsymbol{\mathcal{M}}$ is a Lipschitz function. The subscript notation $t|t-1$ is used to emphasize the fact that $\boldsymbol{\mu}_{t|t-1}$ is an approximation of the dynamic location process at time t given the past, that is equivalent to say that $\boldsymbol{\mu}_{t|t-1}$ is $\mathcal{F}_{t-1}$-measurable. Therefore, based on past observations and a starting value $\boldsymbol{\mu}_{1|0} \in \boldsymbol{\mathcal{M}}$, one can approximate the unobserved path of $\boldsymbol{\mu}_t$ in (1) by mimicking the recursion in (2). It is typically assumed that a parameter value $\boldsymbol{\theta}_0$ exists, at which the true location can be recovered, i.e. $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0) = \boldsymbol{\mu}_t$ (assumption 1 of correct specification).

In this paper, we approximate the temporal changes of the dynamic location by relying on the score-driven framework of Creal et al. (2013) and Harvey (2013). Specifically, we assume that, conditional on the past, the distribution of $\boldsymbol{y}_t$ is Student's t, with $\nu > 0$ degrees of freedom and conditional location equal to $\boldsymbol{\mu}_{t|t-1}$, i.e.

$$f(\boldsymbol{y}_t|\mathcal{F}_{t-1}) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{N/2}}|\boldsymbol{\Omega}|^{-1/2}\left[1 + \frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})}{\nu}\right]^{-(\nu+N)/2} \tag{3}$$

and specify the SRE in (2) as follows,

$$\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega} = \boldsymbol{\Phi}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\boldsymbol{u}_t, \tag{4}$$

i.e.

$$\phi(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) := \boldsymbol{\omega} + \boldsymbol{\Phi}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\boldsymbol{u}_t,$$

where $\boldsymbol{\omega}$ is a $\mathbb{R}^N$ vector of unconditional means, $\boldsymbol{\Phi}$ and $\boldsymbol{K}$ are $\mathbb{R}^{N\times N}$ matrices of coefficients and the driving force $\boldsymbol{u}_t$ is proportional to the score of conditional density in (3). Indeed, the conditional score with respect to the time varying location filter is

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

where

$$\boldsymbol{u}_t = (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/w_t, \tag{5}$$

with $w_t = 1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu$, is a martingale difference sequence, i.e. $\mathbb{E}_{t-1}[\boldsymbol{u}_t] = \mathbf{0}_N$, under correct specification, where the shorthand notation $\mathbb{E}_{t-1}[X]$ is used for the conditional expectation $\mathbb{E}[X|\mathcal{F}_{t-1}]$. The score as the driving force in an updating equation for a time varying parameter is the key feature of score-driven models. The rationale behind the recursion (4) is very intuitive. Analogously to the Gauss-Newton algorithm, it improves the model fit by pointing in the direction of the greatest increase of the likelihood. Optimality of score driven updates in observation-driven models is discussed by Blasques et al. (2015).

In the context of location estimation under the Student's t assumption, a further relevant motivation for the score-driven methodology lies in the robustness of the implied filters. Indeed, the positive scaling factors w_t in equation (5) are scalar weights that involve the Mahalanobis distance. They possess the role of re-weighting the large deviation from the mean incorporated in the innovation error

$$\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}. \tag{6}$$

Robustness comes precisely from winsorizing the innovation error $\boldsymbol{v}_t$. Note that when $\nu \to \infty$, $\boldsymbol{u}_t$ converges to $\boldsymbol{v}_t$ and equations (4) and (6) coincide with the steady state innovation form of a linear Gaussian state-space model.

A formal proof of the robustness of the method is in the following Lemma, which provides sufficient conditions for a filter to be robust, in line with Calvet et al. (2015). We first enounce the correct specification assumption.

Assumption 1. *The filter in* (4) *is correctly specified, i.e. when* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, *where* $\boldsymbol{\theta}_0$ *is the true parameter vector,* $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0) = \boldsymbol{\mu}_t$.

Lemma 1. *Under assumption 1, for* $0 < \nu < \infty$, *the vector sequence* $\{\boldsymbol{u}_t\}_{t\in\mathbb{Z}}$ *is uniformly bounded, that is* $\sup_t \mathbb{E}[\|\boldsymbol{u}_t\|] < \infty$ *and possesses all the even moments*

$$\mathbb{E}[\|\boldsymbol{u}_t\|^{2s}] = \|\boldsymbol{\Omega}\|^s \frac{B\big(\frac{N+2s}{2}, \frac{\nu+2s}{2}\big)}{B\big(\frac{N}{2}, \frac{\nu}{2}\big)}\Big(\frac{\nu}{N}\Big)^s,$$

for $s = 1, 2, \ldots$ *and where* $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ *is the beta function and* $\|\boldsymbol{\Omega}\| = \sqrt{\mathrm{tr}(\boldsymbol{\Omega}^\top\boldsymbol{\Omega})}$. *The odd moments of* $\boldsymbol{u}_t$ *are all equal to zero.*

The moment structure reveals important features of the driving force $\boldsymbol{u}_t$, that turns out to be an an *IID* sequence with zero mean vector and (vec)-variance covariance matrix,

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

3 Properties of the Filter

Let us combine equations (4) and (5) and write the filter explicitly, as follows,

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{\omega} + \boldsymbol{\Phi}(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega}) + \boldsymbol{K}\frac{\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}}{1 + (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})/\nu}. \tag{7}$$

By starting at some initial value, $\boldsymbol{\mu}_{1|0} \in \boldsymbol{\mathcal{M}}$, and using equation (7) for $t = 1, \ldots, T$, with $T \in \mathbb{N}$, one can recover a unique filtered path $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. A desirable property is that the values used to initialise the process are asymptotically negligible, in the sense that as the time t increases, the impact of the chosen $\boldsymbol{\mu}_{1|0}$ eventually vanishes and the process will converge to a unique stationary and ergodic sequence. This stability property of the filtered sequence $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ is known as invertibility, see Straumann and Mikosch (2006) and Blasques et al. (2018). Existence of the unique stationary and ergodic solution to the SRE (7) is established by Lemma 2. Invertibility of the filter is proved in Lemma 3.

Lemma 2. *Let us consider equation* (7)*, evaluated at* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*. Assume that* $0 < \nu < \infty$ *and* $\varrho(\boldsymbol{\Phi}) < 1$*, where* $\varrho(\boldsymbol{\Phi})$ *denotes the spectral radius of* $\boldsymbol{\Phi}$*. Then, there exists a unique vector sequence* $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$ *satisfying the recursion in* (7)*, which is strictly stationary and ergodic with* $\mathbb{E}[\|\tilde{\boldsymbol{\mu}}_{t|t-1}\|^m] < \infty$ *for every* $m > 0$.

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

With the next Lemma, the relevant conditions under which the SRE in (7) is contractive on average are given so that the convergence of the filtered sequence $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ to a unique $\mathcal{F}_{t-1}$-measurable stationary and ergodic solution $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$, irrespective of the initialization $\boldsymbol{\mu}_{1|0}$, is obtained as a corollary of Theorem 3.1 of Bougerol (1993) or, equivalently, of Theorem 2.8 of Straumann and Mikosch (2006). Moreover, as a consequence of Lemma 1, both $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ and $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$ have bounded moments.

Lemma 3. *Let the conditions of Lemma 2 hold and assume that*

$$\mathbb{E}\left[\ln \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\mu}\in\boldsymbol{\mathcal{M}}} \left\| \prod_{j=1}^{k} \boldsymbol{X}_{k-j+1} \right\|\right] < 0, \tag{8}$$

for $k \geq 1$*, where* $\boldsymbol{\Theta}$ *is a compact parameter space and* $\boldsymbol{X}_t = \boldsymbol{\Phi} + \boldsymbol{K}\partial \boldsymbol{u}_t / \partial \boldsymbol{\mu}_{t|t-1}^\top$*. Then, the filtered location vector* $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ *is invertible and converges exponentially fast almost surely (e.a.s.) to the unique stationary ergodic sequence* $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$ *for any initialization of the filtering recursion,* $\boldsymbol{\mu}_{1|0} \in \boldsymbol{\mathcal{M}}$*, that is,*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \tilde{\boldsymbol{\mu}}_{t|t-1}\| \xrightarrow{e.a.s.} 0 \qquad \textit{as} \qquad t \to \infty, \tag{9}$$

Furthermore, $\sup_t \mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1}\|^m] < \infty$ *and* $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\tilde{\boldsymbol{\mu}}_{t|t-1}\|^m] < \infty, \forall m \geq 1$.

Remark 1. *The invertibility property derived in Lemma 3, is not an alternative to the stationarity and ergodicity obtained in Lemma 2, that are properties of the random process* $\{\boldsymbol{y}_t\}_{t\in\mathbb{N}}$*. In fact, invertibility is a property of the mapping from* $\{\boldsymbol{y}_t\}_{t\in\mathbb{N}}$ *to* $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ *that becomes crucial when* $\{\boldsymbol{y}_t\}_{t\in\mathbb{N}}$ *is a stationary and ergodic process, because then* $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ *will inherit these characteristics.*

The contraction condition in equation (8) imposes restrictions on the parameter space $\boldsymbol{\Theta}$ that cannot be checked directly. Also, the expectation in the same equation cannot be verified in practice, since it depends on the unconditional, unknown, distribution of $\boldsymbol{y}_t$, see also the discussion in Blasques et al. (2018). Thus, we rely on sufficient conditions which are typically more restrictive than (8) and that we discuss in the following, similarly to Linton and Wu (2020). Specifically, the contraction condition in (8) is satisfied if

$$\mathbb{E}\left[\ln \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\mu}_{1|0}\in\boldsymbol{\mathcal{M}}} \left\|\boldsymbol{X}_1\right\|\right] < 0. \tag{10}$$

Motivated by Example 3.8 of Straumann and Mikosch (2006), we rewrite $\boldsymbol{X}_1$ at $\boldsymbol{\theta}_0$, so that equation (10) becomes

$$\mathbb{E}\left[\ln \left\|\boldsymbol{\Phi}_0 + \frac{\boldsymbol{K}_0}{1+\boldsymbol{\epsilon}_1^\top\boldsymbol{\epsilon}_1/\nu_0}\left(\frac{2\boldsymbol{\Omega}_0^{1/2}\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top\boldsymbol{\Omega}_0^{-1/2}/\nu_0}{1+\boldsymbol{\epsilon}_1^\top\boldsymbol{\epsilon}_1/\nu_0} - \boldsymbol{I}_N\right)\right\|\right] < 0. \tag{11}$$

Since $\boldsymbol{\epsilon}_1 \sim \boldsymbol{t}_{\nu_0}(\boldsymbol{0}_N, \boldsymbol{I}_N)$, based on Monte Carlo simulations, Figure 1 displays a region for a bivariate model ($N = 2$) that satisfies the condition (11) on a grid of values $(\|\boldsymbol{\Phi}_0\|, \|\boldsymbol{K}_0\|) \in (0,1)^2$, with $\nu_0 = 7$ and $\boldsymbol{\Omega}_0 = \boldsymbol{I}_2$.

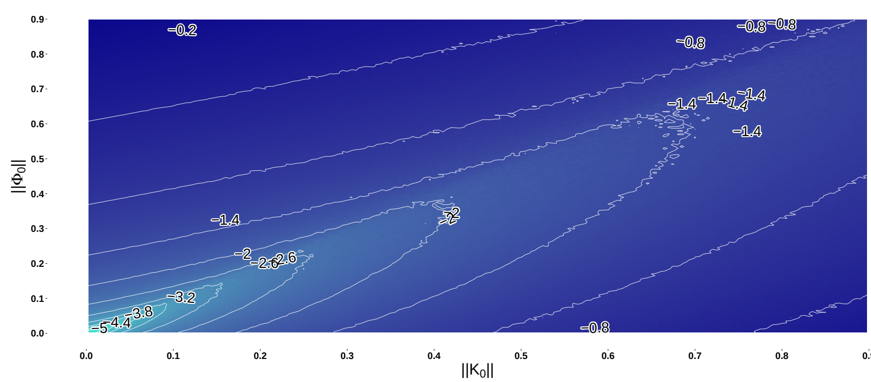

Figure 1: Contour plot of the domain for invertibility.

As expected, the restrictions that need to be imposed on $\|\boldsymbol{\Phi}_0\|$ and $\|\boldsymbol{K}_0\|$ are stronger than those required for strict stationarity and ergodicity, see Lemma 2. Neverthless, the region depicted in Figure 1 shows that a subset $\boldsymbol{\Theta}^*$ of the parameter space $\boldsymbol{\Theta}$ exists, with $\|\boldsymbol{\Phi}\| < 1$ and $\|\boldsymbol{K}\|$ sufficiently small such that (10) is satisfied $\forall\boldsymbol{\theta} \in \boldsymbol{\Theta}^* \subset \boldsymbol{\Theta}$, producing a non degenerate invertibility region.

The simulation-based method described so far involves a posteriori checking that the estimated parameters lie in the implied invertibility region. As an alternative, a restricted estimation procedure can be implemented, where parameters are constrained to satisfy the empirical version of the invertibility condition in equation (8), as in Wintenberger (2013) and Blasques et al. (2018). The empirical counterpart of (8) with $k = 1$ is

$$\frac{1}{T}\sum_{t=1}^{T} \ln \left\| \boldsymbol{\Phi} + \frac{\boldsymbol{K}}{1 + \boldsymbol{v}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{v}_t/\nu} \left(\frac{2\boldsymbol{v}_t \boldsymbol{v}_t^\top \boldsymbol{\Omega}^{-1}/\nu}{1 + \boldsymbol{v}_t^\top \boldsymbol{\Omega}^{-1} \boldsymbol{v}_t/\nu} - \boldsymbol{I}_N \right) \right\| < -\delta,$$

for some $\delta > 0$ arbitrarily small.

To conclude, we note that the process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ inherits some properties from those of the filter evaluated at the true parameter value. As a consequence of Lemma 1 and Lemma 2, we obtain the following result.

Lemma 4. *Under the conditions of Lemma 1 and Lemma 2, $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic. Moreover, $\forall m > \nu - \delta$, $\delta > 0$, $\mathbb{E}[\|\boldsymbol{y}_t\|^m] < \infty$.*

Finally, the multi-step forecasts can be straightforwardly obtained as

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

4 Maximum Likelihood Estimation

Let $\ell_t(\boldsymbol{\theta})$ denote the conditional log-likelihood function for a single observation, obtained by taking the logarithm of (3) considered as a function of the parameter $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top)^\top \in \boldsymbol{\Theta} \subset \mathbb{R}^p$, $\boldsymbol{\xi} = (\nu, (\text{vech}(\boldsymbol{\Omega}))^\top, \boldsymbol{\omega}^\top)^\top \in \mathbb{R}^s$, with $s = 1 + \frac{1}{2}N(N+1) + N$ and $\boldsymbol{\psi} = ((\text{vec}\,\boldsymbol{\Phi})^\top, (\text{vec}\,\boldsymbol{K})^\top)^\top \in \mathbb{R}^d$, with $d = (N \times N) + (N \times N)$ and hence, $p = s + d$.

Lemma 3 ensures that any choices of the initial condition $\boldsymbol{\mu}_{1|0} \in \boldsymbol{\mathcal{M}}$ used for starting the filtering process are asymptotically equivalent, such that, once an initial value has been fixed, it is possible to obtain an approximated version of the conditional log-likelihood, $\hat{\ell}_t(\boldsymbol{\theta})$, by replacing $\boldsymbol{\mu}_{t|t-1}$ in $\ell_t(\boldsymbol{\theta})$ by the filtered dynamic location $\hat{\boldsymbol{\mu}}_{t|t-1}$. Thus, for the whole sample, we obtain $\hat{\ell}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta})$ and the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is

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

We now discuss strong consistency and asymptotic normality of the MLE. The following assumptions are standard regularity condition in likelihood theory of non linear observation driven models.

Assumption 2.

1. *The data generating process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary and ergodic.*
2. $\mathbb{E}[\ln \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\mu}\in\boldsymbol{\mathcal{M}}} \| \prod_{j=1}^{k} \boldsymbol{X}_{k-j+1} \|] < 0$ *for* $k \geq 1$.
3. *The parameter space $\boldsymbol{\Theta}$ is compact with* $0 < \nu < \infty$ *and* $\det \boldsymbol{K} \neq 0$.
4. *The true parameter vector $\boldsymbol{\theta}_0$ belongs to the interior of $\boldsymbol{\Theta}$, i.e. $\boldsymbol{\theta}_0 \in$ int$(\boldsymbol{\Theta})$.*

5. $\mathbb{E}[\|\boldsymbol{X}_t \otimes \boldsymbol{X}_t\|] < 1$.

In Assumption 2, we first remark that condition 1 can be replaced by the conditions of Lemma 4, whereas condition 2 ensures that the filtered sequence $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ converges to a stationary ergodic limit sequence, irrespective of the initial conditions. Moreover, conditions 3 and 4 of Assumption 2 ensure the existence of the MLE and the validity of first order asymptotics. Finally, condition 5 of Assumption 2 guarantees the existence of the information matrix.

Theorem 4.1. *Under conditions 1–4 in Assumption 2,*

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

Theorem 4.2. *Under conditions 1–5 in Assumption 2,*

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

where,

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

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

By Theorem 4.1, $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ can be consistently estimated by

$$\widehat{\boldsymbol{\mathcal{I}}}(\hat{\boldsymbol{\theta}}_T) = -\frac{1}{T}\sum_{t=1}^{T}\left[\frac{d^2\hat{\ell}_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_T}\right]. \tag{12}$$

As the dynamic location and its derivatives are nonlinear functions of the parameter $\boldsymbol{\theta}$, the general formula for the second derivatives in (12) evaluated at the stationary and ergodic $\boldsymbol{\mu}_{t|t-1}$, has the form below

$$\begin{aligned}\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top} =& \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} + \left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)^\top \frac{\partial^2\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top}\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\\ &+ \frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}_{t|t-1}^\top}\frac{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}.\end{aligned} \tag{13}$$

To avoid the recursive evaluation of the second derivatives of the dynamic location vector, a simpler consistent estimator can be obtained based on the analytical form of the conditional information matrix $\boldsymbol{\mathcal{I}}_t(\boldsymbol{\theta})$, as in Fiorentini et al. (2003), defined as

$$\boldsymbol{\mathcal{I}}_t(\boldsymbol{\theta}) = -\mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}\right]. \tag{14}$$

Indeed, by the law of iterated expectations, one has

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

Given the assumption of correct specification, the score vector evaluated at the true parameter vector $\boldsymbol{\theta}_0$ forms a martingale difference sequence, so that, under the assumptions of Theorem 4.2, asymptotic results for martingale difference sequences can be applied. In addition, the dynamic location (and its derivatives) are $\mathcal{F}_{t-1}$-measurable functions and therefore, after taking the conditional expectation, the last term in the right-hand-side of equation (13) will cancel out.

It follows that, by Theorem 4.1, $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ can be consistently estimated by

$$\widehat{\boldsymbol{\mathcal{I}}}(\widehat{\boldsymbol{\theta}}_T) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\mathcal{I}}}_t(\widehat{\boldsymbol{\theta}}_T),$$

where $\widehat{\boldsymbol{\mathcal{I}}}_t(\widehat{\boldsymbol{\theta}}_T)$ is the conditional information matrix in (14) evaluated at the filtered dynamic location $\hat{\boldsymbol{\mu}}_{t|t-1}$ and at the MLE $\widehat{\boldsymbol{\theta}}_T$. The analytical form of $\boldsymbol{\mathcal{I}}_t(\boldsymbol{\theta})$, requires lengthy calculations, and so, the detailed derivations are deferred to Section S2 of the online supplementary material. However, since this is one of the main result of the paper, its expression is reported in Appendix B.

4.1 Departure from assumptions

We discuss how the results derived so far change, if some of the underlying assumptions are modified. In particular, we shall consider the impact of misspecification on asymptotic normality and on invertibility, the case when the marginal distributions have different degrees of freedom and the possible extension to explanatory variables.

When the model is not correctly specified, the conditional expectation of the score of the (quasi) log-likelihood function evaluated at the true parameter vector $\boldsymbol{\theta}_0$ is different from zero. Therefore, the conditional expectation of the last term in the right-hand-side of equation (13) will not cancel out. In such case, both the asymptotic and the conditional covariance matrix of the MLE assume the classical sandwich from, see Bollerslev and Wooldridge (1992).

The situation changes if one wants to allow the degrees of freedom vary across the the time series, e.g. by assuming marginal Student's t distributions with ν_i degrees of freedom, $i = 1, \ldots, N$ and then construct the multivariate joint distribution using a Student's t copula (Demarta and McNeil, 2005). In this case, the one-step MLE proposed in the paper is not applicable and a two-step estimation procedure has to be considered. In particular, when margins are estimated parametrically, inferential procedures about the copula model using a ML approach can be tackled with the inference-functions for margins (IFM) approach proposed by Joe (1997). In the first step, the IFM method maximizes each marginal log-likelihood with respect to each unknown parameter vector. These estimates are then used in the second step to maximize the copula log-likelihood function. In this case, the analytical formulae derived for the conditional Fisher information matrix reported in Section 4 are not valid as the information matrices of each marginal variables have to be computed, which inevitably complicates the computational burden. Moreover, if the degrees of freedom parameters turn out to be the same and equal to ν, then this model reduces to the proposed one but the two-step MLE would be less efficient than the one-step. A detailed discussion and comparisons of the asymptotic efficiency of the one-step and two-step MLE for copula functions can be found in Joe (2005). In addition, Harvey (2013), Blasques et al. (2015) and Koopman et al. (2016) pointed out that score-driven models are particularly suited to filter complex nonlinear dynamics, that are frequently experienced in applications, irrespective of the severity of model misspecification.

On the other hand, the invertibility of the multivariate score-driven filters remains valid under misspecification, and that is an essential ingredient to derive the asymptotic properties of the MLE. In practice, the MLE is based on the approximate log-likelihood $\hat{\ell}_t(\boldsymbol{\theta})$, which, for $t = 1, \ldots, T$, is a function of the filtered process $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ initialized at some fixed $\boldsymbol{\mu}_{1|0}$. The key result of Lemma 3 ensures that the impact of starting values for the filters become asymptotically negligible. This allows us to retrieve a reliable approximate sequence for the unobserved dynamic location vector and, eventually, to establish strong consistency and asymptotic normality of the MLE. Similar discussions can be found in Harvey (2013), Blasques et al. (2015) and Koopman et al. (2016).

Finally, in line with the results obtained by Harvey (2013) and Harvey and Luati (2014) for the univariate dynamic location model, it is in principle possible to augment the model with exogenous variables. Provided that the exogenous variables are Markov chains and satisfy a set of separate conditions for their transition mechanism, the conditions for stationarity, ergodicity and invertibility and the asymptotic properties of MLE stated in the paper remain similar. What changes are the closed-form formulæ that lead to the conditional Fisher's information matrix $\boldsymbol{\mathcal{I}}_{t|t-1}(\boldsymbol{\theta})$ as well as the matrix itself. Indeed, the latter has to be extended to account for the uncertainty in the estimation of a larger vector of parameters which includes the coefficients associated with the exogenous variables.

4.2 Computational Aspects

ML estimation and inference are carried out by means of Fisher's scoring method. A strongly reliable algorithm based on analytical formulae for the score vector and the Hessian matrix (reported in Appendix S2) is developed, which can be directly implemented in any statistical package through the following steps:

1. Choose a starting value $\widehat{\boldsymbol{\theta}}_T^{(0)} = (\nu^{(0)}, (\mathrm{vech}(\boldsymbol{\Omega}^{(0)}))^\top, (\boldsymbol{\omega}^{(0)})^\top, (\mathrm{vec}(\boldsymbol{\Phi}^{(0)}))^\top, (\mathrm{vec}(\boldsymbol{K}^{(0)}))^\top)^\top$
2. For $h > 0$, update $\widehat{\boldsymbol{\theta}}_T^{(h)}$ using the scoring rule $\widehat{\boldsymbol{\theta}}_T^{(h+1)} = \widehat{\boldsymbol{\theta}}_T^{(h)} + \big[\widehat{\boldsymbol{\mathcal{I}}}_T(\widehat{\boldsymbol{\theta}}_T^{(h)})\big]^{-1} \widehat{\boldsymbol{s}}_T(\widehat{\boldsymbol{\theta}}_T^{(h)})$, where $\boldsymbol{s}_T(\boldsymbol{\theta}) = \sum_{t=1}^T \frac{d\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ *and* $\boldsymbol{\mathcal{I}}_T(\boldsymbol{\theta}) = -\sum_{t=1}^T \mathbb{E}_{t-1}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta} d\boldsymbol{\theta}^\top}\right]$.
3. Repeat until convergence, i.e., $\|\widehat{\boldsymbol{\theta}}_T^{(h+1)} - \widehat{\boldsymbol{\theta}}_T^{(h)}\| / \|\widehat{\boldsymbol{\theta}}_T^{(h)}\| < \delta$ for some fixed $\delta > 0$.

The analytical expressions for the conditional information matrix used in step 2 are given in Appendix B, whereas the formulae of the score vector together with the detailed derivations are deferred to Section S2 of the online supplementary material.

5 Simulation Study

In this Section we report the details of an extensive simulation study aimed to investigate two main aspects of the specification, i.e. the impact of initial conditions on the filtered estimates, addressed in Section 5.1, and the effect of potential misspecification of the underlying distribution, discussed in Section 5.2.

5.1 Impact of Initial Conditions

We aim to assess the sensitivity of the proposed multivariate filtering procedure to initial conditions. In particular, we are interested in measuring the sensitivity of the filtering and estimation procedures to $\boldsymbol{\mu}_{1|0}$ and $\widehat{\boldsymbol{\theta}}_T^{(0)}$, respectively.

As far as the former is concerned, as discussed in Section 3, in practice, the multivariate score-driven filter is initialized at some fixed $\boldsymbol{\mu}_{1|0}$, to eventually retrieve the approximate sequence $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ from the observed data using the filtering recursion in equation (7). To highlight the relevance of the invertibility property from an applied point of view, we conduct a Monte Carlo experiment that makes also use of a real quarterly macroeconomic time series dataset. Specifically, we simulate $N = 3$ time series $\{\boldsymbol{y}_t\}_{t=1}^T$ of $T = 100$ observations from a multivariate Student's t distribution with $\nu = 5$ degrees of freedom. To generate time series with empirically relevant dynamics and correlation structure, we estimate the parameters of our multivariate model for three series coming from Greek's macroeconomic quarterly data, namely, the log-differences of GDP, import and export data, from 1995 to 2021. The log-differenced series are shown in Figure 2 (left panel). The dimension of the simulated time series equals that of the multivariate real dataset.

To investigate the "forgetting memory" property of the sequence $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ initialized at some fixed value $\boldsymbol{\mu}_{1|0}$ along with the relevance of working with invertible multivariate DCS-t models, in the right panel of Figure 2 we show the impact of the initialization on a number of fixed points.

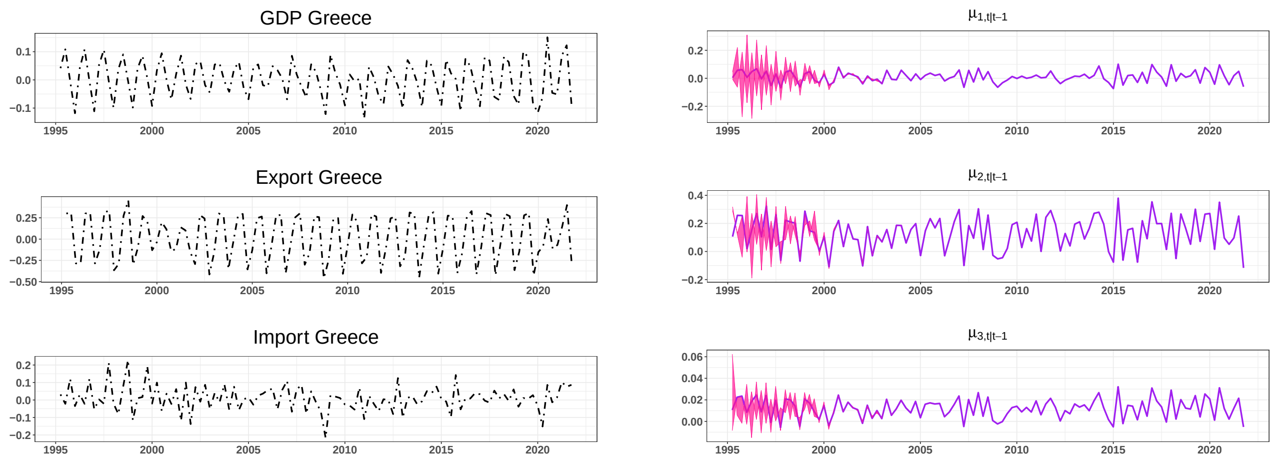

Figure 2: log-difference of the original series (left panel), and filtered dynamic locations for the case satisfying the invertibility condition (right panel). The pink area at the beginning of the series in the right panel indicate the different initial conditions used to start the multivariate filtering procedure, whereas the purple lines, depict the stable trajectories of the filtered dynamic locations to which the filters converge for each of the fixed initial conditions.

It is interesting to note that the chaotic behaviour of the initialized filtering process $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ vanishes after a fair, low, number of data points and eventually converges to the unique "true" and stationary counterpart. This Monte Carlo experiment highlights the fact that the inference procedure based on the MLE procedure remains reliable even with a rather low and realistic number of observations.

Besides initializing the filtered sequence, it is also necessary to initialise the estimation procedure detailed in Subsection 4.2 at some starting value $\widehat{\boldsymbol{\theta}}_T^{(0)}$. To this end, we follow the approach suggested in Fiorentini et al. (2003). First, a consistent estimator of the restricted version of the parameter vector $\tilde{\boldsymbol{\theta}}_T$ is obtained by the Gaussian quasi-ML procedure in Bollerslev and Wooldridge (1992). Second, a consistent method of moments is adopted for the degrees of freedom ν, by making use of the empirical coefficient of excess kurtosis $\tilde{\kappa}$ on the standardized residuals and of the relation $\tilde{\nu} = (4\tilde{\kappa} + 6)/\tilde{\kappa}$. Convergence is fast, in that usually few iterations of that procedure are needed, which makes scoring methods particularly appealing for estimation purposes.

5.2 Model Misspecification

The results derived in the paper assume a correct specification of the underlying distribution. To demonstrate the reliability of the proposed score-driven filters under misspecification of the innovation density, we perform a simulation experiment where we assume that the data generating process follows: a skew elliptical distribution, to account for asymmetry (Section 5.2.1), a Meta t distribution, to account for heterogeneous degrees of freedom (Section 5.2.2), a Gaussian distribution, to assess robustness in the estimation of the degrees of freedom parameter (Section 5.2.3).

5.2.1 Asymmetry

We first consider an asymmetric distribution as the true data generating process. Specifically, we consider the case where the error vector is skew-elliptically distributed, see Azzalini (2005), i.e. it is given by the following equations:

$$\boldsymbol{y}_t = \boldsymbol{\xi}_{t|t-1} + \boldsymbol{V}^{1/2}\boldsymbol{W}_t, \tag{15}$$

where, following the notation of Azzalini (2005), $\boldsymbol{W}_t$ follows a N-dimensional skewed-normal distribution, i.e., $\boldsymbol{W}_t \sim \text{SN}(\boldsymbol{0}_N, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$ with some non-singular correlation matrix $\bar{\boldsymbol{\Omega}} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{\alpha} \in \mathbb{R}^N$, which regulates the slant of the density, and is independent of $\boldsymbol{V} \sim \chi^2_\nu/\nu$. We assume that the N-dimensional dynamic location vector $\boldsymbol{\xi}_{t|t-1}$ evolves as a vector autoregressive model of order one

$$\boldsymbol{\xi}_{t|t-1} = \bar{\boldsymbol{\Phi}}\boldsymbol{y}_{t-1}, \tag{16}$$

with $\bar{\boldsymbol{\Phi}} \in \mathbb{R}^{N\times N}$ denoting the matrix that contains the autoregressive coefficients. We set $N = 3$, and moreover, to evaluate the benefit the benefit from capturing the cross-series correlations with a multivariate model, we generate time series in case of full correlation matrix with $\text{vec}\bar{\boldsymbol{\Omega}} = (1, 0.3, 0.4, 0.3, 1, 0.3, 0.4, 0.3, 1)^\top$, together with the case that $\text{vec}\bar{\boldsymbol{\Omega}} = \text{vec}\boldsymbol{I}_3$, that is, with zero correlation between the variables. For the level of kurtosis we set $\nu = 6$ and we specify different values of $\boldsymbol{\alpha}$, in order to assess the robustness of the proposed filtering procedure with respect to either negative and positive skewness, together with no skewness (i.e. symmetric case). More precisely, in our simulations, we consider three cases, negative skewness with $\boldsymbol{\alpha} = (\text{-}0.5,\text{-}0.5,\text{-}0.5)^\top$, no skewness with $\boldsymbol{\alpha} = (0, 0, 0)^\top$, and positive skewness, $\boldsymbol{\alpha} = (0.5, 0.5, 0.5)^\top\}$. As concern the matrix of autocorrelation coefficients $\bar{\boldsymbol{\Phi}}$, we consider $\text{vec}\bar{\boldsymbol{\Phi}} = (0.8, 0.2, 0.1, 0.1, 0.5, 0.1, 0.1, 0.2, 0.7)^\top$, such that its spectral radius is $\rho(\bar{\boldsymbol{\Phi}}) \approx 0.95$ which ensures a high rate of persistence in the simulated time series. Therefore, from equations (15) and (16) we generate $M = 500$ time series of length $T = 1000$. For all the $M = 500$ replication, we estimate both the multivariate DCS-t model described in Section 2 and the univariate DCS-t model introduced by Harvey and Luati (2014), and we compare the filtered paths of $\{\hat{\mu}_{i,t|t-1}\}_{t=1}^{1000}$ for $i = 1, 2, 3$, provided by the two approaches against the true paths simulated according to equations (15) and (16). To evaluate the performance of the two models, we consider the mean squared error (MSE) and the mean absolute error (MAE),

$$MSE = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\mu}_{i,t|t-1} - \xi_{i,t|t-1}\right)^2, \qquad MAE = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|\hat{\mu}_{i,t|t-1} - \xi_{i,t|t-1}\right|. \tag{17}$$

The results in Panel A of Table 1 present a clear outcome: In case of correlated time series with negative, absence or positive skewness, and across the two performance measures, the new multivariate score-driven model provides more accurate filtered paths than the univariate counterpart of the unobserved dynamic location vector. Figure 3 further corroborates this feature of the multivariate model, since the median of the 500 filtered paths of $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t=1}^{1000}$ are very close to the medians of the 500 true paths $\{\boldsymbol{\xi}_{t|t-1}\}_{t=1}^{1000}$. On the other hand, the results reported in Panel B of Table 1 are less sharp, which clearly confirm the fact that the multivariate DCS-t model cannot benefit from capturing

Table 1: *Mean squared error (MSE) and mean absolute error (MAE) between the median of the true simulated trajectories, and the median of the filtered dynamic locations provided by the multivariate DCS-t (mDCS-t) and the univariate DCS-t (uDCS-t) across all the $M = 500$ trajectories with $T = 1000$ generated according to equations* (15) *snd* (16)*. The results given in case of presence of both correlated (Panel A) and uncorrelated (Panel B) time series. In each case, we consider negative, absence and positive degrees of skewness.*

Panel A: Correlated time series						
	MSE			*MAE*		
	Neg. Skew.	*No Skew.*	*Pos. Skew.*	*Neg. Skew.*	*No Skew.*	*Pos. Skew.*
mDSC-t	**1.1114**	**1.4725**	**1.0208**	**1.1423**	**2.0983**	**2.5099**
uDSC-t	3.7985	3.0987	3.8645	3.0568	3.6632	3.7654

Panel B: Uncorrelated time series						
	MSE			*MAE*		
	Neg. Skew.	*No Skew.*	*Pos. Skew.*	*Neg. Skew.*	*No Skew.*	*Pos. Skew.*
mDSC-t	**0.9381**	**0.0620**	**0.8686**	**1.5112**	1.5113	**1.2743**
uDSC-t	1.2111	0.0987	0.9234	1.8393	**1.5101**	1.5648

the cross-series correlation.

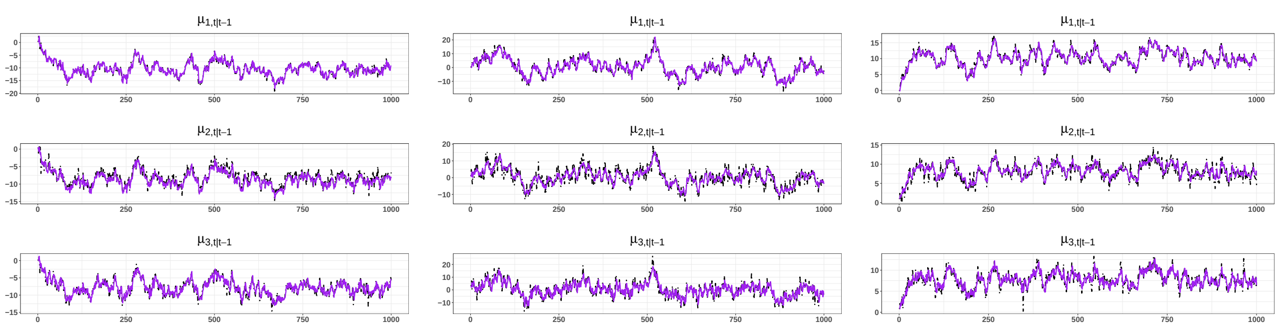

Figure 3: Median of the true simulated trajectories (dot-dashed black line), and median of the filtered dynamic locations (purple line) across the $M = 500$ trajectories with $T = 1000$ data points, generated according to equations (15) snd (16), for the case of negative skewness (left panel) with $\boldsymbol{\alpha}$ = (-0.5,-0.5,-0.5)$^\top$, no skewness (middle panel) with $\boldsymbol{\alpha} = (0, 0, 0)^\top$ and positive skewness (right panel) with $\boldsymbol{\alpha} = (0.5, 0.5, 0.5)^\top$.

5.2.2 Heterogeneous degrees of freedom

We investigate the flexibility of the multivariate filter when the time series are correlated and heavy-tailed but with heterogeneous degrees of freedom. We carry on a Monte Carlo experiment, where we generate the data from a dynamic version of the Meta-t distribution introduced by Fang et al. (2002). More precisely, we consider the same observation equation defined in (15) but we replace the skewed distributional assumption $\boldsymbol{W}_t \sim \text{SN}(\mathbf{0}_N, \bar{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$ with $\boldsymbol{W}_t \sim Meta\text{-}t(\mathbf{0}_N, \bar{\boldsymbol{\Omega}}, \nu, \nu_1, \ldots, \nu_N)$, i.e., a Meta-$t$ distribution with $\mathbf{0}_N$ location vector and $\bar{\boldsymbol{\Omega}}$ correlation matrix. Here ν denotes the degrees of freedom of the Student's t copula whereas $\nu_1, \ldots, \nu_N$ denote the (possibly heterogeneous) degrees of freedom of the univariate Student's t distribution of the marginals. Also in this case, we assume that the dynamic location vector $\boldsymbol{\xi}_{t|t-1}$ evolves as a zero-mean vector autoregressive model as given by equation (16), and we set again $N = 3$, $\text{vec}\bar{\boldsymbol{\Omega}} = (1, 0.3, 0.4, 0.3, 1, 0.3, 0.4, 0.3, 1)^\top$, $\nu = 6$ and $\text{vec}\bar{\boldsymbol{\Phi}} = (0.8, 0.2, 0.1, 0.1, 0.5, 0.1, 0.1, 0.2, 0.7)^\top$. However, since we the interest is in measuring the robustness of the proposed multivariate filtering procedure under heterogeneous degrees of freedom, we consider four different cases: (i) Correlated time series with same degrees of freedom, such that $\nu = \nu_1 = \nu_2 = \nu_3 = 6$, (ii) correlated time series with different degrees of freedom so that $\nu = 6$ whereas $\nu_1 = 5$, $\nu_2 = 10$ and $\nu_3 = 15$ for the marginal distributions, and furthermore, (iii) uncorrelated time series with same degrees of freedom as in case (i), and lastly (iv) uncorrelated time series with different degrees of freedom as in case (ii).

Similarly to the previous simulation exercise, from equations (15) and (16) we generate $T = 1000$ observations for $\boldsymbol{y}_t$ and then, for each $t = 1, \ldots, 1000$, we repeat the generation $M = 500$ times. Again, we estimate both our multivariate DCS-t model described in Section 2 of the paper and the univariate DCS-t model introduced by Harvey and Luati (2014), and for the $M = 500$ replications, we compare the filtered paths of $\{\hat{\mu}_{i,t|t-1}\}_{t=1}^{1000}$ for $i = 1, 2, 3$, provided by the two approaches against the true paths simulated according to equations (15) and (16) with the new distributional assumptions. Then, to evaluate the performance of the two models, we consider the mean squared error (MSE) and the mean absolute error (MAE), as defined in (17).

Table 2: *Mean squared error (MSE) and mean absolute error (MAE) between the median of the true simulated trajectories, and the median of the filtered dynamic locations provided by the multivariate DCS-t (mDCS-t) and the univariate DCS-t (uDCS-t) across all the $M = 500$ trajectories with $T = 1000$ generated according to equations* (15) *snd* (16)*. Results are given for* $\boldsymbol{W}_t \sim \text{Meta-t}(\mathbf{0}_N, \bar{\boldsymbol{\Omega}}, \nu, \nu_1, \ldots, \nu_N)$ *with same degrees of freedom $\nu = \nu_1 = \nu_2 = \nu_3 = 6$, and different degrees of freedom, such that $\nu = 6$ and $\nu_1 = 5$, $\nu_2 = 10$ and $\nu_3 = 15$.*

Panel A: Correlated time series				
	MSE		*MAE*	
	Same DoF	*Different DoF*	*Same DoF*	*Different DoF*
mDSC-t	**1.4644**	**1.4584**	**0.9254**	**0.9189**
uDSC-t	2.0876	2.9214	2.9871	2.6541
Panel B: Uncorrelated time series				
	MSE		*MAE*	
	Same DoF	*Different DoF*	*Same DoF*	*Different DoF*
mDSC-t	**0.9048**	1.3581	**0.8882**	1.3721
uDSC-t	0.9374	**0.9647**	0.8930	**1.0312**

The results in Table 2 confirm that, in case of correlated time series and for all two performance measures, the new multivariate score-driven model provides a more accurate filtered paths than the univariate counterpart of the unobserved dynamic location vector, even in the case of heavy-tailed data with heterogeneous degrees of freedom in the marginals. For a graphical inspection, in Figure 4 we plot the median of the 500 filtered paths of $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t=1}^{1000}$, which are very close to the median of the 500 true paths $\{\boldsymbol{\xi}_{t|t-1}\}_{t=1}^{1000}$. Therefore, we conclude that the multivariate model provides a good trade-off between allowing for correlation across the time series and imposing the same degrees of freedom to all the series. However, we now note that in case of uncorrelated time series and different degrees of freedom, the multivariate DCS-t is heavily misspecified, and then, the performances of the multivariate filter start to deteriorate. Obviously, this is not the case for the univariate DCS-t, as the univariate filers do not suffer from a multivariate structure and then, they can adapt to each of the uncorrelated time series.

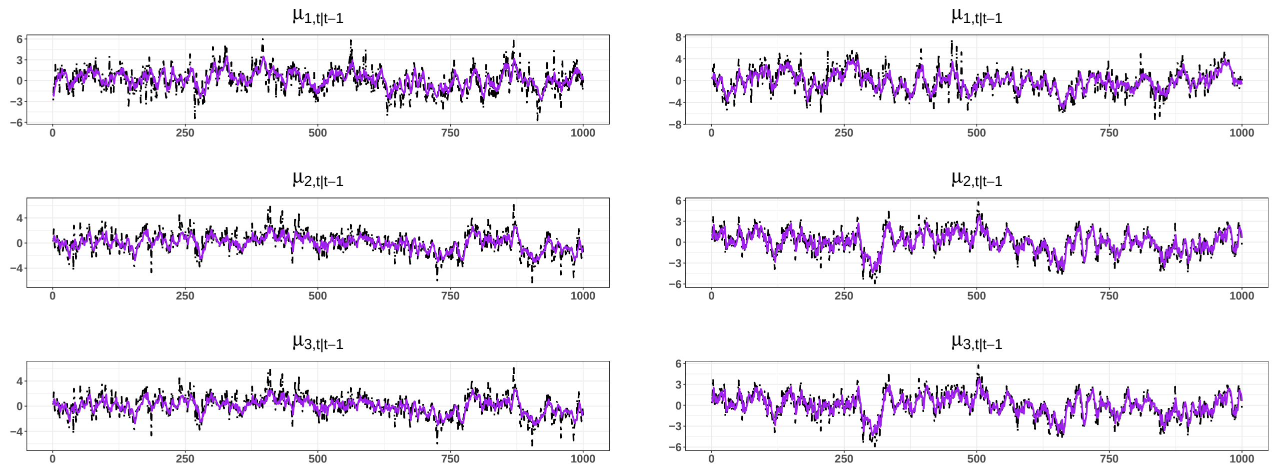

Figure 4: Median of the true simulated trajectories (dot-dashed black line), and median of the filtered dynamic locations (purple line) across all the $M = 500$ trajectories with $T = 1000$ generated according to equations (15) snd (16), for the case of Meta-t distribution with same degrees of freedom $\nu = \nu_1 = \nu_2 = \nu_3 = 6$ (left panel), and different degrees of freedom, such that $\nu = 6$ and $\nu_1 = 5$, $\nu_2 = 10$ and $\nu_3 = 15$ (right panel).

5.2.3 Finite sample properties and Gaussian assumption

We conclude by discussing a Monte Carlo study aimed to assess the finite sample properties of the MLE based on the Fisher's scoring method detailed in the above section. The results are reported in section S1 of the online supplementary material to save space. Here we highlight that our approach performs well in terms of bias and root mean square errors for a wide range of time series, from the most severe heavy-tailed case (i.e., ν very small) to the Gaussian case (i.e. for $\nu \to \infty$). Most importantly, when the underlying process is Gaussian, the degrees of freedom parameter is always estimated of the order of hundreds. In addition, we note that the algorithm delivers satisfactory results even when the number of iterations is limited to ten rounds.

6 Empirical Analysis of Homescan Data Consumer Prices

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

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

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

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

To illustrate the potential contribution of the proposed method, we exploit a data-set that has been recently used to evaluate the effects of a tax on sugar-sweetened beveraged introduced in France in

2012 Capacci et al. (2019). Our data for the empirical application were previously used for evaluating the impact of the French soda tax through a quasi-experimental difference-in-difference design, where the Italian data served as a counter-factual, see Capacci et al. (2019), and consists of weekly scanner price data for food and non-alcoholic drinks. The data were collected in a single region, within the Italian GfK homescan consumer panel, based on purchase information on 318 households surveyed in the Piedmont region, over the period between January 2011 and December 2012. The regional scope and the relatively small sample provide an ideal setting to test the applicability and effectiveness of the multivariate filtering approach.

Table 3: Average unit values, € per kilogram, Piedmont homescan data (standard deviations in brackets)

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

6.1 Data

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

Average unit values are shown in Table 3. Food and non-alcoholic drinks are composite aggregates, hence they are potentially subject to fluctuations in response to changes in the consumer basket even when prices are stable. Instead, Coca-Cola is a relatively homogeneous good, with little variability due to different packaging sizes.

6.2 Results

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

$$\hat{\boldsymbol{\omega}} = \begin{bmatrix} \underset{(0.000)}{0.443} \\ \underset{(0.000)}{4.394} \\ \underset{(0.000)}{-1.070} \end{bmatrix} \qquad \hat{\boldsymbol{\Phi}} = \begin{bmatrix} \underset{(0.011)}{0.839} & \underset{(0.002)}{0.015} & \underset{(0.005)}{0.007} \\ \underset{(0.059)}{-0.528} & \underset{(0.009)}{0.912} & \underset{(0.025)}{0.342} \\ \underset{(0.020)}{0.222} & \underset{(0.003)}{0.023} & \underset{(0.009)}{0.847} \end{bmatrix} \qquad \hat{\boldsymbol{K}} = \begin{bmatrix} \underset{(0.017)}{0.442} & \underset{(0.003)}{-0.023} & \underset{(0.007)}{0.007} \\ \underset{(0.079)}{0.334} & \underset{(0.014)}{0.216} & \underset{(0.038)}{-0.631} \\ \underset{(0.030)}{-0.290} & \underset{(0.005)}{-0.098} & \underset{(0.014)}{-0.014} \end{bmatrix}$$

where the values in parenthesis are the standard errors and with

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

The estimated degrees of freedom are approximately 7. We remark that the assumption of a (conditional) multivariate Student's t distribution implies that all the univariate marginal distributions are tail equivalent, see Resnick (2004). This requires the implicit underlying assumption that the level of heavy-tailedness across the observed time series vector is fairly homogeneous. To investigate this issue, and for the sake of comparisons, we have carried out a univariate analysis, as in Harvey and Luati (2014), from which it resulted that the estimated degrees of freedom were very low for Coca-Cola (about 4) and medium size (smaller than 30) for the other two series, as expected. Hence, the multivariate score-driven model developed in the paper reveals to be a good compromise between a multivariate non-robust filter, based on a linear Gaussian model, and a robust univariate estimator. Indeed, a multivariate Portmanteau test on the residuals obtained from the three univariate models is carried out to test the null hypothesis $H_0 : \boldsymbol{R}_1 = \cdots = \boldsymbol{R}_m = \boldsymbol{0}$, where $\boldsymbol{R}_i$ is the sample cross-correlation matrix for some $i \in \{1, \ldots, m\}$ against the alternative $H_1 : \boldsymbol{R}_i \neq \boldsymbol{0}$. The results of Table 4 indicate rejection of the null hypothesis of absence of serial dependence in the trivariate series at the 5% significance level.

Table 4: *Multivariate Portmanteau test.*

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

We also remark that the estimated degrees of freedom close to 7 rule out the hypothesis that the data come from a linear Gaussian state-space model, in which case the estimated degrees of freedom would be definitely higher. Nevertheless, we have fitted a misspecified linear Gaussian state-space model estimated with the Kalman filter and, as expected, along with a higher sensitivity to extreme values, in particular in the last period of the Coca-Cola series, likelihood and information criteria are in favour of the multivariate model based on the conditional Student's t distribution.

Table 5: *Likelihood, Akaike and Bayesian information criteria.*

	log-*Lik*	*AIC*	*BIC*
KF	241.16	-434.32	-370.85
mDCS-t	**257.93**	**-465.69**	**-402.23**
uDCS-t	230.87	-441.74	-392.07

To enforce our results, we explore the robustness to misspecification with respect to a univariate model and in particular to the univariate model by Harvey and Luati (2014). The aim is to assess the trade-off between (i) allowing for correlations among series at the price of imposing one common degree of freedom, against (ii) a univariate approach that allows for heterogeneous degrees of freedom but does not directly model the series correlation structure.

To this end, we perform the specification test for nonlinear dynamic models discussed in Bai and Chen (2008) and Kheifets (2015). Specifically, using the notation of Kheifets (2015), we test the following null hypothesis:

H_0^M: The distribution of $\boldsymbol{y}_t|\mathcal{F}_{t-1}$ is in the parametric family $F_t(\,\cdot\,|\mathcal{F}_{t-1}, \boldsymbol{\theta})$ for some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

In our case, this null hypothesis translates in testing if the N-vector stochastic process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ has multivariate conditional t-distribution with ν degrees of freedom, $\boldsymbol{\mu}_{t|t-1}$ dynamic conditional location and $\boldsymbol{\Omega}$ scale matrix. It is well-known that, analogously to the multivariate Gaussian case, when a random vector has a multivariate t-distribution, any of its subvector has multivariate t-distribution, see Zellner (1971) or Kotz and Nadarajah (2004). Furthermore, as discussed in Spanos (1986) and Ding (2016) the conditional distributions are also multivariate t.

To apply test, we need to compute the univariate series of probability integral transforms $\{U_{tk}\}_{t=1}^{T}$, $k = 1, \dots, N$. Then, by the location-scale representation of the model in equation (1), we have

$$U_{tk} = \mathbb{P}(\epsilon_{tk} \le \epsilon) = t_{\nu_k}\left(\frac{y_{tk} - \mu_{t|t-1,k|k-1}}{\sqrt{\Omega_{k|k-1}}}\right),$$

with $\boldsymbol{y}_{t,k-1} = (y_{t,1}, \dots, y_{t,k-1})^\top$ for $k = 1, \dots, N$, conditional mean $\mathbb{E}[y_{tk}|\boldsymbol{y}_{t,k-1}, \mathcal{F}_{t-1}] = \mu_{t|t-1,k|k-1}$, conditional variance $\mathbb{V}[y_{tk}|\boldsymbol{y}_{t,k-1}, \mathcal{F}_{t-1}] = \frac{\nu}{\nu-2}\Omega_{k|k-1}$, and $\nu_k = \nu + N - 1$ degrees of freedom.

Explicit formulae for the multivariate probability integral transform of Student's t distributed random vector can be found in Bai and Chen (2008) and Ding (2016). In particular, by applying the results of Ding (2016) into the multivariate Student's t setting, we have that

$$y_{tk}|\boldsymbol{y}_{t,k-1}, \mathcal{F}_{t-1} \sim t_{\nu_k}(\mu_{t|t-1,k|k-1}, \Omega_{k|k-1}),$$

where for $k = 1, \ldots, N$

$$\mu_{t|t-1,k|k-1} = \mu_{t|t-1,k} + \boldsymbol{\Omega}_{k,k-1}\Omega_{k,k}^{-1}\boldsymbol{\Omega}_{k-1,k}(y_{t,k-1} - \boldsymbol{\mu}_{t|t-1,k-1}),$$

with $\mathbb{E}[y_{tk}|\mathcal{F}_{t-1}] = \mu_{t|t-1,k}$ and $\mathbb{V}[y_{tk}|\boldsymbol{y}_{t,k-1}, \mathcal{F}_{t-1}] = \frac{\nu}{\nu-2}\Omega_{k,k}$, and moreover, with the conditional vectors $\mathbb{E}[\boldsymbol{y}_{t,k-1}|\mathcal{F}_{t-1}] = \boldsymbol{\mu}_{t|t-1,k-1}$ and $\mathbb{C}ov[y_{tk}, \boldsymbol{y}_{t,k-1}|\mathcal{F}_{t-1}] = \frac{\nu}{\nu-2}\boldsymbol{\Omega}_{k,k-1}$ conditional matrices, such that

$$\Omega_{k|k-1} = \frac{\nu + (\boldsymbol{y}_{t,k-1} - \boldsymbol{\mu}_{t|t-1,k-1})^\top \boldsymbol{\Omega}_{k-1,k-1}^{-1}(\boldsymbol{y}_{t,k-1} - \boldsymbol{\mu}_{t|t-1,k-1})}{\nu + k - 1}(\Omega_{k,k} - \boldsymbol{\Omega}_{k,k-1}\boldsymbol{\Omega}_{k-1,k-1}^{-1}\boldsymbol{\Omega}_{k-1,k}).$$

By stacking the U_{tk} in a vector, for $t = 1, \ldots, T$ and $k = 1, \ldots, N$, we obtain a new vector $\boldsymbol{U}_\tau$ of univariate random variables of length $T \times N$ which, under H_0^M are *IID* and uniformly distributed on $[0, 1]$. This allow us to apply the test statistics based on $V_{2T,j}$ for $j = 1, 2, \ldots$ of Kheifets (2015), that tests the j-lag pairwise independence using the fact that, under H_0^M, the distance to this null hypothesis is measured by

$$V_{2T,j}(r) = \frac{1}{\sqrt{T-j}} \sum_{t=j+1}^{T} \left(\mathbb{1}_{\{\boldsymbol{U}_\tau \leq r_1\}} \mathbb{1}_{\{\boldsymbol{U}_{\tau-j} \leq r_2\}} - r_1 r_2 \right),$$

for each $r_1, r_2 \in [0, 1]^2$.

Since, for all $t = 1, \ldots, T$ and $k = 1, \ldots, N$, U_{tk} is not observable, we estimate $\hat{U}_{tk}$ by replacing each $\mu_{t|t-1,k|k-1}$ by the filtered counterpart $\hat{\mu}_{t|t-1,k|k-1}$, and the static parameters $\Omega_{k|k-1}$ and ν by their MLE $\hat{\Omega}_{k|k-1}$ and $\hat{\nu}$, respectively. Based on $\hat{U}_{tk}$ we eventually get $\hat{V}_{2T,j(r)}$.

To test the null hypothesis H_0^M, we consider both Cramer-von Mises (CvM) and Kolmogorov-Smirnov (KS) statistics proposed by Kheifets (2015), defined as

$$D_{2T}^{CvM} = \int_{[0,1]^2} \hat{V}_{2T,j}(r)\mathrm{d}r, \qquad D_{2T}^{KS} = \sup_{[0,1]^2} \left| \hat{V}_{2T,j}(r) \right|, \tag{18}$$

whose approximate distribution is obtained by a parametric bootstrap procedure as described in Kheifets (2015). Results are gathered in Table 6.

Table 6: *The p-values of specification tests for the multivariate DCS-t (mDCS-t) and the univariate DCS-t (uDCS-t).*

	D_{2T}^{CvM}	D_{2T}^{KS}
mDCS-t	0.144	0.254
uDCS-t	0.001***	0.001***

From Table 6 we conclude that the tests are not rejected in the multivariate DCS-t case at any usual significance level, such as 1%, 5% or 10%. Conversely, H_0^M is rejected at any of these significant level in the univariate DCS-t case, suggesting that neither the full dynamics nor auto-correlation structures are captured by the simpler univariate approach.

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

In order to explore the relation among the series, we implement an impulse response analysis. We follow the approach of Gallant et al. (1993) and Lin (1997), and define the impulse response function for dynamic conditional locations as the impact of a small perturbation of the i-th variable on the future predicted location vector $\boldsymbol{\mu}_{t+1|t}$, i.e., the i-th element of the vector $\boldsymbol{y}_t$, for $i = 1, \ldots, N$, i.e.

$$\boldsymbol{Q}_{s,t}(i) = \mathbb{E}[\boldsymbol{\mu}_{t+1|t}|\boldsymbol{y}_t = \boldsymbol{y}_0 + \boldsymbol{\delta}_i, \mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{\mu}_{t+1|t}|\boldsymbol{y}_t = \boldsymbol{y}_0, \mathcal{F}_{t-1}],$$

where $\boldsymbol{y}_0$ is a vector of zeros defined as the initial condition without any impulse, and $\boldsymbol{\delta}_i$ is a zero vector with i-th element equal to 1 (or -1). The nonlinear impulse are computed by using the local projections approach of Óscar Jordá (2005), and the confidence bands are obtained by using the Newey-West corrected standard-errors, see Newey and West (1987).

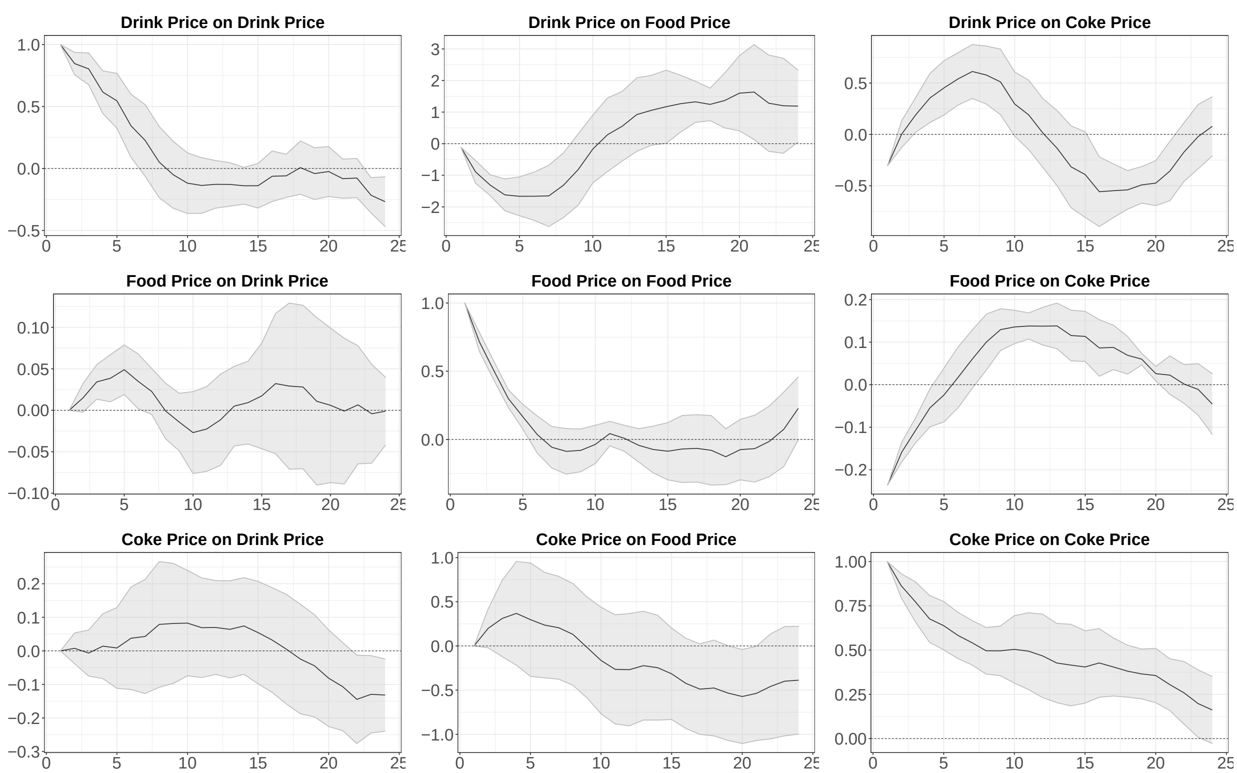

Figure 5: Estimated impulse response functions of the filtered $\hat{\boldsymbol{\mu}}_{t|t-1}$ for a unit shock.

Figure 5 shows the estimated impulse response functions. What emerges is a negative relation between drink and food prices: a unit shock in drink prices will produce a negative shock in food prices. This may adjustments in purchasing decisions by the households aimed at mitigating the rising cost of their shopping basket. This would be evidence that univariate signals are likely to suffer from the unit value bias. Similarly, a non trivial negative relation exists between food and Coca-Cola prices. A unit shock on food prices yields a concurrent negative impact on Coca-Cola prices, which is also noted from the analysis of the cross-correlations. As one might expect, a positive correlation exists between Coca-Cola prices and drink prices, as the former product belongs to the latter category. Instead, unit shocks on food prices seem to have negligible correlation (if any) on drink prices.

6.3 Interpretation

Figure 6 shows the original unit value time series and the corresponding signals extracted through the multivariate score-driven filter. Noise and outliers, as well as some irregular periodic pattern, are clearly visible in the drinks and food series. On the other hand, the Coca-Cola series is relatively regular, with the exception of few peaks, including a couple of large outliers in the second year. Given the homogeneous nature of the good, it is reasonable to believe that those extreme values are the results of measurement error. The estimates illustrate an effective noise reduction and return patterns that are smoother and more consistent with a regular price time series. As one would expect, the Coca-Cola DCS-t series is very flat, and suggests a relatively stable price over the two-years time window, with no outliers.

Figure 7 shows the monthly natural logarithm differences of the raw homescan prices (HSP) and the estimated signals, together with changes in the official Regional CPIs (R-CPI) for food and non-alcoholic drinks, whereas no CPI to the brand detail is produced. The R-CPIs are provided by the National Statistical Institute (ISTAT). They have a monthly frequency and are built with a traditional survey-based approach on retailers. The comparison between the score-driven filtered values and the R-CPIs is purely indicative, as the unit values from the homescan data are weekly, whereas the official CPIs are monthly. This frequency difference may lead to biased comparisons Diewert et al. (2016). Nevertheless, the graphs confirm that the score-driven signals are effective in reducing the noise in the data. This is especially true for the food series, whose CPIs are more volatile compared to drinks. The correlation between the raw homescan log-differenced unit value and the log-differenced food CPI is 0.05, against 0.44 when the filtered time series is considered. For the non-alcoholic drinks price series the gain is less conspicuous, as prices evolve very regularly over the time window. Still, an inexistent correlation between the HSP and the R-CPI (-0.02) turns into a positive one (+0.11) when considering the score-driven estimates and the R-CPI.

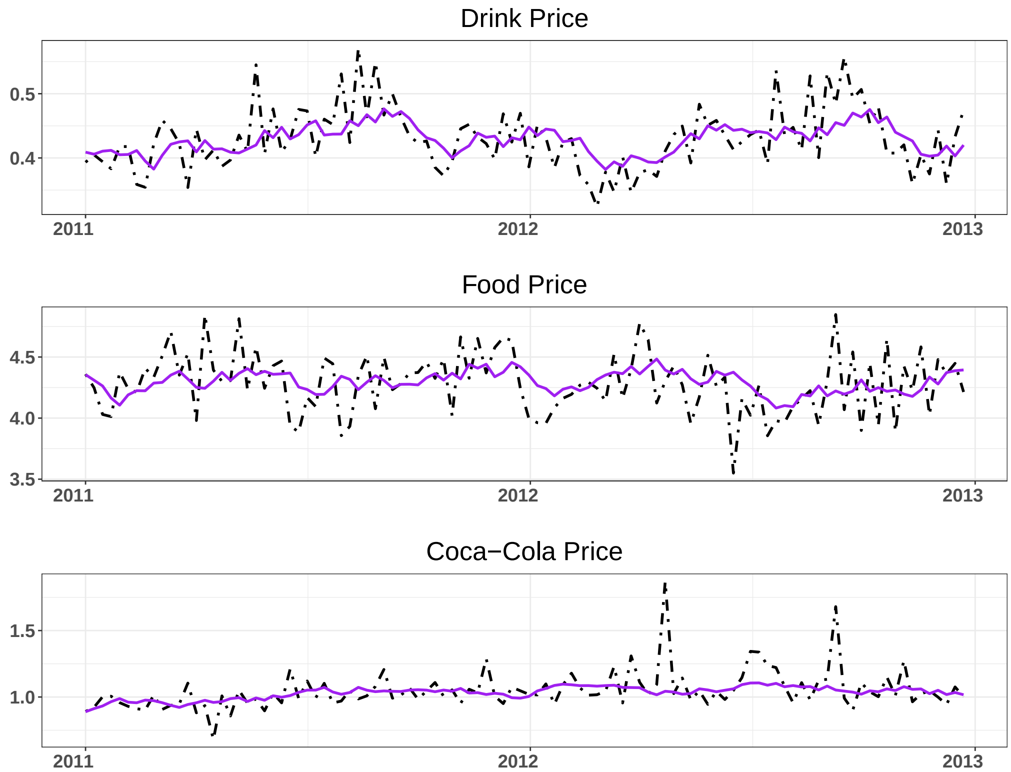

Figure 6: Original series (dotted line) and estimated signals

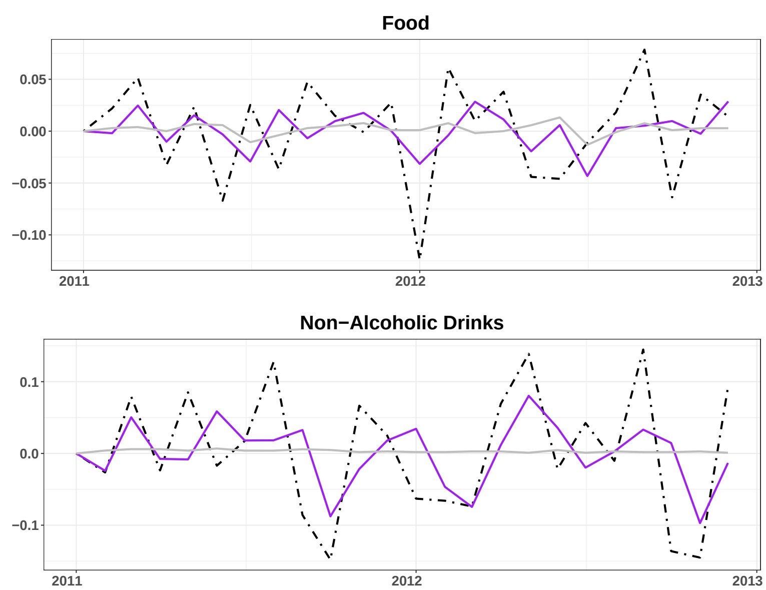

Figure 7: Raw unit value series (dotted line), estimated signal (purple line) and Regional CPIs (log differences, grey line)

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

7 Concluding Remarks

We developed a nonlinear and multivariate dynamic location filter which enables the extraction of reliable signals from vector processes affected by outliers and possibly non-Gaussian errors. Its peculiarity lies in the specification of a score-robust updating equation for the time-varying conditional location vector. Compared to the existing literature on observation driven models for time varying parameters, the model has two innovative features: (a) it extends the univariate first-order dynamic conditional location score by Harvey and Luati (2014) to the multivariate setting; and (b) it extends the dynamic model for time varying volatilities and correlations by Creal et al. (2011) to the location case.

We derived the stochastic properties of the filter: bounded moments, stationarity, ergodicity, and filter invertibility. Parameters are estimated by the method of maximum likelihood and we provided analytic closed-form formulæ for the score vector and the Hessian matrix. The latter can be directly used to design a scoring algorithm that is naturally much faster than the usual ones based on numerical derivatives. Consistency and asymptotic normality have been proved and a large scale Monte-Carlo study analyzed the finite sample properties of the estimation procedure. In particular, we investigated: the sensitivity to initial conditions of the filtered sequence and of the estimation algorithm; the robustness against misspecification in several directions, such as asymmetry of the underlying distribution, heterogeneous degrees of freedom; the performance compared to its corresponding univariate counterpart. In the case when the degrees of freedom tend to infinity, or, in practice, their estimate is of the order of hundreds, our specification converges to a linear and Gaussian model.

The empirical application showed that robust filtering may lead to satisfactory estimates of price signals from homescan data, in the case when the multivariate dimension is low. We contribute to research in this area with two promising results. First, we show that robust modeling allowing for heavy tails is more effective in dealing with noisy series affected by outliers or extreme observations. Second, the multivariate extension of the DCS-t model has shown more appropriate than the robust univariate filtering approach in the case of scanner price data, as price time series are expected to have a good degree of correlation.

References

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics 32*(2), 159–188.

Baba, Y., R. F. Engle, D. F. Kraft, and K. F. Kroner (1990). Multivariate simultaneous generalized arch. *Manuscript, University of California, San Diego, Department of Economics*.

Bai, J. and Z. Chen (2008). Testing multivariate distributions in garch models. *Journal of Econometrics 143*(1), 19–36.

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., S. J. Koopman, and A. Lucas (2015, 03). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika 102*(2), 325–343.

Blazsek, S., A. Escribano, and A. Licht (2017). Score-driven non-linear multivariate dynamic location models. *UC3M Working paper*.

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.

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

Broda, C., E. Leibtag, and D. E. Weinstein (2009). The role of prices in measuring the poor's living standards. *Journal of Economic Perspectives 23*, 77–97.

Caivano, M., A. Harvey, and A. Luati (2016). Robust time series models with trend and seasonal components. *SERIEs 7*(1), 99–120.

Calvet, L. E., V. Czellar, and E. Ronchetti (2015). Robust filtering. *Journal of the American Statistical Association 110*(512), 1591–1606.

Capacci, S., O. Allais, C. Bonnet, and M. Mazzocchi (2019). The impact of the French soda tax on prices, purchases and tastes. An ex post evaluation. *Plos One*.

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

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scand. J. Stat. 8*, 93–115.

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.

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

Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Diewert, W. E., K. J. Fox, and J. de Haan (2016). A newly identified source of potential CPI bias: Weekly versus monthly unit value price indexes. *Economics Letters 141*, 169–172.

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

Durbin, J. and S. Koopman (2012). *Time series Analysis by State Space Methods* (Second ed.). Oxford University Press.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric Theory 11*(1), 122–150.

Fang, H.-B., K.-T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis 82*(1), 1 – 16.

Fang, K., S. Kotz, and K. Ng (1990). *Symmetric multivariate and related distributions.* Number 36 in Monographs on statistics and applied probability. London [u.a.]: Chapman & Hall.

Feenstra, R. C. and M. D. Shapiro (2003). High-frequency substitution and the measurement of price indexes. In R. C. Feenstra and M. D. Shapiro (Eds.), *Scanner Data and Price Indexes*, Volume 64 of *Studies in Income and Wealth*, pp. 123–150. The University of Chicago Press.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with student t innovations. *Journal of Business & Economic Statistics 21*(4), 532–546.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993). Nonlinear dynamic structures. *Econometrica 61*(4), 871–907.

Gasperoni, F., A. Luati, L. Paci, and E. D'Innocenzo (2021). Score-driven modeling of spatio-temporal data. *Journal of the American Statistical Association 0*(0), 1–12.

Hafner, C. M. and A. Preminger (2009). On asymptotic theory for multivariate garch models. *Journal of Multivariate Analysis 100*(9), 2044–2054.

Hamilton, J. (1994). *Time series analysis.* Princeton, NJ: Princeton Univ. Press.

Hannan, E. J. (1970). *Multiple Time Series.* Wiley Series in Probability and Statistics. Wiley.

Hannan, E. J. and M. Deistler (1987). *The Statistical Theory of Linear Systems.* New York, NY, USA: John Wiley & Sons, Inc.

Hansen, P. R. and M. Schmidtblaicher (2019). A dynamic model of vaccine compliance: How fake news undermined the danish hpv vaccine program. *Journal of Business & Economic Statistics 0*(0), 1–21.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Great Britain: Cambridge University Press.

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

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Ivancic, L., W. E. Diewert, and K. J. Fox (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics 161*, 24–35.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis 94*(2), 401–419.

Kaplan, G. and S. Schulhofer-Wohl (2017). Inflation at the household level. *Journal of Monetary Economics 91*, 19–38.

Kheifets, I. L. (2015). Specification tests for nonlinear dynamic models. *The Econometrics Journal 18*(1), 67–94.

Koopman, S. J., A. Lucas, and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *The Review of Economics and Statistics 98*(1), 97–110.

Kotz, S. and S. Nadarajah (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press.

Krengel, U. and A. Brunel (1985). *Ergodic Theorems*. De Gruyter studies in mathematics. W. de Gruyter.

Lin, W.-L. (1997). Impulse response function for conditional volatility in garch models. *Journal of Business & Economic Statistics 15*(1), 15–25.

Linton, O. and J. Wu (2020). A coupled component dcs-egarch model for intraday and overnight volatility. *Journal of Econometrics 217*, 176–201.

Lucas, A., J. Schaumburg, and B. Schwaab (2019). Bank business models at zero interest rates. *Journal of Business & Economic Statistics 37*(3), 542–555.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg.

Masreliez, J. C. (1975). Approximate non-gaussian filtering with linear state and observation relations. *IEEE Transactions on Automatic Control 20*, 107–110.

Masreliez, J. C. and D. R. Martin (1977). Robust bayesian estimation for the linear model and robustifying the kalman filter. *IEEE Transactions on Automatic Control 22*(3), 361–371.

Melser, D. (2018). Scanner data price indexes: Addressing some unresolved issues. *Journal of Business & Economic Statistics 36*(3), 516–522.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Óscar Jordá (2005). Estimation and inference of impulse responses by local projections. *The American Economic Review 95*(1), 161–182.

Prucha, I. R. and H. H. Kelejian (1984). The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances. *Econometrica 52*(3), 721–736.

Resnick, S. (2004). On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability 41*, 191–212.

Shapiro, M. D. and R. C. Feenstra (Eds.) (2003). *Scanner Data and Price Indexes*, Volume 64 of *Studies in Income and Wealth*, Chicago and London. The University of Chicago Press.

Silver, M. (1995). Elementary aggregates, micro-indices and scanner data: Some issues in the compilation of consumer price indices. *Review of Income and Wealth 41*(4), 427–438.

Silver, M. and S. Heravi (2001). Scanner data and the measurement of inflation. *The Economic Journal 111*(June), 383–404.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge University Press.

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.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models (2Nd Ed.)*. Berlin, Heidelberg: Springer-Verlag.

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

White, H. (2001). *Asymptotic Theory for Econometricians*. Economic Theory, Econometrics, and Mathematical Economics. Emerald Group Publishing Limited.

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

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. Wiley.

SUPPLEMENTARY MATERIAL

Additional supporting information may be found in the online appendix for this article at the publisher's website.

Appendix A: Main Proofs

Proof of Lemma 1

The score $\boldsymbol{u}_t$ in equation (5) can be written as

$$\boldsymbol{u_t} = \boldsymbol{v}_t(1 - b_t) \tag{19}$$

with $b_t = 1 - 1/w_t$ and where, conditional to $\mathcal{F}_{t-1}$,

$$b_t = \frac{\boldsymbol{v}_t^\top \boldsymbol{\Omega}^{-1}\boldsymbol{v}_t/\nu}{1 + \boldsymbol{v}_t^\top \boldsymbol{\Omega}^{-1}\boldsymbol{v}_t/\nu}, \quad 0 \leq b_t \leq 1, \quad \text{with} \quad b_t \sim \mathcal{B}eta\Big(\frac{N}{2}, \frac{\nu}{2}\Big), \tag{20}$$

i.e. the driving force $\boldsymbol{u}_t$ is a continuous function of a beta distributed random variable, see Pag. 19 of Kotz and Nadarajah (2004) or Proposition 39 of Harvey (2013). For $0 < \nu < \infty$, $\|\boldsymbol{u}_t\| = 0$ if $\|\boldsymbol{v}_t\| = 0$, while $\|\boldsymbol{u}_t\| \to 0$ if $\|\boldsymbol{v}_t\| \to \infty$ because $b_t \to 1$. Therefore, we achieve that $\sup_t \mathbb{E}[\|\boldsymbol{u}_t\|] < \infty$.

Second, we retrieve the moment structure of $\boldsymbol{u}_t$. Under assumption 1, the following stochastic representation is valid for the driving force

$$\boldsymbol{u}_t = \sqrt{\nu}\sqrt{b_t(1-b_t)}\boldsymbol{\Omega}^{1/2}\mathbf{z}_t, \tag{21}$$

where $\mathbf{z}_t$ is uniformly distributed on the unit sphere in $\mathbb{R}^N$ independently of b_t, see Fang et al. (1990). It follows that for even integers $m = 2s, s = 1, 2, \ldots$, the moments of $\boldsymbol{u}_t$ can be expressed as

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

Proof of Lemma 2

It follows from Lemma 1, that, at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and given the filtration at time $t-1$, i.e. $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \boldsymbol{y}_{t-3}, \ldots\}$, the score $\boldsymbol{u}_t$ forms a martingale difference sequence with zero mean and time-invariant covariance matrix. As, in addition, Lemma 1 shows that all the moments of $\boldsymbol{u}_t$ exist and do not depend on time, but only on the degrees of freedom ν and the vector time series dimension N, the process $\{\boldsymbol{u}_t\}_{t\in\mathbb{Z}}$ is *IID* and independently distributed of $\boldsymbol{\mu}_{t|t-1}$. Therefore, by using recursive arguments, for each starting value $\boldsymbol{\mu}_{s|s-1}$, where s is a fixed time point, one has that $\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega} = \boldsymbol{\Phi}^{t-s}(\boldsymbol{\mu}_{s|s-1} - \boldsymbol{\omega}) + \sum_{j=0}^{t-1}\boldsymbol{\Phi}^j\boldsymbol{K}\boldsymbol{u}_{t-j}$. Consequently, according to the theory of linear systems, see Hannan and Deistler (1987), the condition $\varrho(\boldsymbol{\Phi}) < 1$ is sufficient for the existence and uniqueness of a strictly stationary and ergodic solution $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$.

Then, when the process starts from the infinite past, we can write $\tilde{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\omega} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{K} \boldsymbol{u}_{t-j}$, so that, from Lemma 1, by taking the unconditional expectation and applying the triangle, Hölder and Minkowsky inequalities, we get

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

where $\bar{c} = N\|\boldsymbol{K}\|$ and $\bar{\rho} < 1$. The first inequality follows from a standard result in linear algebra, as $\|\boldsymbol{\Phi}\| = \|\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\| = \text{tr}(\boldsymbol{\Lambda}) = \sum_{i=1}^{N} \rho_i$ where ρ_i are the eigenvalues of $\boldsymbol{\Phi}$. $\square$

Proof of Lemma 3

The stationary and ergodic solution of equation (7) can be embedded in a first order nonlinear dynamic system. $\tilde{\boldsymbol{\mu}}_{t+1|t} = \phi(\tilde{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}), t \in \mathbb{Z}$. Let us define inductively, for $k \geq 1$ and any initialization $\hat{\boldsymbol{\mu}}_{1|0} \in \boldsymbol{\mathcal{M}}$, a sequence of Lipschitz maps $\phi^{(k+1)} : \boldsymbol{\mathcal{M}} \times \mathbb{R}^N \times \boldsymbol{\Theta} \mapsto \boldsymbol{\mathcal{M}}$ for $k \geq 1$ such that $\phi^{(k+1)}(\hat{\boldsymbol{\mu}}_{1|0}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k+1}, \boldsymbol{\theta}) = \phi\left(\phi^{(k)}(\hat{\boldsymbol{\mu}}_{1|0}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k, \boldsymbol{\theta}), \boldsymbol{y}_{k+1}, \boldsymbol{\theta}\right)$. By applying the mean value theorem to $\phi(\hat{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})$, that is, the nonstationary Lipschitz map, we obtain

$$\hat{\boldsymbol{\mu}}_{t+1|t} = \widehat{\boldsymbol{X}}_t^{\star} \hat{\boldsymbol{\mu}}_{t|t-1} + \varphi(\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}), \tag{22}$$

where $\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}$ denotes a set of points between $\hat{\boldsymbol{\mu}}_{t|t-1}$ and $\tilde{\boldsymbol{\mu}}_{t|t-1}$. Moreover, we have that $\widehat{\boldsymbol{X}}_t^{\star} = \phi'(\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta})$, where ϕ' denotes the first partial derivatives of ϕ with respect to the transpose of the vector $\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}$, and $\varphi(\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) = \phi(\tilde{\boldsymbol{\mu}}_{t|t-1}, \boldsymbol{y}_t, \boldsymbol{\theta}) - \widehat{\boldsymbol{X}}_t^{\star} \tilde{\boldsymbol{\mu}}_{t|t-1}$. Equation (22) is a multivariate SRE, that can be viewed as vector autoregressive process with random coefficients. The sufficient conditions for invertibility given by Bougerol (1993) and Straumann and Mikosch (2006) then become

$$\mathbb{E}\left[\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\phi(\tilde{\boldsymbol{\mu}}_{1|0}, \boldsymbol{y}_1, \boldsymbol{\theta}) - \tilde{\boldsymbol{\mu}}_{1|0}\right\|\right] < \infty, \qquad \mathbb{E}\left[\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{X}_1\right\|\right] < \infty, \tag{23}$$

for any $\tilde{\boldsymbol{\mu}}_{1|0} \in \boldsymbol{\mathcal{M}}$ and

$$\mathbb{E}\left[\ln \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\mu}\in\boldsymbol{\mathcal{M}}} \left\|\prod_{j=1}^{k} \boldsymbol{X}_{k-j+1}\right\|\right] < 0, \tag{24}$$

for $k \geq 1$ and where $\ln^+ x = \max\{0, \ln x\}$.

Let us consider condition (23). One has

$$\begin{aligned}\mathbb{E}\left[\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\phi(\tilde{\boldsymbol{\mu}}_{1|0}, \boldsymbol{y}_1, \boldsymbol{\theta}) - \tilde{\boldsymbol{\mu}}_{1|0}\right\|\right] &\leq 2\ln 2 + \ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{\Phi}\right\| + 2\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\tilde{\boldsymbol{\mu}}_{1|0} - \boldsymbol{\omega}\right\| \\ &\quad + \ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{K}\right\| + \mathbb{E}\left[\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\boldsymbol{u}_1\right\|\right] < \infty\end{aligned}$$

by compactness of $\boldsymbol{\Theta}$ and since $\boldsymbol{u}_t$ is uniformly bounded $\forall t$ in both $\boldsymbol{\mu}_{t|t-1} \in \boldsymbol{\mathcal{M}}$ and $\boldsymbol{y}_t \in \mathbb{R}^N$. In particular, for any $\boldsymbol{\mu}_{t|t-1} \in \boldsymbol{\mathcal{M}}$, as $\|\boldsymbol{y}_t\| \to \infty$ we obtain that $\|\boldsymbol{u}_t\| \to 0$. Thus, $\sup_t \mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{u}_t\|] < \infty$

which clearly implies $\mathbb{E}[\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{u}_1\|] < \infty$.

Moreover, note that $\mathbb{E}\big[\ln^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{X}_1\big\|\big] < \infty$ directly follows from the contraction condition $\mathbb{E}\left[\ln \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\tilde{\boldsymbol{\mu}}_{1|0}\in\boldsymbol{\mathcal{M}}} \|\boldsymbol{X}_1\|\right] < 0$. Therefore, condition (23) is fulfilled.

As far as condition (24) is concerned, the exponentially fast almost sure convergence of the filtered $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$ is obtained as an application of Theorem 3.1 in Bougerol (1993) or Theorem 2.8 in Straumann and Mikosch (2006), since the contraction condition (24) implies that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t+1|t} - \tilde{\boldsymbol{\mu}}_{t+1|t}\| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\left(\prod_{i=0}^{t-1} \widehat{\boldsymbol{X}}^{\star}_{t-i}\right)\left(\hat{\boldsymbol{\mu}}_{1|0} - \tilde{\boldsymbol{\mu}}_{1|0}\right)\right\| \leq \varrho^t c$, where $c > 0$ and $0 < \varrho < 1$ are constants.

Therefore, all the requirements of Bougerol (1993)'s Theorem are satisfied. Additionally, the claim that the moments are bounded follow from the fact that, as noted above, $\boldsymbol{u}_t$ is uniformly bounded. □

Proof of Lemma 4

Under the correct specification assumption 1, for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ the stationary and ergodic solution $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$ coincide with $\{\boldsymbol{\mu}_{t|t-1}\}_{t\in\mathbb{Z}}$ in (4), and, consequently, with $\boldsymbol{\mu}_t$, since Lemma 2 ensures that the SE solution is unique. As a consequence of Lemma 1 and Lemma 2, the process $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ is stationary by continuity and its moments are bounded. Ergodicity of $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ under the same assumptions follows by Proposition 4.3 of Krengel and Brunel (1985) □

To prove consistency and asymptotic normality of the MLE, additional quantities are introduced. Let us define the empirical average log-likelihood function based on the chosen initial value $\boldsymbol{\mu}_{1|0}$ and on the filtered sequence $\{\hat{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{N}}$

$$\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \widehat{\ell}_t(\boldsymbol{\theta}), \tag{25}$$

and the likelihood based on the stationary sequence $\{\tilde{\boldsymbol{\mu}}_{t|t-1}\}_{t\in\mathbb{Z}}$

$$\mathcal{L}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}), \tag{26}$$

with the following limit

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}[\ell_t(\boldsymbol{\theta})]. \tag{27}$$

The first and second derivatives of the above quantities with respect of the parameter will be denoted as $\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}), \mathcal{L}'_T(\boldsymbol{\theta}), \mathcal{L}'(\boldsymbol{\theta})$ and as $\widehat{\mathcal{L}}''_T(\boldsymbol{\theta}), \mathcal{L}''_T(\boldsymbol{\theta}), \mathcal{L}''(\boldsymbol{\theta})$, respectively.

The proof of consistency is based on some Lemmata that we report here for sake of clarity. The proofs of the Lemmata are in the online appendix.

Lemma 5. *Assume that conditions 1, 2 and 3 in Assumption 2 are satisfied. Then* $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta})|\right] < \infty$ *and* $\mathbb{E}\big[|\ell_t(\boldsymbol{\theta}_0)|\big] < \infty$. *Furthermore, under condition 4, for every* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, $\mathbb{E}\big[|\ell_t(\boldsymbol{\theta})|\big] < \mathbb{E}\big[|\ell_t(\boldsymbol{\theta}_0)|\big]$.

Lemma 6. *Assume that conditions 1, 2 and 3 in Assumption 2 are satisfied. Then,* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ *as* $t \to \infty$, *and* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ *as* $t \to \infty$, *where* $\widehat{\mathcal{L}}_T(\boldsymbol{\theta})$, $\mathcal{L}_T(\boldsymbol{\theta})$ *and* $\mathcal{L}(\boldsymbol{\theta})$ *and are defined in* (25), (26) *and* (27), *respectively.*

Proof of Theorem 4.1

One has,

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

By Lemma 6 and the Ergodic Theorem, $\lim_{T\to\infty} \widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0) = \lim_{T\to\infty} \mathcal{L}_T(\boldsymbol{\theta}_0) = \mathcal{L}(\boldsymbol{\theta}_0)$, and, by Lemma 5, $\mathcal{L}(\boldsymbol{\theta}) < \mathcal{L}(\boldsymbol{\theta}_0), \forall \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Following similar arguments of Theorem 3.4 in White (1994), one can show that strong consistency holds if $\forall\ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\exists\ \mathcal{B}_\eta(\boldsymbol{\theta})$, where $\mathcal{B}_\eta(\boldsymbol{\theta}) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \eta, \eta > 0\}$ s.t. for any $\boldsymbol{\theta}^\star \in \mathcal{B}_\eta(\boldsymbol{\theta})$,

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

With a similar reasoning, by the reverse Fatou's Lemma and the Ergodic Theorem

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

and therefore, $\forall\ \varepsilon > 0\ \exists\ \eta > 0$ s.t. $\mathbb{E}\left[\sup_{\boldsymbol{\theta}^\star\in\mathcal{B}_\eta(\boldsymbol{\theta})} \ell_t(\boldsymbol{\theta})\right] < \mathbb{E}\left[\ell_t(\boldsymbol{\theta})\right] + \varepsilon = \mathcal{L}(\boldsymbol{\theta}) + \varepsilon$.Note that ε can be made arbitrarily small. Therefore, the uniqueness and identifiability of the maximizer $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, is ensured by the uniqueness of $\boldsymbol{\theta}_0$ as the maximizer of the likelihood, see Lemma 5, the compactness of the parameter space $\boldsymbol{\Theta}$ and finally, the continuity of the limit $\mathcal{L}(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ which is ensured from the continuity of $\mathcal{L}_T(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\forall T \in \mathbb{N}$ and the uniform convergence in Lemma 6. Then, strong consistency follows by Theorem 3.4 in White (1994). □

The proof of asymptotic normality requires the following Lemmata, proved in the online appendix.

Lemma 7. *Assume that conditions 1, 2 and 3 in Assumption 2 are satisfied. Then, the first derivatives of the* log*-likelihood* $\mathcal{L}'_T(\boldsymbol{\theta}_0)$ *obeys the CLT for martingale difference sequences, that is* $\sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}_0) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{V})$ *as* $t \to \infty$*, where* $\boldsymbol{V} = \mathbb{E}\left[(\mathcal{L}'_T(\boldsymbol{\theta}_0))(\mathcal{L}'_T(\boldsymbol{\theta}_0))^\top\right]$.

Lemma 8. *Assume that conditions 1, 2, 3 and 4 in Assumption 2 are satisfied. Then, we obtain* $\sqrt{T}\|\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0) - \mathcal{L}'_T(\boldsymbol{\theta}_0)\| \xrightarrow{P} 0$ *as* $T \to \infty$.

Lemma 9. *Assume that conditions 1, 2 and 3, in Assumption 2 are satisfied. Then, we obtain* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\widehat{\mathcal{L}}''_T(\boldsymbol{\theta}) - \mathcal{L}''_T(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ *as* $t \to \infty$.

Lemma 10. *Assume that conditions 1, 2, 3 and 4 in Assumption 2 are satisfied. Then, we obtain* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}''_T(\boldsymbol{\theta}) - \mathcal{L}''(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ *as* $t \to \infty$,

Lemma 11. *Assume that conditions 1, 2, 3, 4 and 5 in Assumption 2 are satisfied. Then, the second derivative processes of the likelihood* $\left\{\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}\right\}_{t\in\mathbb{Z}}$ *are stationary ergodic with bounded moments. In particular,* $\mathbb{E}\left[\frac{d^2\ell_t(\boldsymbol{\theta})}{d\boldsymbol{\theta}d\boldsymbol{\theta}^\top}\right] < \infty$, *and is nonsingular.*

Proof of Theorem 4.2 (Asymptotic Normality)

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

$$\begin{aligned}\mathbf{0} =&\sqrt{T}\widehat{\mathcal{L}}'_T(\hat{\boldsymbol{\theta}}_T) = \sqrt{T}\Big[\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0) - \mathcal{L}'_T(\boldsymbol{\theta}_0)\Big] + \sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}_0)\\ &+ \Big[\big(\mathcal{L}''_T(\boldsymbol{\theta}_0) - \mathcal{L}''(\boldsymbol{\theta}_0)\big) + \big(\widehat{\mathcal{L}}''_T(\boldsymbol{\theta}^\star) - \mathcal{L}''_T(\boldsymbol{\theta}_0)\big) + \mathcal{L}''(\boldsymbol{\theta}_0)\Big]\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),\end{aligned} \tag{28}$$

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

First, the fact that $\sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}_0)$ obeys the CLT for martingales is entailed in Lemma 7. Convergence of the first difference in square brackets of equation (28) is ensured by Lemma 8. Thus, by the asymptotic equivalence (see Lemma 4.7 in White (2001)) $\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0)$ has the same asymptotic distribution of $\sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}_0)$. As regards the second line, we have that the middle term vanishes almost surely and exponentially fast, since Lemma 9 demonstrates that the initial conditions for the likelihood's second derivatives are asymptotically irrelevant and the consistency theorem further ensures the convergence in the same point by continuity arguments of the likelihood's second derivatives. In addition, the first term in the brackets of the second line vanishes as well by the Uniform Law of Large Numbers discussed in Lemma 10. Finally, with Lemma 11 at hand, we can easily solve equation (28), since $\mathcal{L}''(\boldsymbol{\theta}_0)$ is nonsingular. Slustky's Lemma (see Lemma 2.8 *(iii)* of van der Vaart (1998)) completes the proof. □

Appendix B: The Conditional Fisher Information Matrix

As we show in Section S2 of the online supplementary material, the conditional information matrix may be represented as follows,

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

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

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

The second,

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

Third, the unconditional mean

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

By symmetry, the fourth and last block are composed by

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

We note that the formulae above require the calculations of the derivative of the dynamic location process $\boldsymbol{\mu}_{t|t-1}$, which are given by

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

where $\boldsymbol{X}_t = \boldsymbol{\Phi} + \boldsymbol{K}\boldsymbol{\mathcal{C}}_t$ and with

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

and finally

$$
\begin{aligned}
\boldsymbol{C} &= \frac{\partial(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} = \boldsymbol{I}_N - \boldsymbol{\Phi},\\
\boldsymbol{D}_t &= \frac{\partial(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{\partial(\mathrm{vec}\,\boldsymbol{\Phi})^\top} = (\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})^\top \otimes \boldsymbol{I}_N,\\
\boldsymbol{E}_t &= \frac{\partial(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{\partial(\mathrm{vec}\,\boldsymbol{K})^\top} = \boldsymbol{u}_t^\top \otimes \boldsymbol{I}_N.
\end{aligned}
$$

It is worth to remark that $\boldsymbol{\omega}$ is asymptotically independent of the other parameters. Moreover, none of the terms of the conditional information matrix involves the second derivatives of the dynamic location. This result is a direct consequence of the asymptotic properties of the proposed MLE under the assumption of correct specification of the model and some regularity conditions.

Online supplement for the paper: A Robust Score-Driven Filter for Multivariate Time Series

Enzo D'Innocenzo*
Department of Economics, University of Bologna
Piazza Antonio Scaravilli 2, 40126 Bologna, Italy
e-mail: enzo.dinnocenzo2@unibo.it

Alessandra Luati
Department of Mathematics, Imperial College London
180 Queen's Gate, SW7 2AZ, London
e-mail: a.luati@imperial.ac.uk
and

Mario Mazzocchi
Department of Statistical Sciences, University of Bologna,
Via delle Belle Arti 41, 40126 Bologna, Italy,
e-mail: m.mazzocchi@unibo.it

March 17, 2023

Abstract

In this online supplementary materials, we provide details of a Monte Carlo study aimed to assess the finite sample properties of the MLE derived in the paper (Section S1), the relevant quantities for the implementation of the Fisher scoring algorithm (Section S2), the proofs of Lemma 5-11 in the main paper, as well as some additional auxiliary Lemmata (Section S3).

S1 Monte Carlo

The finite-sample properties of the MLE are investigated via Monte-Carlo simulations. We assume that the distribution of the heavy-tailed *IID* errors will be $\boldsymbol{\epsilon}_t \sim \boldsymbol{t}_{\nu_0}(\mathbf{0}_2, \boldsymbol{I}_2)$, where $\nu_0 \in \{3, 5, 10, 100\}$, that is, a standard bivariate Student's t with three, five and ten degrees of freedom, while the case when $\nu_0 = 100$ we cover the case when $\boldsymbol{\epsilon}_t$ tends to behave like a standard Gaussian noise. A property of the multivariate model introduced so far is that it estimates a linear Gaussian model when the errors are actually Gaussian. In this sense, the filter is robust to misspecification if normality holds. On the other hand, it is important to remark that we are assuming that all the time series share the

*Corresponding author.

same degrees of freedom ν_0. It is well-known that estimating the degrees of freedom in Student's t distributions can be quite challenging, since the implied profile likelihood is remarkably flat, see the discussion in Breusch et al. (1997).

In practice, we simulate data from the different specifications of the standard bivariate Student's t and for each of the realized paths of the time series we consider the recursion in (7), which satisfies the conditions of Lemma (2). During the process which generates the data, we use a burn-in period of $1,000$ replications and we store $T = 250$, 500 and $1,000$ observations. This ensures that the collected $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ are stationary ergodic.

With this simulated data at hand, we start the Fisher's scoring algorithm based on the analytical formulae described in Section 4.2. We repeat this simulation scheme $M = 1,000$ times for each case and we use the empirical measures of bias and root mean square error to quantify the accuracy of our proposed estimators. Formally, the empirical bias measure and the empirical root mean square error of $\hat{\nu}$ over the $M = 1,000$ replications are computed as

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

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

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

thus $\boldsymbol{\theta} \in \mathbb{R}^{14}$, which means that a complete bivariate system is characterized by 14 parameters. The true parameters of the considered DGP are

$$\nu_0 \in \{3, 5, 10, 100\},\ \boldsymbol{\Omega}_0 = \boldsymbol{I}_2,\ \boldsymbol{\omega}_0 = \begin{bmatrix}-3 & 5\end{bmatrix},\ \boldsymbol{\Phi}_0 = \begin{bmatrix}0.85 & 0.00\\ 0.00 & 0.80\end{bmatrix},\ \boldsymbol{K}_0 = \begin{bmatrix}0.95 & 0.05\\ 0.05 & 0.90\end{bmatrix}.$$

The Monte-Carlo results are reported in Tables S1 to S4 according to the values of the degrees of freedom $\nu_0 \in \{3, 5, 10, 100\}$, respectively. Each table contains three columns: *Estimate* reports the Monte-Carlo average of the point estimates obtained from the simulations, while *Bias* and *RMSE* report the Monte-Carlo deviations from the true values as described above, which are associated with the time series dimensions, that is $T = 250$, 500 and $1,000$.

The first evident result, common to all the tables, is that as the time series dimension increases, the values of the empirical *Bias* and *RMSE* tend to reduce sharply, which is line with the consistency Theorem 4.1. In particular, we note that even if the value of ν_0 is very low, namely $\nu_0 = 3$, the results are still satisfactory.

In general, estimation of the number of degrees of freedom is rather accurate. In the approximately Gaussian case when $\nu_0 = 100$ the filter collapses to the steady state form of the Kalman filter and the degrees of freedom parameter is recovered already in the case of the smallest sample size. Moreover, the decreasing bias and *RMSE* patterns may be due to the fixed initial value of the dynamic location vector $\boldsymbol{\mu}_{1|0}$ which was used to start the filter recursions. However, the invertibility conditions introduced in Lemma 3 ensure that for $T \to \infty$, this initial estimation bias will eventually tapers off. In conclusion, the *ML* estimations deliver satisfactory results in terms of bias and root mean square error, hence the reliability of the Fisher-scoring method.

Table S1: Monte-Carlo Simulation results for $\nu_0 = 3$.

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

Table S2: Monte-Carlo Simulation results for $\nu_0 = 5$.

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

Table S3: Monte-Carlo Simulation results for $\nu_0 = 10$.

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

Table S4: Monte-Carlo Simulation results for $\nu_0 = 100$.

	$T = 250$			$T = 500$			$T = 1000$		
	Estimate	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*	*Estimate*	*Bias*	*RMSE*
ν	96.708	3.290	25.729	98.782	1.218	13.782	100.855	-0.855	12.866
Ω_{11}	0.999	0.001	0.119	1.016	-0.017	0.085	1.002	-0.002	0.064
Ω_{12}	-0.005	0.005	0.068	0.006	-0.006	0.056	0.000	-0.001	0.030
Ω_{22}	0.991	0.009	0.106	1.008	-0.008	0.084	1.003	-0.003	0.061
ω_1	-2.974	-0.026	0.380	-2.964	-0.035	0.302	-3.032	0.032	0.268
ω_2	4.944	0.056	0.304	5.042	-0.042	0.230	5.009	-0.009	0.118
Φ_{11}	0.826	0.023	0.054	0.834	0.016	0.039	0.841	0.009	0.039
Φ_{12}	0.004	-0.004	0.060	0.002	-0.002	0.040	0.001	-0.001	0.022
Φ_{21}	-0.001	0.001	0.045	-0.004	0.004	0.033	0.001	-0.001	0.022
Φ_{22}	0.780	0.019	0.061	0.785	0.015	0.043	0.785	0.015	0.034
κ_{11}	0.945	0.004	0.121	0.946	0.004	0.093	0.947	0.003	0.049
κ_{12}	0.048	0.002	0.094	0.052	-0.002	0.062	0.055	-0.005	0.039
κ_{21}	0.064	-0.014	0.098	0.049	0.001	0.069	0.049	0.001	0.037
κ_{22}	0.910	-0.010	0.108	0.903	-0.003	0.090	0.907	-0.008	0.038

S2 Computational Aspects

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

S2.1 The Score Vector

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

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

yielding the recursions for the static parameters

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

for the unconditional mean

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

and for the remaining parameters determining the dynamics of the location vector

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

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

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

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

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

The second,

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

Third, the unconditional mean

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

By symmetry, the fourth and last block are composed by

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

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

To construct the score vector, we take the first differential of the likelihood function

$$\begin{aligned}\mathrm{d}\ell_t(\boldsymbol{\theta}) =& \frac{1}{2}\left[\psi\left(\frac{\nu+N}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{N}{\nu} + \frac{\nu+N}{\nu}\, b_t - \ln w_t\right](\mathrm{d}\nu) \\ &+ \frac{1}{2}(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}))^\top \boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\left[\frac{\nu+N}{\nu}\frac{1}{w_t}(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t) - \mathrm{vec}\,\boldsymbol{I}_N\right] \\ &+ \frac{\nu+N}{\nu}\frac{1}{w_t}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}),\end{aligned} \tag{S1}$$

where $\psi(x) = d\ln\Gamma(x)/d(x)$ is the digamma function and $\boldsymbol{\mathcal{D}}_N$ the duplication matrix, which allow us to write $\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega} = \boldsymbol{\mathcal{D}}_N(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}))$, since the scale matrix is symmetric. Secondly, we define $\boldsymbol{s}_t(\boldsymbol{\theta}) = d\ell_t(\boldsymbol{\theta})/d\boldsymbol{\theta}$ and partition the parameter as $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\psi}^\top)^\top$, so that the score vector can be partitioned into two blocks and two distinct applications of the chain rule are required. Specifically, for $\boldsymbol{\xi} = (\boldsymbol{\omega}^\top, (\mathrm{vech}(\boldsymbol{\Omega}))^\top, \nu)^\top$, we have

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

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

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

Let us start by considering the first differential of the dynamic location

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

where

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

Let us embed the dynamic differential as an SRE

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

where

$$\boldsymbol{X}_t = \boldsymbol{\Phi} + \boldsymbol{K}\boldsymbol{\mathcal{C}}_t, \tag{S2}$$

and

$$\boldsymbol{R}_t = \boldsymbol{K}\boldsymbol{a}_t\mathrm{d}\nu + \boldsymbol{K}\boldsymbol{B}_t\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega} + \boldsymbol{D}_t\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Phi} + \boldsymbol{E}_t\mathrm{d}\,\mathrm{vec}\,\boldsymbol{K}. \tag{S3}$$

The terms of the latter equations are

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

which we write, for convenience, also in their vectorised form

$$\begin{aligned}
\boldsymbol{a}_t &= b_t^{3/2}(1-b_t)^{1/2}/\nu \boldsymbol{\Omega}^{1/2}\mathbf{z}_t, \\
\mathrm{vec}\, \boldsymbol{B}_t &= \nu b_t^{3/2}(1-b_t)^{1/2}(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{1/2})(\mathbf{z}_t \otimes \mathbf{z}_t \otimes \mathbf{z}_t), \\
\mathrm{vec}\, \boldsymbol{\mathcal{C}}_t &= 2b_t(1-b_t)\,(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{1/2})(\mathbf{z}_t \otimes \mathbf{z}_t) - (1-b_t)\,\mathrm{vec}\, \boldsymbol{I}_N.
\end{aligned} \tag{S4}$$

The partial derivatives

$$\begin{aligned}
\boldsymbol{C} &= \frac{\partial(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\top} = (\boldsymbol{I}_N - \boldsymbol{\Phi}), \\
\boldsymbol{D}_t &= \frac{\partial(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{\partial (\mathrm{vec}\, \boldsymbol{\Phi})^\top} = \left[(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})^\top \otimes \boldsymbol{I}_N\right], \\
\boldsymbol{E}_t &= \frac{\partial(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{\partial (\mathrm{vec}\, \boldsymbol{K})^\top} = \left[(\boldsymbol{u}_t)^\top \otimes \boldsymbol{I}_N\right],
\end{aligned}$$

are required to obtain the final recursions, necessary for the iterative procedure

$$\begin{aligned}
\frac{d(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d\nu} &= \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\nu} + \boldsymbol{K}\boldsymbol{a}_t, \\
\frac{d(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top} &= \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vech}(\boldsymbol{\Omega}))^\top} + \boldsymbol{K}\boldsymbol{B}_t, \\
\frac{d(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d\boldsymbol{\omega}^\top} &= \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d\boldsymbol{\omega}^\top} + \boldsymbol{C}, \\
\frac{d(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d(\mathrm{vec}\, \boldsymbol{\Phi})^\top} &= \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\, \boldsymbol{\Phi})^\top} + \boldsymbol{D}_t, \\
\frac{d(\boldsymbol{\mu}_{t+1|t} - \boldsymbol{\omega})}{d(\mathrm{vec}\, \boldsymbol{K})^\top} &= \boldsymbol{X}_t \frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\, \boldsymbol{K})^\top} + \boldsymbol{E}_t.
\end{aligned} \tag{S5}$$

The discussion on the required partial derivatives of the log-likelihood function is similarly tackled. From (S1) the calculation are straightforward, we define

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

which completes the construction of the score vector.

S2.2 The Hessian Matrix

Like in the previous section, we obtain the second differential of the conditional log-likelihood by differentiating (S1), which yields

$$
\begin{aligned}
\mathrm{d}^2\ell_t(\boldsymbol{\theta}) =& \frac{1}{2}\left[\frac{1}{2}\psi'\left(\frac{\nu+N}{2}\right) - \frac{1}{2}\psi'\left(\frac{\nu}{2}\right) + \frac{N}{\nu^2} - \frac{N}{\nu^2}b_t - \frac{\nu+N}{\nu^2}b_t(1-b_t) + \frac{1}{\nu}b_t\right](\mathrm{d}^2\nu) \\
&+ \left[\frac{\nu+N}{2\nu^2}(1-b_t)^2(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2}) \right.\\
&\qquad \left. \times(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\otimes\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top)(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right] \\
&- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\
&- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\right] \\
&+ \left[\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right] \\
&- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right] \\
&+ \left[2\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top(\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right] \\
&+ \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{-1/2}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\
&+ \left[\frac{1}{2}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})\right] \\
&+ \left[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t + \frac{1}{2}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Omega})^\top(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)\right] \\
&\times\left[\frac{\nu+N}{\nu^2}b_t(1-b_t) - \frac{N}{\nu^2}(1-b_t)\right](\mathrm{d}\nu),
\end{aligned}
\tag{S6}
$$

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

We thus define the Hessian matrix

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

Similar arguments as those used in the computation of the score vector lead us to decompose the Hessian into four blocks and then apply the chain rule separately to each block. The first set is $\boldsymbol{\xi} = (\boldsymbol{\omega}^\top, (\mathrm{vech}(\boldsymbol{\Omega}))^\top, \nu)^\top$,

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

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

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

and finally, by symmetry, we get the remaining blocks

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

As far as the second differentials of the dynamic equation are concerned, we have

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

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

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

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

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad \times (\operatorname{vec}\boldsymbol{\Omega}^{-1})(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})\Big\}\\
&- (1-b_t)^2/\nu\Big\{\big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top \otimes \boldsymbol{I}_N\big](\operatorname{vec}\boldsymbol{I}_N)\big[(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})\big]\Big\}\\
&- 2(1-b_t)^2/\nu\Big\{\big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top \otimes \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\big](\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})\Big\}\\
&+ \Big\{[(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})^\top \otimes (\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)^\top]\operatorname{vec}(\boldsymbol{\Omega}^{-1/2}\otimes\boldsymbol{\Omega}^{-1/2})\Big\}\\
&\qquad\qquad\qquad\qquad \times \big[2b_t(1-b_t)^2/\nu^2 - (1-b_t)^2/\nu^2)\big](\mathrm{d}\nu)\\
&+ 2\Big\{\big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top \otimes (\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\big](\operatorname{vec}\boldsymbol{\Omega}^{-1})\Big\}\\
&\qquad\qquad\qquad\qquad \times \big[2b_t(1-b_t)/\nu^2 - (1-b_t)^2/\nu^2\big](\mathrm{d}\nu)\\
&- \Big\{\big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top \otimes \boldsymbol{I}_N\big](\operatorname{vec}\boldsymbol{I}_N)\Big\}\big[b_t(1-b_t)/\nu\big](\mathrm{d}\nu).
\end{aligned}
$$

Let us write

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

where $\boldsymbol{X}_t$ is as in (S2) and

$$
\begin{aligned}
\boldsymbol{Q}_t =& \boldsymbol{K}\boldsymbol{a}_t'\mathrm{d}^2\nu + \boldsymbol{K}\boldsymbol{B}_t'\mathrm{d}^2\operatorname{vec}\boldsymbol{\Omega} + \boldsymbol{K}(\mathrm{d}\operatorname{vec}\boldsymbol{\Omega})^\top\widehat{\boldsymbol{a}\boldsymbol{B}}_t'\mathrm{d}\nu\\
&+ \boldsymbol{D}_t'\mathrm{d}^2\operatorname{vec}\boldsymbol{\Phi} + \boldsymbol{E}_t'\mathrm{d}^2\operatorname{vec}\boldsymbol{K} + (\mathrm{d}\operatorname{vec}\boldsymbol{\Phi})^\top\widehat{\boldsymbol{D}\boldsymbol{E}}_t'(\mathrm{d}\operatorname{vec}\boldsymbol{K}).
\end{aligned} \tag{S7}
$$

We now derive the terms of recursion (S7). We first need a set of partial derivative

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

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

$$
\begin{aligned}
\boldsymbol{\mathcal{C}}_t' = \frac{\partial^2\boldsymbol{u}_t}{\partial\boldsymbol{\mu}_{t|t-1}\partial\boldsymbol{\mu}_{t|t-1}^\top} =& 8(1-b_t)^3/\nu^2\Big\{\big[\boldsymbol{I}_N\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\big](\operatorname{vec}\boldsymbol{\Omega}^{-1})\Big\}\\
&\times\big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}\big]2(1-b_t)^2/\nu\Big\{\big[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_N\big]\\
&\times\big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\otimes\boldsymbol{I}_N + \boldsymbol{I}_N\otimes(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\big]\Big\}\\
&- 2(1-b_t)^2/\nu\Big\{\big[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_N\big]\big[(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})\otimes\boldsymbol{I}_N\big]\Big\}.
\end{aligned} \tag{S8}
$$

Secondly, a set of partial cross-derivatives

$$
\begin{aligned}
\widehat{\boldsymbol{aB}}'_t = \frac{\partial^2 \boldsymbol{u}_t}{\partial(\mathrm{vech}(\boldsymbol{\Omega}))\partial\nu} =& [\boldsymbol{I}_N \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top] \,\mathrm{vec}(\boldsymbol{\Omega}^{-1/2} \otimes \boldsymbol{\Omega}^{-1/2}) \\
& \times \big[2b_t(1-b_t)^2/\nu^2 - (1-b_t)^2/\nu^2)\big], \\
\widehat{\boldsymbol{a\mathcal{C}}}'_t = \frac{\partial^2 \boldsymbol{u}_t}{\partial\boldsymbol{\mu}_{t|t-1}\partial\nu} =& 2\Big\{\big[\boldsymbol{I}_N \otimes (\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})(\boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1})^\top\big](\mathrm{vec}\,\boldsymbol{\Omega}^{-1})\Big\} \\
& \times \big[2b_t(1-b_t)/\nu^2 - (1-b_t)^2/\nu^2\big] \\
& - \Big\{\big[(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top \otimes \boldsymbol{I}_N\big](\mathrm{vec}\,\boldsymbol{I}_N)\Big\} \\
& \times \big[b_t(1-b_t)/\nu\big],
\end{aligned} \tag{S9}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{B\mathcal{C}}}'_t = \frac{\partial^2 \boldsymbol{u}_t}{\partial\boldsymbol{\mu}_{t|t-1}\partial(\mathrm{vech}(\boldsymbol{\Omega}))^\top} =& 4(1-b_t)^3/\nu^2\Big\{\big[\boldsymbol{I}_N \otimes \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\big] \\
& \times (\mathrm{vec}\,\boldsymbol{\Omega}^{-1})(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\Big\} \\
& - (1-b_t)^2/\nu\Big\{\big[\boldsymbol{I}_N \otimes \boldsymbol{I}_N\big](\mathrm{vec}\,\boldsymbol{I}_N) \\
& \times \big[(\boldsymbol{\epsilon}_t \otimes \boldsymbol{\epsilon}_t)^\top(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\big]\Big\} \\
& - 2(1-b_t)^2/\nu\Big\{\big[\boldsymbol{I}_N \otimes \boldsymbol{\Omega}^{1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top\boldsymbol{\Omega}^{1/2}\big] \\
& \times (\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\Big\}.
\end{aligned} \tag{S10}
$$

In addition, we a new set of partial derivatives defined by

$$
\begin{aligned}
\boldsymbol{D}'_t = \frac{\partial[d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})]}{\partial(\mathrm{vec}\,\boldsymbol{\Phi})d(\mathrm{vec}\,\boldsymbol{\Phi})^\top} =& 2\left[\left(\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{\Phi})^\top}\right)^\top \otimes \boldsymbol{I}_N\right], \\
\boldsymbol{E}'_t = \frac{\partial[d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})]}{\partial(\mathrm{vec}\,\boldsymbol{K})d(\mathrm{vec}\,\boldsymbol{K})^\top} =& 2\left[\left(\boldsymbol{\mathcal{C}}_t^\top\frac{d(\boldsymbol{\mu}_{t|t-1} - \boldsymbol{\omega})}{d(\mathrm{vec}\,\boldsymbol{K})^\top}\right)^\top \otimes \boldsymbol{I}_N\right],
\end{aligned}
$$

and finally, we conclude the derivations with

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

We therefore have obtained a new set of recursions composed by

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

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

which continue with

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

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

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

and conclude with

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

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

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

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

The construction of the Hessian can now be completed by deriving the remaining second-order partial derivatives of the second differential in (S6).

By virtue of this representation, one can show that

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

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

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

and

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

which completes the construction of the Hessian matrix.

S2.3 The Conditional Information Matrix

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

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

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

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

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

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

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

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

Now, by equation (S5) the calculations boils down to the static terms of the matrix. Specifically,

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

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

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

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

and lastly the cross terms

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

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

S2.4 Third differentials

This section derives the third differential of the conditional log-likelihood with respect to the dynamic location, auxiliary to the proof of the asymptotic normality of the MLE, see Lemma 9. By differentiating equation (S6) with respect $\boldsymbol{\mu}_{t|t-1}$ one obtains

$$\begin{aligned}
\mathrm{d}^3_{\boldsymbol{\mu}_{t|t-1}}\ell_t(\boldsymbol{\theta}) =& \left[8\frac{\nu+N}{\nu^3}(1-b_t)^3(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\
&+ \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top[\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\otimes\boldsymbol{I}_N+\boldsymbol{I}_N\otimes\boldsymbol{\epsilon}_t\boldsymbol{\Omega}^{-1/2}](\mathrm{d}\boldsymbol{\mu}_{t|t-1})^2\right] \\
&- \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\
&- \left[2\frac{\nu+N}{\nu^2}(1-b_t)^2(\mathrm{d}\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\right] \\
&- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}^2\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}(\mathrm{d}\boldsymbol{\mu}_{t|t-1})\right] \\
&- \left[\frac{\nu+N}{\nu}(1-b_t)(\mathrm{d}^3\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1/2}\boldsymbol{\epsilon}_t\right]. \qquad \text{(S11)}
\end{aligned}$$

S3 Lemmata

This Appendix contains the proofs of the auxiliary lemmata used to establish consistency and asymptotic normality of the MLE of Section 4.

Lemmata for the Proof of Consistency

Proof of Lemma 5

Consider the t-th contribution to the log-likelihood, $\ell_t(\boldsymbol{\theta})$. We have that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta})|\right] \\
&\quad \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ln\Gamma\left(\frac{\nu+N}{2}\right)\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ln\Gamma\left(\frac{\nu}{2}\right)\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{N}{2}\ln(\pi\nu)\right| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{1}{2}\ln|\boldsymbol{\Omega}|\right| \\
&\qquad + \frac{\nu+N}{2}\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ln\left(1+\frac{(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})}{\nu}\right)\right|\right] < \infty,
\end{aligned}$$

since the compactness of the parameter space $\boldsymbol{\Theta}$ with $0<\nu<\infty$ ensures that the first three terms are finite, there exist $\Omega_- > 0$ and $\Omega_+ < \infty$ such that $\Omega_- < |\boldsymbol{\Omega}| < \Omega_+$ and moreover, the logarithmic moment in the last term exists as a consequence of Lemmata 2, 4 and 3 with $m>0$. In particular, we can show that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\quad\boldsymbol{\Theta}} \left|(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{t|t-1})/\nu\right|^m\right] < \infty,$$

is always satisfied for some $m>0$ and with $\nu>0$, implying the existence of the required logarithmic moment.

Clearly, the result obtained above also implies that $\mathbb{E}\big[|\ell_t(\boldsymbol{\theta}_0)|\big] < \infty$, and then, we can turn to the last statement.

To prove the uniqueness and identifiability of $\boldsymbol{\theta}_0$ it is sufficient to consider the sequence $\{\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ under the assumption that $(\nu, \operatorname{vech}\boldsymbol{\Omega})^\top = (\nu_0, \operatorname{vech}\boldsymbol{\Omega}_0)^\top$. We prove the argument by contradiction.

By denoting with $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta})$ and $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0)$ the dynamic location vector evaluated at $\boldsymbol{\theta}$ and the true parameter vector $\boldsymbol{\theta}_0$ respectively, and as $\boldsymbol{v}_t(\boldsymbol{\theta}) = \boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta})$ and $\boldsymbol{v}_t(\boldsymbol{\theta}_0) = \boldsymbol{y}_t - \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0)$ the difference becomes

$$\begin{aligned}
&\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0) \\
&\propto \ln\Big[1+(\boldsymbol{v}_t(\boldsymbol{\theta}))^\top\boldsymbol{\Omega}_0^{-1}(\boldsymbol{v}_t(\boldsymbol{\theta}))/\nu_0\Big] - \ln\Big[1+(\boldsymbol{v}_t(\boldsymbol{\theta}_0))^\top\boldsymbol{\Omega}_0^{-1}(\boldsymbol{v}_t(\boldsymbol{\theta}_0))/\nu_0\Big] \\
&= \ln\left(\Big[1+(\boldsymbol{v}_t(\boldsymbol{\theta}))^\top\boldsymbol{\Omega}_0^{-1}(\boldsymbol{v}_t(\boldsymbol{\theta}))/\nu_0\Big]\Big/\Big[1+(\boldsymbol{v}_t(\boldsymbol{\theta}_0))^\top\boldsymbol{\Omega}_0^{-1}(\boldsymbol{v}_t(\boldsymbol{\theta}_0))/\nu_0\Big]\right),
\end{aligned}$$

where the latter equation holds if and only if $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0)$ almost surely since $\boldsymbol{\Omega}_0$ is symmetric positive definite and $0<\nu<\infty$. Under maintained assumptions, it is clear that $\{\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ are stationary and ergodic sequences, which implies that the same holds true for the sequence by $\{(\boldsymbol{\mu}_{t+1|t}(\boldsymbol{\theta}) - \boldsymbol{\mu}_{t+1|t}(\boldsymbol{\theta}_0))\}_{t\in\mathbb{Z}}$. Thus, it is possible to write the difference recursion as

$$\begin{aligned}
&(\boldsymbol{\mu}_{t+1|t}(\boldsymbol{\theta}) - \boldsymbol{\mu}_{t+1|t}(\boldsymbol{\theta}_0)) \\
&\qquad = (\boldsymbol{\omega}-\boldsymbol{\omega}_0) + (\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\boldsymbol{\omega}_0 + (\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0) + (\boldsymbol{K}-\boldsymbol{K}_0)\boldsymbol{u}_t,
\end{aligned}$$

and the relation above entails the fact that if $\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}_0)$ $\forall t$ almost surely, then

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

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

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

Proof of Lemma 6

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

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

where by direct calculation and the triangle inequality we get

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

Note that the compactness of the parameter space imposed by condition 3 is crucial here. Moreover, if we treat the dynamic location vector as a fixed parameter with value $\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}$ and let $\boldsymbol{y}_t \to \infty$ the entire term in the right hand side of the latter inequality will vanish. Hence, we obtain that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial \hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1}} \right\| = O_p(1)$, which is enough to ensure the existence of log-moments. Furthermore, conditions 1 and 2 are needed in order to keep the data stationary and ergodic and the filter invertible, respectively. Thus, we can apply Lemma 3 and obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}\| \xrightarrow{\text{e.a.s.}} 0$. In conclusion, by Lemma 2.1 in Straumann and Mikosch (2006) the claimed almost sure convergence holds.

Now, for the second result, we note that Theorem A.2.2 of White (1994), i.e. the Uniform Law of Large Numbers in its version for stationary and ergodic processes, applies straightforwardly to our case since: (1) the parameter space is compact, (2) the empirical likelihood function $\mathcal{L}_T(\boldsymbol{\theta})$ defined in (26) is continuous in $\boldsymbol{\theta}$ $\forall$ $\boldsymbol{y}_t$ and $\forall$ $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is measurable in $\boldsymbol{y}_t$, which is stationary and ergodic, and (3) by Lemma 5 we obtain the moment bound $\mathbb{E}\big[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta})|\big] < \infty$ which ensure the dominance condition.

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

Lemmata for the Proof of Asymptotic Normality

Proof of Lemma 7

The first derivatives of the log-likelihood contribution at time t with respect to the true parameter vector $\boldsymbol{\theta}_0$ can be retrieved from the differential in equation (S1).

We have

$$\begin{aligned}
&\mathbb{E}_{t-1}[\mathrm{d}\ell_t(\boldsymbol{\theta}_0)] \\
&\quad=\frac{1}{2}\left[\psi\left(\frac{\nu_0+N}{2}\right)-\psi\left(\frac{\nu_0}{2}\right)-\frac{N}{\nu_0}+\frac{\nu_0+N}{\nu_0}\,\mathbb{E}_{t-1}[b_t]-\mathbb{E}_{t-1}[\ln w_t]\right](\mathrm{d}\nu_0) \\
&\qquad+\frac{1}{2}(\mathrm{d}\,\mathrm{vech}(\boldsymbol{\Omega}_0))^\top\boldsymbol{\mathcal{D}}_N^\top(\boldsymbol{\Omega}_0^{-1/2}\otimes\boldsymbol{\Omega}_0^{-1/2})\left[\frac{\nu_0+N}{\nu_0}\mathbb{E}_{t-1}[(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)/w_t]-\mathrm{vec}\,\boldsymbol{I}_N\right] \\
&\qquad+\frac{\nu_0+N}{\nu}(\mathrm{d}\tilde{\boldsymbol{\mu}}_{t|t-1})^\top\boldsymbol{\Omega}_0^{-1}\mathbb{E}_{t-1}[(\boldsymbol{y}_t-\tilde{\boldsymbol{\mu}}_{t|t-1})/w_t],
\end{aligned}$$

since the derivatives obtained from $\mathrm{d}\tilde{\boldsymbol{\mu}}_{t|t-1}$ are $\mathcal{F}_{t-1}$-measurables.

Then, one has

$$\begin{aligned}
\mathbb{E}_{t-1}[b_t]&=\frac{N}{\nu+N},\\
\mathbb{E}_{t-1}[\ln(1/w_t)]=\mathbb{E}_{t-1}[\ln(1-b_t)]&=\psi\left(\frac{\nu}{2}\right)-\psi\left(\frac{\nu+N}{2}\right),\\
\mathbb{E}_{t-1}[(\boldsymbol{\epsilon}_t\otimes\boldsymbol{\epsilon}_t)/w_t]=\nu\mathbb{E}_{t-1}[(\mathbf{z}_t\otimes\mathbf{z}_t)]\mathbb{E}_{t-1}[b_t]&=\frac{\nu}{\nu+N}\,\mathrm{vec}\,\boldsymbol{I}_N,\\
\mathbb{E}_{t-1}[(\boldsymbol{y}_t-\tilde{\boldsymbol{\mu}}_{t|t-1})/w_t]=\sqrt{\nu}\mathbb{E}_{t-1}[\sqrt{b_t(1-b_t)}]\,\boldsymbol{\Omega}^{1/2}\mathbb{E}_{t-1}[\mathbf{z}_t]&=\mathbf{0},
\end{aligned}$$

where the first and the second equality follow from the properties of the beta distribution, see equation (20). The third and the fourth equalities are obtained based on the stochastic representation of the model given in equation (21). Thus, by substitutions, we obtain the martingale difference property.

The second claim follows by Lemmata 1, 2, 3, 4, 5 and by an application of the continuous mapping theorem to $d\ell_t(\boldsymbol{\theta}_0)$.

With the support of the Cramér-Wold device (see van der Vaart (1998) pag. 16) the CLT for martingales of Billingsley (1961) directly applies to the linear combination $\sqrt{T}\mathcal{L}'_T(\boldsymbol{\theta}_0)=\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\frac{d\ell_t(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}_0}\Rightarrow\mathcal{N}(\mathbf{0},\boldsymbol{V})$. □

Proof of Lemma 8

The claimed convergence in probability can be proved based on the invertibility of the location filter, see Lemma 3, and its derivatives, see Lemma S3.3. Invertibility also ensures that the perturbed first differential of the dynamic location evaluated at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ will converge to the unique stationary ergodic solution,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\hat{\boldsymbol{\mu}}_{t|t-1}-\tilde{\boldsymbol{\mu}}_{t|t-1}\|\xrightarrow{\text{e.a.s.}}0\quad\text{and}\\
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|d\hat{\boldsymbol{\mu}}_{t|t-1}-d\tilde{\boldsymbol{\mu}}_{t|t-1}\|\xrightarrow{\text{e.a.s.}}0\quad\text{as}\quad t\to\infty.
\end{aligned}\tag{S12}$$

Hence, we can rely on a multivariate mean value expansion about all the elements of the vectors $\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}$ and $d\hat{\boldsymbol{\mu}}^{\star}_{t|t-1}$, which lie on the chords between $(\hat{\boldsymbol{\mu}}_{t|t-1}, \tilde{\boldsymbol{\mu}}_{t|t-1})$ and $(d\hat{\boldsymbol{\mu}}_{t|t-1}, d\tilde{\boldsymbol{\mu}}_{t|t-1})$ respectively, yielding

$$\sqrt{T}\|\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0) - \mathcal{L}'_T(\boldsymbol{\theta}_0)\| \\ \leq \sqrt{T}\left\|\begin{matrix}\frac{\partial(\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0))}{\partial\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1}} \\ \frac{\partial(\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0))}{\partial(d\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1})}\end{matrix}\right\| \left\|\begin{matrix}(\hat{\boldsymbol{\mu}}_{t|t-1} - \tilde{\boldsymbol{\mu}}_{t|t-1}) \\ (d\hat{\boldsymbol{\mu}}_{t|t-1} - d\tilde{\boldsymbol{\mu}}_{t|t-1})\end{matrix}\right\|.$$

The first term on the right hand of the inequality is uniformly bounded. Exponentially fast almost sure convergence of the second term in the right hand side is obtained by Lemma S3.3.

By means of analogous arguments as in Lemma 6 we can show that

$$\left\|\frac{\partial(\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0))}{\partial\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1}}\right\| = O_p(1), \quad \text{and} \quad \left\|\frac{\partial(\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0))}{\partial(d\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1})}\right\| = O_p(1).$$

Moreover, the results obtained in (S12) imply that for t large enough

$$\max\{\|\hat{\boldsymbol{\mu}}_{t|t-1} - \tilde{\boldsymbol{\mu}}_{t|t-1}\|, \|d\hat{\boldsymbol{\mu}}_{t|t-1} - d\tilde{\boldsymbol{\mu}}_{t|t-1}\|\} < 1.$$

By using the Chebyshev and the c_m inequalities we then have that for $\varepsilon > 0$ and some $m > 0$

$$\begin{aligned}\mathbb{P}\Big(\sqrt{T}\|\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0) - \mathcal{L}'_T(\boldsymbol{\theta}_0)\| > \varepsilon\Big) \leq & \frac{\sqrt{T}}{\varepsilon^m}\mathbb{E}[\|\widehat{\mathcal{L}}'_T(\boldsymbol{\theta}_0) - \mathcal{L}'_T(\boldsymbol{\theta}_0)\|^m] \\ \leq & \frac{1}{T^{m/2}\varepsilon^m}\sum_{t=1}^{T}\mathbb{E}\left[\left\|\frac{d\widehat{\ell}_t(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}_0} - \frac{d\ell_t(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}_0}\right\|^m\right] \\ \leq & \frac{1}{T^{m/2}\varepsilon^m}O_p(t\varrho^t),\end{aligned}$$

which is $O_p(T^{-m/2})$ and this implies the claimed convergence in probability. □

Proof of Lemma 9

The second derivatives of the likelihood are nonlinear functions of the filtered location vector and its first end second derivatives. Hence, the mean value theorem is applied for each dynamic equation. As a result,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\widehat{\mathcal{L}}''_T(\boldsymbol{\theta}) - \mathcal{L}''_T(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\begin{matrix}\frac{\partial\widehat{\mathcal{L}}''_T(\boldsymbol{\theta})}{\partial\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1}} \\ \frac{\partial\widehat{\mathcal{L}}''_T(\boldsymbol{\theta})}{\partial(d\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1})} \\ \frac{\partial\widehat{\mathcal{L}}''_T(\boldsymbol{\theta})}{\partial(d^2\hat{\boldsymbol{\mu}}^{\star\top}_{t|t-1})}\end{matrix}\right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\begin{matrix}(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\mu}_{t|t-1}) \\ (d\hat{\boldsymbol{\mu}}_{t|t-1} - d\boldsymbol{\mu}_{t|t-1}) \\ (d^2\hat{\boldsymbol{\mu}}_{t|t-1} - d^2\boldsymbol{\mu}_{t|t-1})\end{matrix}\right\|.$$

Thus, the proof follows by the same arguments of the proof of Lemma 6, i.e. by obtaining the uniformly boundedness of the first term and the exponentially fast convergence of the second term in the right hand side respectively. Note that the last component of the first term in the right hand side involves a third order differential, which can be found in (S11) and is uniformly bounded. Subsequent applications of Lemma 2.1 of Straumann and Mikosch (2006) yield the desired result. □

Proof of Lemma 10

Note that $\mathcal{L}''_T(\boldsymbol{\theta})$ is a continuous function of $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ and therefore stationary and ergodic. The Lemma follows straightforwardly from Lemma 11 and The Uniform Law of Large Numbers for ergodic stationary processes, see Theorem A.2.2 in White (1994) and Lemma 6. □

Proof of Lemma 11

This Lemma is a multivariate extension of the Theorem 5 of Harvey (2013). Thus, we only discuss the relevant arguments.

The complete equation of the second differential is more subtle than the first, thus we leave it in (S6). We prove the arguments by considering equation (13), namely

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

By taking the expectation, we get a finite and static term in the first summand on the right hand side, while by the independence and the martingale difference sequence properties of the score vector, the last term becomes null. Thus, we can focus our attention on the middle term. Define

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

Note that, by independence, we can express the vectorized counterpart as

$$\operatorname{vec}\boldsymbol{\mathcal{I}}^{(\boldsymbol{\mu}_{t|t-1})}(\boldsymbol{\theta}) = \mathbb{E}\left[\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\otimes\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right]^\top \operatorname{vec}\boldsymbol{\mathcal{I}}^{(\boldsymbol{\mu})}(\boldsymbol{\theta}).$$

By Lemmata 3 and S3.3, the dynamic location filter and its differentials are invertible and achieve their own unique stationary ergodic solution with an unbounded number of finite moments.

Thus, we obtain the desired result by repeated applications of the law of iterated expectation to the following equality

$$\begin{aligned}&\mathbb{E}_{t-1}\left[\left(\frac{d(\boldsymbol{\mu}_{t+1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\otimes\left(\frac{d(\boldsymbol{\mu}_{t+1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right]^\top\\ &=\mathbb{E}_{t-1}\left[\left(\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}+\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\otimes\left(\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}+\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\right]^\top\\ &=\left(\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\otimes\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)^\top\mathbb{E}_{t-1}\left[\left(\boldsymbol{X}_t\otimes\boldsymbol{X}_t\right)\right]^\top\\ &\quad+\mathbb{E}_{t-1}\left[\left(\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\otimes\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\right]^\top+\mathbb{E}_{t-1}\left[\left(\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\otimes\boldsymbol{X}_t\frac{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})}{d\boldsymbol{\theta}^\top}\right)\right]^\top\\ &\quad+\mathbb{E}_{t-1}\left[\left(\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\otimes\frac{d\boldsymbol{R}_t}{d\boldsymbol{\theta}^\top}\right)\right]^\top.\end{aligned}$$

Note that the contraction conditions 2 and 5 are more than enough to ensure the stability of the recursions, while Lemma S3.1 ensures the existence of the required moments. □

Auxiliary Lemmata

Lemma S3.1. *Consider the stochastic difference equation*

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

where $\boldsymbol{X}_t$ *and* $\boldsymbol{R}_t$ *are defined in* (S2) *and* (S3)*, respectively.*

Assume that conditions 1, 2 and 3 in Assumption 2 are satisfied. Then, there exist a unique sequence $\{d(\tilde{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\omega})\}_{t\in\mathbb{Z}}$ *which is stationary and ergodic. A causal stationary solution exists and can be expressed as*

$$d(\tilde{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\omega}) = \sum_{j=0}^{\infty}\Big(\prod_{k=1}^{j}\boldsymbol{X}_{t-k}\Big)\boldsymbol{R}_{t-j}.$$

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

Proof. The proof follows the arguments of the proof of Lemma 2, which now applies by rewriting $\boldsymbol{X}_t$ and all the components of $\boldsymbol{R}_t$ in terms of the innovations and independently of $\tilde{\boldsymbol{\mu}}_{t|t-1}$ so that a stationary ergodic sequence $\{(\boldsymbol{X}_t, \boldsymbol{R}_t)\}_{t\in\mathbb{Z}}$ can be generated.

It follows from Lemma 2 that the first condition is used in order to keep the multivariate system stable and the matrices $\boldsymbol{X}_t$ random, while the contraction condition 2 for linear stochastic difference equations gives us the sufficient condition which ensures that $d(\tilde{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\omega})$ is the unique stationary and ergodic solution, see Bougerol (1993).

Moreover, the Hölder and Minkowsky inequalities imply that

$$\mathbb{E}\Big[\|d(\tilde{\boldsymbol{\mu}}_{t+1|t} - \boldsymbol{\omega})\|^m\Big] \le \Big\{\sum_{j=0}^{\infty}\mathbb{E}\Big[\Big\|\prod_{k=0}^{j}\boldsymbol{X}_{t-k}\Big\|^m\Big]^{1/m}\mathbb{E}\Big[\|\boldsymbol{R}_{t-j}\|^m\Big]^{1/m}\Big\}^m.$$

In addition, from equation (S4) we note that when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$\begin{aligned}
&\mathbb{E}\Big[\|\boldsymbol{X}_t\|^m\Big]\\
&\quad\le \|\boldsymbol{\Phi}\|^m + \mathbb{E}\Big[\|\boldsymbol{K}\boldsymbol{\mathcal{C}}_t\|^m\Big]\\
&\quad\le \bar{\rho}^m + c_{\boldsymbol{K}}\mathbb{E}\Big[b_t^{m/2}(1-b_t)^{m/2}\Big]\mathbb{E}\big[\|(\mathbf{z}_t\otimes\mathbf{z}_t)\|^m\big] + c_{\boldsymbol{K}}N^{m/2}\mathbb{E}\big[(1-b_t)^{m/2}\big]\\
&\quad= \bar{\rho}^m + c_{\boldsymbol{K}}\mathbb{E}\big[\|\mathbf{z}_t\|^{2m}\big]\frac{B\big(\frac{N+m}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu}{2}\big)} + c_{\boldsymbol{K}}N^{m/2}\frac{B\big(\frac{N}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu}{2}\big)}\\
&\quad= \bar{\rho}^m + \frac{c_{\boldsymbol{K}}}{N^m}\frac{B\big(\frac{N+m}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu_0}{2}\big)} + c_{\boldsymbol{K}}N^{m/2}\frac{B\big(\frac{N}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu}{2}\big)} < \infty,
\end{aligned}$$

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

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

$$\begin{aligned}
\mathbb{E}\Big[\|\boldsymbol{a}_t\|^m\Big] &= \mathbb{E}\Big[b_t^{3m/2}(1-b_t)^{m/2}/\nu^{m/2}\Big]\mathbb{E}\Big[\|\boldsymbol{\Omega}^{1/2}\mathbf{z}_t\|^m\Big]\\
&\le \frac{c_{\boldsymbol{\Omega}}}{N^{m/2}}\frac{B\big(\frac{N+3m}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu_0}{2}\big)} < \infty,
\end{aligned}$$

In addition,

$$
\begin{aligned}
&\mathbb{E}\Big[\|\operatorname{vec}\boldsymbol{B}_t\|^m\Big]\\
=&\mathbb{E}\Big[\nu^{m/2}b_t^{3m/2}(1-b_t)^{m/2}\Big]\mathbb{E}\Big[\Big\|(\boldsymbol{\Omega}^{-1/2}\mathbf{z}_t\otimes\boldsymbol{\Omega}^{-1/2}\mathbf{z}_t\otimes\boldsymbol{\Omega}^{1/2}\mathbf{z}_t)\Big\|^m\Big]\\
\leq& c_{\boldsymbol{\Omega}}\mathbb{E}\big[\|\mathbf{z}_t\|^{3m}\big]\frac{B\big(\frac{N+3m}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu}{2}\big)}=\frac{c_{\boldsymbol{\Omega}}}{N^{3m/2}}\frac{B\big(\frac{N+3m}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu}{2}\big)}<\infty,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\Big[\|\boldsymbol{D}_t\|^m\Big]=&\mathbb{E}\Big[\big\|\big[(\tilde{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})^\top\otimes\boldsymbol{I}_N\big]\big\|^m\Big]\\
\leq&\Big\{\sqrt{N}\bar{c}\sum_{j=0}^{\infty}\bar{\rho}^j\Big(\mathbb{E}\Big[\|\boldsymbol{u}_{t-j}\|^m\Big]\Big)^{1/m}\Big\}^m<\infty,
\end{aligned}
$$

by Lemma 2, and finally, by Lemma 1,

$$
\begin{aligned}
\mathbb{E}\Big[\|\boldsymbol{E}_t\|^m\Big]&=\mathbb{E}\Big[\big\|\big[(\boldsymbol{u}_t)^\top\otimes\boldsymbol{I}_N\big]\big\|^m\Big]\leq N^{m/2}\mathbb{E}\Big[\|\boldsymbol{u}_t\|^m\Big]\\
&\leq c_{\boldsymbol{\Omega}}\nu^{m/2}\frac{B\big(\frac{N+m}{2},\frac{\nu+m}{2}\big)}{B\big(\frac{N}{2},\frac{\nu}{2}\big)}.<\infty.
\end{aligned}
$$

□

Lemma S3.2. *Consider the stochastic difference equation*

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

where $\boldsymbol{X}_t$, $\boldsymbol{Q}_t$ and $\boldsymbol{C}_t'$ are defined in (S2), (S7) *and* (S8), *respectively.*

Assume that conditions 1 and 2 in Assumption 2 are satisfied. Then, there exist a unique sequence $\{d^2(\tilde{\boldsymbol{\mu}}_{t+1|t}-\boldsymbol{\omega})\}_{t\in\mathbb{Z}}$ which is stationary and ergodic. A causal stationary solution exists and can be expressed as

$$
\begin{aligned}
d^2(\tilde{\boldsymbol{\mu}}_{t+1|t}-\boldsymbol{\omega})=&\sum_{j=0}^{\infty}\Big\{\Big(\prod_{k=1}^{j}\boldsymbol{X}_{t-k}\Big)\\
&\times\Big[\boldsymbol{K}d(\tilde{\boldsymbol{\mu}}_{t-j|t-j-1}-\boldsymbol{\omega})^\top\boldsymbol{C}_{t-j}' d(\tilde{\boldsymbol{\mu}}_{t-j|t-j-1}-\boldsymbol{\omega})+\boldsymbol{Q}_{t-j}\Big]\Big\}
\end{aligned}
$$

Furthermore, $\mathbb{E}[\|d^2(\tilde{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\|^m]<\infty$ for every $m>0$.

Proof. From Lemma 2, the first two conditions ensure the existence of a unique stationary and ergodic sequence $\{d^2(\tilde{\boldsymbol{\mu}}_{t+1|t}-\boldsymbol{\omega})\}_{t\in\mathbb{Z}}$.

Moreover, by the Hölder and Minkowsky inequalities along with the independence between each component, imply that

$$\mathbb{E}\bigg[\|d^2(\tilde{\boldsymbol{\mu}}_{t+1|t}-\boldsymbol{\omega})\|^m\bigg] \leq \bigg\{\sum_{j=0}^{\infty}\mathbb{E}\bigg[\bigg\|\prod_{k=0}^{j}\boldsymbol{X}_{t-k}\bigg\|^{2m}\bigg]^{1/2m}$$
$$\times\bigg(c_{\boldsymbol{K}}\mathbb{E}\bigg[\|d(\tilde{\boldsymbol{\mu}}_{t-j|t-j-1}-\boldsymbol{\omega})\|^{4m}\bigg]^{1/4m}\mathbb{E}\bigg[\|\boldsymbol{\mathcal{C}}'_{t-j}\|^{2m}\bigg]^{1/2m}$$
$$+\mathbb{E}\bigg[\|\boldsymbol{Q}_{t-j}\|^m\bigg]^{1/m}\bigg)\bigg\}^m,$$

from which we can see that, by Lemma S3.1, the first two terms are uniformly bounded and the third is the second derivative of the driving force with respect to the dynamic location vector.

In the same spirit of Lemma S3.1, let us consider equation (S8). Then, when $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, the c_m-inequality establishes that

$$\mathbb{E}\bigg[\|\boldsymbol{\mathcal{C}}'_t\|^m\bigg] \leq \mathbb{E}\bigg[[8(1-b_t)^3/\nu^2]^m\bigg\|\Big\{\big[\boldsymbol{I}_N\otimes\boldsymbol{v}_t\boldsymbol{v}_t^\top\big]\operatorname{vec}\boldsymbol{\Omega}^{-1}\Big\}\bigg\|^m\bigg\|\big[\boldsymbol{v}_t^\top\boldsymbol{\Omega}^{-1}\big]\bigg\|^m\bigg]$$
$$+\mathbb{E}\bigg[[2(1-b_t)^2/\nu]^m\bigg\|\Big\{\big[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_N\big]\big[\boldsymbol{v}_t\otimes\boldsymbol{I}_N+\boldsymbol{I}_N\otimes\boldsymbol{v}_t\big]\Big\}\bigg\|^m\bigg]$$
$$+\mathbb{E}\bigg[[2(1-b_t)^2/\nu]^m\bigg\|\Big\{\big[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{I}_N\big]\big[\boldsymbol{v}_t\otimes\boldsymbol{I}_N\big]\Big\}\bigg\|^m\bigg]$$
$$\leq C_4\mathbb{E}\big[\|\mathbf{z}_t\|^{3m}\big]+C_3\mathbb{E}\big[\|\mathbf{z}_t\|^{2m}\big]+C_3\mathbb{E}\big[\|\mathbf{z}_t\|^{2m}\big]<\infty.$$

Straightforward calculations show that analogous results hold for each component of $\boldsymbol{Q}_t$, so that $\boldsymbol{Q}_t$ is uniformly bounded. □

Lemma S3.3. *Let the conditions of Lemmata 2, S3.1 and S3.2 hold true. Consider further the filtering equation* (7) *under the condition of Lemma 3. Then, for any initialization of the filter* $\hat{\boldsymbol{\mu}}_{1|0}$ *and its first derivatives in* $d\hat{\boldsymbol{\mu}}_{1|0}$*, the perturbed first and second derivative sequences of the dynamic location filter, i.e.* $\{d(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\}_{t\in\mathbb{N}}$ *and* $\{d^2(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\}_{t\in\mathbb{N}}$*, converge exponentially fast almost surely to the unique stationary ergodic solution* $\{d(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})\}_{t\in\mathbb{Z}}$ *and* $\{d^2(\boldsymbol{\mu}_{t|t-1}-\boldsymbol{\omega})\}_{t\in\mathbb{Z}}$*.*

Furthermore, for any $m>0$

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

Proof. We provide a detailed discussion for the first case, that is the convergence of the perturbed first derivatives, since the proof for the convergence of the perturbed second derivatives follows the same line.

The proof of this Lemma builds upon the arguments of Theorem 2.10 in Straumann and Mikosch (2006) for perturbed SRE. In particular, the perturbed SREs corresponds to the derivatives in $d(\hat{\boldsymbol{\mu}}_{t+1|t}-\boldsymbol{\omega})=\widehat{\boldsymbol{X}}_t d(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})+\widehat{\boldsymbol{R}}_t$, which are nonlinear functions of the initialized filtered sequence $\{(\hat{\boldsymbol{\mu}}_{t|t-1}-\boldsymbol{\omega})\}_{t\in\mathbb{N}}$. The relevant contraction condition (8) of Lemma 3 holds and the required convergence of the recurrence equation is obtained if

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\widehat{\boldsymbol{X}}_t-\widetilde{\boldsymbol{X}}_t\|\xrightarrow{\text{e.a.s.}}0 \quad\text{and}\quad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\widehat{\boldsymbol{R}}_t-\widetilde{\boldsymbol{R}}_t\|\xrightarrow{\text{e.a.s.}}0 \quad\text{as}\quad t\to\infty. \tag{S13}$$

In order to verify these conditions, we use the mean value theorem, giving

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{X}}_t - \widetilde{\boldsymbol{X}}_t\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\mathcal{C}}'_t\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \tilde{\boldsymbol{\mu}}_{t|t-1}\|, \tag{S14}$$

and

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{R}}_t - \widetilde{\boldsymbol{R}}_t\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \begin{matrix} \boldsymbol{\mathcal{C}}'_t \\ \widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t \\ \widehat{\boldsymbol{a}\boldsymbol{\mathcal{C}}}'_t \end{matrix} \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \tilde{\boldsymbol{\mu}}_{t|t-1}\|,$$

where the expression for $\boldsymbol{\mathcal{C}}'_t$, $\widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t$ and $\widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t$ can be found in (S8), (S9) and (S10) respectively. We can combine the results obtained in Lemma S3.2 together with the almost sure exponentially fast convergence (9) in Lemma 3, in order to achieve the required convergences in (S13). As in Lemma S3.1 we can show by direct calculations that the property of uniformly boundedness applies to each these derivatives, since they are continuous functions of w_t in equation (5).

We obtain that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\mathcal{C}}'_t\| = O_p(1), \quad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \begin{matrix} \boldsymbol{\mathcal{C}}'_t \\ \widehat{\boldsymbol{B}\boldsymbol{\mathcal{C}}}'_t \\ \widehat{\boldsymbol{a}\boldsymbol{\mathcal{C}}}'_t \end{matrix} \right\| = O_p(1)$$

and

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\mu}}_{t|t-1} - \tilde{\boldsymbol{\mu}}_{t|t-1}\| = o_{e.a.s.}(1) \quad \text{as} \quad t \to \infty.$$

Thus, repeated applications of Lemma 2.1 in Straumann and Mikosch (2006) ensure the required convergence in (S14).

Summarising, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|d(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega}) - d(\tilde{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega})\| \xrightarrow{\text{e.a.s.}} 0 \qquad \text{as} \qquad t \to \infty.$$

Since the sequence $\{d^2(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega})\}_{t\in\mathbb{N}}$ is a nonlinear function of both the perturbed recurrence $\{d(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega})\}_{t\in\mathbb{N}}$ and the filter $\{(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega})\}_{t\in\mathbb{N}}$ the same arguments apply sequentially, yielding

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|d^2(\hat{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega}) - d^2(\tilde{\boldsymbol{\mu}}_{t|t-1} - \boldsymbol{\omega})\| \xrightarrow{\text{e.a.s.}} 0 \qquad \text{as} \qquad t \to \infty.$$

The second claim for the moment bounds follows by the continuous mapping theorem, since the derivatives are nonlinear continuous functions of $\tilde{\boldsymbol{\mu}}_{t|t-1}$, which has unbounded moments, see Lemma 3. □

References

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

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

Breusch, T. S., J. C. Robertson, and A. H. Welsh (1997). The emperor's new clothes: a critique of the multivariate t regression model. *Statistica Neerlandica 51*(3), 269–286.

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

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

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.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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