TI 2025-042/III
Tinbergen Institute Discussion Paper

Matrix-Valued Spatial Autoregressions with Dynamic and Robust Heterogeneous Spillovers

Yicong Lin[1]
André Lucas[2]
Shiqi Ye[3]

1 Vrije Universiteit Amsterdam, Tinbergen Institute

2 Vrije Universiteit Amsterdam, Tinbergen Institute

3 AMSS Center for Forecasting Science, Chinese Academy of Sciences

Matrix-Valued Spatial Autoregressions with Dynamic and Robust Heterogeneous Spillovers

Yicong Lin[a], André Lucas[a], Shiqi Ye[b*]

[a]: Vrije Universiteit Amsterdam and Tinbergen Institute

[b]: AMSS Center for Forecasting Science, Chinese Academy of Sciences

June 26, 2025

Abstract

We introduce a new time-varying parameter spatial matrix autoregressive model that integrates matrix-valued time series, heterogeneous spillover effects, outlier robustness, and time-varying parameters in one unified framework. The model allows for separate dynamic spatial spillover effects across both the row and column dimensions of the matrix-valued observations. Robustness is introduced through innovations that follow a (conditionally heteroskedastic) matrix Student's t distribution. In addition, the proposed model nests many existing spatial autoregressive models, yet remains easy to estimate using standard maximum likelihood methods. We establish the stationarity and invertibility of the model and the consistency and asymptotic normality of the maximum likelihood estimator. Our simulations reveal that the latent time-varying two-way spatial spillover effects can be successfully recovered, even under severe model misspecification. The model's usefulness is illustrated both in-sample and out-of-sample using two different applications: one in international trade, and the other based on global stock market data.

Keywords: matrix-valued time series; spatial autoregression; time-varying parameters; score-driven dynamics; outlier-robust.

[*]Corresponding author. Email: shiqi.ye.c@gmail.com. Ye gratefully acknowledges support from China NNSF Fundamental Scientific Center Project (No. 71988101); the National Natural Science Foundation of China (No. 723B2020); the China Postdoctoral Science Foundation Funded Project (No. 2024M763465); the Research Project by the General Office of the National Data Bureau (No. SJ-zh2024075).

1 Introduction

Recent years have highlighted some of the limitations of traditional vector-based modeling approaches for complex data structures. Specifically, when the data inherently exhibit a row and column structure, it is natural to move beyond the scope of vector-valued representations and consider matrix-valued time series; see, e.g., Chen and Chen (2022) and Chen et al. (2024) for an example with bilateral trade data between countries and Chang et al. (2023) and Li and Xiao (2024) for a financial portfolio example with data across multiple countries and sectors. One of the core models for matrix-valued time-series is the Matrix Autoregressive (MAR) model proposed by Chen et al. (2021). It substantially reduces the dimensionality of the model's parameter space without compromising interpretability. This model has by now been extended to accommodate non-stationarity and cointegration, non-linearities, and high-dimensional settings (see, for instance, Jiang et al., 2024; Li and Xiao, 2024; Yu et al., 2024; Zhang, 2024).

For many matrix-valued time-series, the rows and columns of the observations relate to some form of spatial dependence, either through geographic connections, trade relations, social connections, or otherwise. Such connections can be exploited to build parsimonious, interpretable spatial (regression) models (for an overview of spatial models in the vector-valued case, see, e.g., LeSage and Pace, 2009; Elhorst et al., 2021; Anselin, 2022). It is unlikely, however, that the strength and impact of the underlying spatial connections remains constant over time, particularly if we consider a longer sample period. This poses challenges to existing matrix-valued time series models, as a spatial version of those models that accounts for time-varying spatial spillovers is thus far lacking from the literature. In this paper, we therefore develop a new time-varying parameter *spatial* MAR (TVP-SMAR) model. The model enjoys three distinguishing features. First, it allows for the inclusion of different observed network effects for the row versus the column dimension of the observations. These network effects are further amplified or mitigated by so-called spatial spillover parameters, which are again allowed to be different across rows and columns. The design ensures a considerable dimension reduction of the parameter space, while also preserving empirical interpretability.

Second, we allow the spatial spillover parameters to be heterogeneous both across time and across the individual rows and columns of the observations. For vector-valued observations, both forms of heterogeneity have been shown to be important (see, e.g., Aquaro et al., 2021; Catania and Billé, 2017; Yang and Lee, 2021; Gasperoni et al., 2023; D'Innocenzo et al., 2024) and we generalize them to the matrix-valued context both across row and column dimensions. In addition, we introduce time variation in the spatial spillover parameters. Allowing for dynamic spatial spillover effects sets our approach substantially apart from models like Yang and Lee (2021) and Pu et al. (2024). These earlier models also study spatial spillovers for matrix observations, but either only consider one-way (column-related) effects, or omit spatial spillover heterogeneity, or (in all cases) require the spatial spillovers to be time-invariant. Relaxing the time-invariance assumption for matrix-valued observations is challenging, but important (see, for instance, Chen et al., 2024; He et al., 2024; Zhang and Chan, 2024; Yu et al., 2025, for examples on time-varying parameters in matrix models outside the spatial context). Our set-up overcomes these challenges by using the score-driven time-varying parameter framework of Creal et al. (2013) and Harvey (2013) and deriving recurrence relations that explicitly exploit the matrix structure of the observations. As established in the literature, score-driven parameter updates result in expected improvements in Kullback-Leibler divergence of the model upon every parameter update (Blasques et al., 2015; Gorgi et al., 2023; Creal et al., 2024; de Punder et al., 2024) and offer robustness to potential dynamic misspecification (Beutner et al., 2023).

Third, our model can robustly cope with outliers and influential observations by using the matrix Student's t distribution of Gupta and Nagar (2018). Again, this matrix Student's t exploits the row and column dimensions of the data and is more than merely a re-shaped version of the well-known vector-valued Student's t distribution. The matrix Student's t distribution also links directly to the dynamics of the spatial spillover parameters via the score-driven framework of Creal et al. (2013) and Harvey (2013): due to the fat-tailedness of the matrix t, these updates automatically downweight outlying observations, resulting in robust dynamics for the spatial spillovers.

Despite its flexibility, the log-likelihood function of the new model is available in closed form. This makes parameter estimation and inference possible via standard maximum likelihood procedures. We show that the model is stationary, ergodic, and invertible, and that the maximum likelihood estimator for its parameters is consistent and asymptotically normal. This allows us to recover the time-varying spatial spillover parameter paths from the data exponentially fast almost surely. Using simulations, we also show that the new model can recover the time-varying two-way spatial spillover effects, including settings where the model is severely misspecified.

We present two applications to support the new model's usefulness. In our first application, we analyze monthly trade growth rates in an international bilateral import-export trade network. Our second application looks at weekly sector-level global stock returns. Both applications showcase the improved in-sample fit and interpretability of our new spatial MAR model with time-varying parameters, as well as its robust out-of-sample predictive performance.

The remainder of this paper is structured as follows. Section 2 introduces the model and estimation framework. Section 3 studies the model's asymptotic properties. Section 4 presents an extensive simulation study, while Section 5 explores two empirical applications. Section 6 concludes. Technical proofs and additional empirical results are available in the supplementary material.

Throughout, we adopt the following notational conventions. Vectors and matrices are in bold, whereas scalars are non-bold. A block-diagonal matrix with, for instance, two blocks $\boldsymbol{A}$ and $\boldsymbol{B}$ is written as $\mathrm{diag}\,(\boldsymbol{A}, \boldsymbol{B})$. We also write $\mathrm{diag}(\boldsymbol{A})$ to denote a column vector holding the diagonal elements of a single square matrix $\boldsymbol{A}$. For any matrix $\boldsymbol{A}$, let $\boldsymbol{A}_{i,j}$ denote its (i, j)th element and let $\boldsymbol{I}_m$ denote the $m \times m$ identity matrix.

2 Modelling framework

We first introduce the general model in Section 2.1. Section 2.2 outlines the distributional assumptions, while Section 2.3 discusses the parameter dynamics and the outlier-robust features of the model. Section 2.4 addresses the estimation of the model's static parameters.

2.1 The TVP-SMAR model

Let $\boldsymbol{Y}_t \in \mathbb{R}^{m \times n}$ denote a matrix-valued observation for $t = 1, \ldots, T$ and some finite m and n, where, for instance, m may be the number of different regions or countries, and n the number of sectors in a production type network. Our time-varying parameter spatial matrix-valued autoregressive (TVP-SMAR) model takes the following form:

$$\boldsymbol{Y}_t = \boldsymbol{R}_t^r\, \boldsymbol{W}_t^r\, \boldsymbol{Y}_t\, \boldsymbol{B}^\top + \boldsymbol{A}\, \boldsymbol{Y}_t\, \boldsymbol{W}_t^{c\top}\, \boldsymbol{R}_t^c + \boldsymbol{C}\, \boldsymbol{X}_t\, \boldsymbol{D}^\top + \boldsymbol{E}_t, \qquad \boldsymbol{E}_t \overset{\text{i.i.d.}}{\sim} p_E\left(\boldsymbol{E}_t; \boldsymbol{\xi}_E\right), \quad (1)$$

where $\boldsymbol{R}_t^r \in \mathbb{R}^{m\times m}$ and $\boldsymbol{R}_t^c \in \mathbb{R}^{n\times n}$ are diagonal parameter matrices holding the heterogeneous, time-varying row and column-related spatial spillovers, respectively, and $\boldsymbol{W}_t^r \in \mathbb{R}^{m\times m}$ and $\boldsymbol{W}_t^c \in \mathbb{R}^{n\times n}$ denote the observed corresponding network connection matrices. The coefficient matrices $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ are assumed to be diagonal, though this assumption can be relaxed if one is willing to impose further identification restrictions. Finally, $\boldsymbol{X}_t \in \mathbb{R}^{k_1\times k_2}$ is a matrix containing observed exogenous variables with associated parameter matrices $\boldsymbol{D} \in \mathbb{R}^{n\times k_2}$ and $\boldsymbol{C} \in \mathbb{R}^{m\times k_1}$. We assume that the matrix-valued error term $\boldsymbol{E}_t \in \mathbb{R}^{m\times n}$ is independent and identically distributed (i.i.d.) with a predetermined density function $p_E\left(\,\cdot\,;\boldsymbol{\xi}_E\right)$, where $\boldsymbol{\xi}_E$ represents the parameters characterizing the distribution.

The proposed TVP-SMAR model in (1) encompasses several existing spatial autoregressive models from the literature, such as the time-varying parameter spatial autoregressive (TVP-SAR) model of D'Innocenzo et al. (2024) using $n = 1$; the TVP-SAR model with a single time-varying spatial spillover parameter of Blasques et al. (2016) using $n = 1$ and $\boldsymbol{R}_t^r = \rho_t \boldsymbol{I}_m$; the static heterogeneous spatial spillover model of Aquaro et al. (2021) using $n = 1$ and $\boldsymbol{R}_t^r = \boldsymbol{R}^r$; and the standard spatial regression model of, for instance, Anselin et al. (2009), Asgharian et al. (2013), Kou et al. (2018), and Denbee et al. (2021), by setting $n = 1$ and $\boldsymbol{R}_t^r = \rho \boldsymbol{I}_m$. At the same time, for $n > 1$, the model in (1) encompasses the dynamic panel spatial vector autoregression model of Yang and Lee (2021), which at the core reduces to $\boldsymbol{Y}_t = \boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B} + \boldsymbol{E}_t$.[1] Our model in Eq. (1) substantially extends this specification by allowing for both row and column-related networks and corresponding heterogeneous spatial spillover strengths, as well as by allowing these spillover parameters to vary over time.

[1]For simplicity, we consider a reduced version of the model in Yang and Lee (2021) to highlight the core differences in the spatial effects. Yang and Lee (2021) also allow for spatial time-lags and lagged terms of $\boldsymbol{Y}_t$, which could be included in a similar way in our specification in (1).

To better understand how the spatial spillover specification in (1) exploits the row and column structure of the data, let $\boldsymbol{Y}_{i,\boldsymbol{\cdot},t}$ and $\boldsymbol{Y}_{\boldsymbol{\cdot},j,t}$ denote the ith row and jth column of $\boldsymbol{Y}_t$, respectively. Consider the jth column of $\boldsymbol{Y}_t$ for a model without exogenous regressors. In that case (1) reduces to

$$
\begin{aligned}
&\boldsymbol{Y}_{\boldsymbol{\cdot},j,t} = \boldsymbol{B}_{j,j}\,\boldsymbol{R}_t^r\,\boldsymbol{W}_t^r\,\boldsymbol{Y}_{\boldsymbol{\cdot},j,t} + \boldsymbol{R}_{j,j,t}^c\,\boldsymbol{A}\,\boldsymbol{Y}_t\,(\boldsymbol{W}_{j,\boldsymbol{\cdot},t}^c)^\top + \boldsymbol{E}_{\boldsymbol{\cdot},j,t} \qquad \Longrightarrow \qquad (2)\\
&\boldsymbol{Y}_{i,j,t} = \boldsymbol{B}_{j,j}\,\boldsymbol{R}_{i,i,t}^r \sum_{k=1}^{n} \boldsymbol{W}_{i,k,t}^r\,\boldsymbol{Y}_{k,j,t} + \boldsymbol{A}_{i,i}\,\boldsymbol{R}_{j,j,t}^c \sum_{k=1}^{m} \boldsymbol{W}_{j,k,t}^c\,\boldsymbol{Y}_{i,k,t} + \boldsymbol{E}_{i,j,t}.
\end{aligned}
$$

The first term in (2) resembles the model for vector-valued time series with heterogeneous row-wise time-varying spillovers (D'Innocenzo et al., 2024). The spatial matrix-valued specification imposes parsimony by requiring that these time-varying row-wise spatial dynamics are the same across columns j, differing only by a proportionality constant $\boldsymbol{B}_{j,j}$. Substantially different from D'Innocenzo et al. (2024), however, the second term in (2) imposes further static, heterogeneous row-wise spillovers via the static diagonal matrix $\boldsymbol{A}$. These somewhat resemble the static heterogeneous spillovers in for instance Aquaro et al. (2021), Yang and Lee (2021), or Heil et al. (2022). In contrast to these earlier papers, however, the static heterogeneous spillovers are scaled by the time-varying scalar spatial spillover coefficient $\boldsymbol{R}_{j,j,t}^c$ and operate on different time-varying linear combinations of the columns of $\boldsymbol{Y}_t$ via the jth column of $\boldsymbol{W}_t^{c\top}$. The scalar effect of $\boldsymbol{R}_{j,j,t}$ bears some resemblance to the scalar spatial dynamics in Blasques et al. (2016), but it is substantially adjusted to allow for cross-sectional heterogeneity via the static matrix $\boldsymbol{A}$, and for more complex row-wise dynamics via the first term in (2). A similar argument can be made if one considers the ith row of $\boldsymbol{Y}_t$ rather than the jth column, obtaining $\boldsymbol{Y}_{i,\boldsymbol{\cdot},t}^\top = \boldsymbol{A}_{i,i}\,\boldsymbol{R}_t^c\,\boldsymbol{W}_t^c\boldsymbol{Y}_{i,\boldsymbol{\cdot},t}^\top + \boldsymbol{R}_{i,i,t}^r\boldsymbol{B}\,\boldsymbol{Y}_t^\top\,\boldsymbol{W}_{i,\boldsymbol{\cdot},t}^{r\;\top} + \boldsymbol{E}_{i,\boldsymbol{\cdot},t}^\top$. In short, the new specification in (1) extends and parsimoniously combines several earlier proposals from the literature to introduce time-variation and heterogeneity in spatial autoregressions, while being specifically tailored to matrix-valued data $\boldsymbol{Y}_t$ through explicit use of its row and column structure.

One special case of the general specification in (1) is worth noting. If one is exclusively focused on the time-varying spatial spillover effects along either the row or column dimension of $\boldsymbol{Y}_t$, the following special form of the model can be considered:

$$\boldsymbol{Y}_t = \boldsymbol{R}_t^r \boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B}^\top + \boldsymbol{C}\boldsymbol{X}_t \boldsymbol{D}^\top + \boldsymbol{E}_t, \qquad \boldsymbol{E}_t \overset{\text{i.i.d.}}{\sim} p_E\left(\boldsymbol{E}_t; \boldsymbol{\xi}_E\right), \tag{3}$$

where $\boldsymbol{B}$ can be a full $n \times n$ matrix (with $\boldsymbol{B}_{1,1} = 1$ for identification). Note that this specification is still more general than the core panel specification of Yang and Lee (2021), which has $\boldsymbol{B}$ as diagonal and static $\boldsymbol{R}_t^r = \boldsymbol{I}_m$ for all t. We also note that model (1) can be further extended to allow for heterogeneity in network structures between columns or rows by considering

$$\boldsymbol{Y}_t = \sum_{j=1}^{n} \boldsymbol{B}_{j,j}\, \boldsymbol{R}_t^r\, \boldsymbol{W}_{j,t}^r\, \boldsymbol{Y}_t + \sum_{i=1}^{m} \boldsymbol{A}_{i,i}\, \boldsymbol{Y}_t\, \boldsymbol{W}_{i,t}^{c\top}\, \boldsymbol{R}_t^c + \boldsymbol{C}\, \boldsymbol{X}_t\, \boldsymbol{D}^\top + \boldsymbol{E}_t, \tag{4}$$

where $\boldsymbol{W}_{j,t}^r \in \mathbb{R}^{m\times m}$ and $\boldsymbol{W}_{i,t}^c \in \mathbb{R}^{n\times n}$ are row (column) connection matrices that vary across columns (rows). Models like (4) can be directly linked to the multi-country, multi-sector specifications in for instance Blasques et al. (2023), where the network connections between countries are allowed to vary across sectors. The TVP-SMAR model can be further extended to incorporate spatial spillover effects in the disturbance $\boldsymbol{E}_t$, by specifying

$$\boldsymbol{E}_t = \boldsymbol{\Lambda}_t^r \boldsymbol{W}_t^r \boldsymbol{E}_t \widetilde{\boldsymbol{B}}^\top + \widetilde{\boldsymbol{A}} \boldsymbol{E}_t \boldsymbol{W}_t^c \boldsymbol{\Lambda}_t^c + \boldsymbol{V}_t, \qquad \boldsymbol{V}_t \overset{\text{i.i.d.}}{\sim} p_V\left(\boldsymbol{V}_t; \boldsymbol{\xi}_V\right), \tag{5}$$

where $\boldsymbol{\Lambda}_t^r$, $\boldsymbol{\Lambda}_t^r$, $\widetilde{\boldsymbol{A}}$, and $\widetilde{\boldsymbol{B}}$ are diagonal matrices, and $\boldsymbol{\xi}_V$ contains the parameters. Note that for $n = 1$ and $\boldsymbol{R}_t^r = \rho_t \boldsymbol{I}_m$, Eq. (5) encompasses the DySARAR model considered in Catania and Billé (2017).

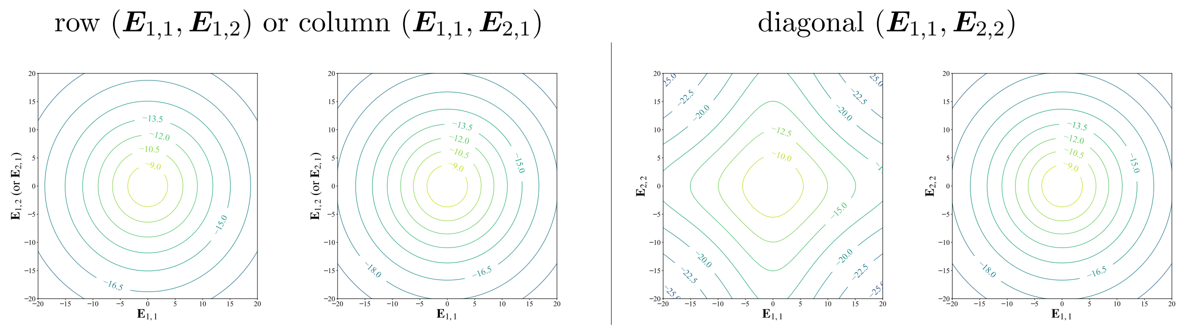
As a final note, not all parameters in (1) are identified. For instance, only the Kronecker products $\boldsymbol{B} \otimes \boldsymbol{R}_t^r \boldsymbol{W}_t^r$, $\boldsymbol{R}_t^c \boldsymbol{W}_t^c \otimes \boldsymbol{A}$, and $\boldsymbol{D} \otimes \boldsymbol{C}$ are identified. To address this issue, we impose the constraint $\boldsymbol{A}_{1,1} = \boldsymbol{B}_{1,1} = 1$. Together with the diagonality assumption for $\boldsymbol{A}$ and $\boldsymbol{B}$, this solves the identification issue for $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{R}_t^c$, and $\boldsymbol{R}_t^r$. As mentioned earlier, relaxing the diagonality assumption for $\boldsymbol{A}$ and $\boldsymbol{B}$ possibly requires further identification restrictions. We also impose $\boldsymbol{C}_{1,1} = 1$ to avoid the identification issue for $\boldsymbol{C}$ and $\boldsymbol{D}$.

2.2 Distributional assumptions

We assume that the error terms $\boldsymbol{E}_t$ follow the matrix Student's t distribution of Gupta and Nagar (2018, Chapter 4). As we see later, this endows the model with a two-fold robustness property. More specifically, for $\nu > 2$, we assume that the probability density function (pdf) of $\boldsymbol{E}_t$ is given by

$$\begin{aligned} p_E\left(\boldsymbol{E}_t | \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu\right) = & \frac{\Gamma_m\left(\frac{1}{2}(\nu + m + n - 1)\right)}{\Gamma_m\left(\frac{1}{2}(\nu + m - 1)\right)\left[(\nu - 2)\pi\right]^{\frac{mn}{2}}} \left|\boldsymbol{\Sigma}\right|^{-\frac{n}{2}} \left|\boldsymbol{\Omega}\right|^{-\frac{m}{2}} \\ & \times \left|\boldsymbol{I}_m + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^\top}{(\nu - 2)}\right|^{-\frac{1}{2}(\nu + m + n - 1)}, \end{aligned} \tag{6}$$

where $\Gamma_m(\cdot) = \pi^{m(m-1/4)} \prod_{i=1}^m \Gamma[a + (1 - i)/2]$ denotes the multivariate gamma function for $a > (m - 1)/2$, and $\Gamma(\cdot)$ the standard gamma function. We denote this zero-mean distribution as $\boldsymbol{E}_t \sim \boldsymbol{T}_{m,n}(\nu, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$. We assume the matrices $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$ to be diagonal, with the identifying restriction $\boldsymbol{\Omega}_{1,1} = 1$ if $m, n > 1$. The matrix-valued Student's t distribution collapses to the multivariate (vector-valued) Student's t distribution if either $m = 1$ or $n = 1$, and to the matrix-valued normal if $\nu \to \infty$. Moreover, $\mathrm{Cov}\left(\mathrm{vec}(\boldsymbol{E}_t)\right) = \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}$ for $\nu > 2$.



(a) matrix t (b) vector t (c) matrix t (d) vector t

Figure 1: Log-pdf contour plots for the matrix and vector-valued t distribution

Note: For the matrix-valued $\boldsymbol{E} \sim \boldsymbol{T}_{2,2}(5, 2\boldsymbol{I}_2, 2\boldsymbol{I}_2)$ and the multivariate t distribution $\text{vec}(\boldsymbol{E}) \sim \boldsymbol{t}_4(5, 4\boldsymbol{I}_4)$, panels 1(a) and 1(b) give contours of the log-pdf for a row (or column) of $\boldsymbol{E}$, the remaining elements set to zero. Panels 1(c) and 1(d) do the same for the diagonal of $\boldsymbol{E}$.

The assumption of a diagonal $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ implies a diagonal covariance structure for $\text{vec}\,(\boldsymbol{E}_t)$. As a result, the shocks $\boldsymbol{E}_t$ can be interpreted as idiosyncratic structural shocks to a particular row and column of $\boldsymbol{Y}_t$. The static ($\boldsymbol{A}$ and $\boldsymbol{B}$) and dynamic ($\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$) spillover matrices, along with the connection networks $\boldsymbol{W}_t^r$ and $\boldsymbol{W}_t^c$, then determine how these shocks propagate through the system.

Interestingly, the distribution of $\boldsymbol{E}_t$ is not merely a reshaped version of a vector-valued multivariate Student's t distribution for $\text{vec}(\boldsymbol{E}_t)$. In particular, if $m, n > 1$ and ν is finite, $\text{vec}(\boldsymbol{E}_t)$ does not follow a multivariate t with mean zero and covariance matrix $\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}$. To illustrate this, consider a toy example with $m = n = 2$, a matrix Student's t random variable $\boldsymbol{E} \sim \boldsymbol{T}_{2,2}(5, 2\boldsymbol{I}_m, 2\boldsymbol{I}_n)$, and a multivariate t $\text{vec}(\boldsymbol{E}) \sim \boldsymbol{t}_4(5, 4\boldsymbol{I}_{mn})$. Figure 1 plots the contours of these two log joint densities if we put two elements to zero and evaluate the density as a function of the other two arguments. Panels 1(a) and 1(b) set the bottom row elements to zero ($\boldsymbol{E}_{2,1} = \boldsymbol{E}_{2,2} = 0$), while panels 1(c) and 1(d) set the off-diagonal entries to zero ($\boldsymbol{E}_{1,2} = \boldsymbol{E}_{2,1} = 0$). Clearly, the first two panels show that the marginal densities for $\boldsymbol{E}_{2,1} = \boldsymbol{E}_{2,2} = 0$ are the same for the matrix and the multivariate t. The last two panels, however, show that the marginal densities for $\boldsymbol{E}_{1,2} = \boldsymbol{E}_{2,1} = 0$ are very different. Whereas the multivariate t retains its elliptical nature, the matrix t resembles much more the case of two independent univariate t distributions. The latter follows immediately from the density specification in (6). For $\boldsymbol{E}_{1,2} = \boldsymbol{E}_{2,1} = 0$ and the example from Figure 1, we obtain a density kernel of the form $\left|\boldsymbol{I}_2 + \frac{4}{3}\mathrm{diag}(\boldsymbol{E}_{1,1}^2, \boldsymbol{E}_{2,2}^2)\right|^{-5}$, which reduces to $\prod_{i=1}^{2} \left(1 + \frac{4}{3}\boldsymbol{E}_{i,i}^2\right)^{-5}$. The distribution thus induces more dependence if two entries $\boldsymbol{Y}_{i,j,t}$ and $\boldsymbol{Y}_{i',j',t}$ of $\boldsymbol{Y}_t$ have an overlapping row ($i = i'$) or column ($j = j'$) index (or both), and less if both their row and column indices differ. The matrix t thus more clearly exploits the matrix nature of the data than its multivariate vector-valued counterpart. We therefore use the matrix Student's t distribution for the error term $\boldsymbol{E}_t$. This also influences the spatial spillover dynamics via the score-driven specification for $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$, as we explain in the next section.

2.3 Score-driven spillover dynamics

To capture the time variation in the spatial spillover parameters, we use the score-driven framework introduced by Creal et al. (2011, 2013) and Harvey (2013). Let $\boldsymbol{f}_t = \left(\boldsymbol{f}_t^{r\top}, \boldsymbol{f}_t^{c\top}\right)^\top$, with $\boldsymbol{f}_t \in \mathbb{F} \subset \mathbb{R}^{m+n}$ containing the time-varying parameters of interest, and $\boldsymbol{f}_t^r$ and $\boldsymbol{f}_t^c$ containing the diagonal elements of $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$, respectively. Let $p_Y(\,\cdot \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s)$ denote the conditional pdf of $\boldsymbol{Y}_t$, where $\boldsymbol{\mathcal{X}}_t = (\boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c)$ collects the observed variables external to $\boldsymbol{Y}_t$, and $\boldsymbol{\theta}_s$ represents the vector of unknown static parameters in $(\boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$. The score-driven update for $\boldsymbol{f}_t$ is then given by

$$
\begin{aligned}
\boldsymbol{f}_{t+1} = \boldsymbol{\psi}_t\big(\boldsymbol{f}_t, \boldsymbol{\theta}\big) &= \boldsymbol{\omega} + \boldsymbol{\Phi}\boldsymbol{f}_t + \boldsymbol{K}\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big), \\
\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) &= \boldsymbol{S}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) \cdot \frac{\partial \log p_Y(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s)}{\partial \boldsymbol{f}_t},
\end{aligned} \tag{7}
$$

with intercept $\boldsymbol{\omega} \in \mathbb{R}^{m+n}$ and parameter matrices $\boldsymbol{\Phi}, \boldsymbol{K} \in \mathbb{R}^{(m+n)\times(m+n)}$ controlling the dynamics of $\boldsymbol{f}_t$. Here, $\boldsymbol{\theta}$ consists of all static parameters, including those in $(\boldsymbol{\omega}, \boldsymbol{\Phi}, \boldsymbol{K})$ as well as $\boldsymbol{\theta}_s$. Furthermore, the scaling matrix $\boldsymbol{S}_t(\cdot)$ is allowed to depend on $(\boldsymbol{Y}_t, \boldsymbol{\mathcal{X}}_t)$. Note that we do not impose any transformations on the elements of $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$. For example, Blasques et al. (2016) and Catania and Billé (2017) use tanh and logistic transformations, respectively, to constrain $\boldsymbol{R}_{i,i,t}^r$ and $\boldsymbol{R}_{i,i,t}^c$ within the unit interval to ensure system stability. However, such constraints can lead to significant losses in flexibility and are unsuitable in our empirical setting, where much more heterogeneity in the elements of $\boldsymbol{R}_t^c$ and $\boldsymbol{R}_t^r$ is called for; compare D'Innocenzo et al. (2024). This can be done without jeopardizing the model's spatial stability, as we see later.

To proceed, let $\boldsymbol{\mathcal{P}}_K^{(j)} \in \mathbb{R}^{K\times K}$ denote a matrix with zeros everywhere except for a single 1 at the jth diagonal position for $j = 1, \ldots, K$. Note that $\boldsymbol{\mathcal{P}}_K^{(j)}\boldsymbol{Q}$ acts as a row selector that zeroes out all but the jth row from a $K \times K$ matrix $\boldsymbol{Q}$. We now have the following result.

Proposition 2.1. *Let $\boldsymbol{Y}_t$ be as defined in (1), with $\boldsymbol{E}_t \sim \boldsymbol{T}_{m,n}(\nu, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ for $\nu > 2$, where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are symmetric, invertible matrices. Recall $\boldsymbol{f}_t = \left(\boldsymbol{f}_t^{r\top}, \boldsymbol{f}_t^{c\top}\right)^\top$ with $\boldsymbol{f}_t^r$ and $\boldsymbol{f}_t^c$ containing the diagonal elements of $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$, respectively, and define*

$$\boldsymbol{G}_t = \boldsymbol{G}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) = \boldsymbol{B} \otimes \boldsymbol{R}_t^r \boldsymbol{W}_t^r + \boldsymbol{R}_t^c \boldsymbol{W}_t^c \otimes \boldsymbol{A}. \tag{8}$$

If the spectral radius $\varrho(\boldsymbol{G}_t)$ of $\boldsymbol{G}_t$ satisfies $\varrho(\boldsymbol{G}_t) < 1$, then the (scaled) score term in (7) with scaling matrix $\boldsymbol{S}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) = \left|\boldsymbol{I}_{mn} - \boldsymbol{G}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s)\right| \cdot \boldsymbol{I}_{mn}$ takes the form

$$\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) = \begin{pmatrix} \boldsymbol{s}_t^r\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) \\ \boldsymbol{s}_t^c\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) \end{pmatrix} = \begin{pmatrix} |\boldsymbol{Z}_t| \operatorname{diag}\big(\boldsymbol{Y}_t^r \boldsymbol{B}^\top \widetilde{\boldsymbol{\mathcal{E}}}_t^\top \boldsymbol{\mathcal{W}}_t\big) - \boldsymbol{b}_t^r(\widetilde{\boldsymbol{Z}}_t, \boldsymbol{B}) \\ |\boldsymbol{Z}_t| \operatorname{diag}\big(\widetilde{\boldsymbol{\mathcal{E}}}_t^\top \boldsymbol{\mathcal{W}}_t \boldsymbol{A} \boldsymbol{Y}_t^c\big) - \boldsymbol{b}_t^c(\widetilde{\boldsymbol{Z}}_t, \boldsymbol{A}) \end{pmatrix}, \tag{9}$$

where $\boldsymbol{Y}_t^r = \boldsymbol{W}_t^r \boldsymbol{Y}_t$, $\boldsymbol{Y}_t^c = \boldsymbol{Y}_t \boldsymbol{W}_t^{c\top}$, $\widetilde{\boldsymbol{\mathcal{E}}}_t^\top = \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top \boldsymbol{\Sigma}^{-1}$, $\widetilde{\boldsymbol{Z}}_t = |\boldsymbol{Z}_t|\, \boldsymbol{Z}_t^{-1}$, $\boldsymbol{Z}_t = \boldsymbol{Z}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) =$

$\boldsymbol{I}_{mn} - \boldsymbol{G}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s)$, *and the residual* $\boldsymbol{\mathcal{E}}_t$ *and weight* $\boldsymbol{\mathcal{W}}_t$ *are defined as*

$$\boldsymbol{\mathcal{E}}_t = \boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) = \boldsymbol{Y}_t - \mathrm{diag}(\boldsymbol{f}_t^r)\boldsymbol{Y}_t^r\boldsymbol{B}^\top - \boldsymbol{A}\boldsymbol{Y}_t^c\mathrm{diag}(\boldsymbol{f}_t^c) - \boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top,$$
$$\boldsymbol{\mathcal{W}}_t = \boldsymbol{\mathcal{W}}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) = \frac{\nu+m+n-1}{\nu-2}\Big(\boldsymbol{I}_m + (\nu-2)^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s)^\top\boldsymbol{\Sigma}^{-1}\Big)^{-1},$$

with the bias-correction terms

$$\boldsymbol{b}_t^r\big(\widetilde{\boldsymbol{Z}}_t, \boldsymbol{B}\big) = \begin{pmatrix} \big[\mathrm{vec}(\boldsymbol{B}^\top \otimes \boldsymbol{W}_t^{r\top}\boldsymbol{\mathcal{P}}_m^{(1)})\big]^\top \\ \vdots \\ \big[\mathrm{vec}(\boldsymbol{B}^\top \otimes \boldsymbol{W}_t^{r\top}\boldsymbol{\mathcal{P}}_m^{(m)})\big]^\top \end{pmatrix} \mathrm{vec}\big(\widetilde{\boldsymbol{Z}}_t\big), \quad \boldsymbol{b}_t^c\big(\widetilde{\boldsymbol{Z}}_t, \boldsymbol{A}\big) = \begin{pmatrix} \big[\mathrm{vec}(\boldsymbol{W}_t^{c\top}\boldsymbol{\mathcal{P}}_n^{(1)} \otimes \boldsymbol{A}^\top)\big]^\top \\ \vdots \\ \big[\mathrm{vec}(\boldsymbol{W}_t^{c\top}\boldsymbol{\mathcal{P}}_n^{(n)} \otimes \boldsymbol{A}^\top)\big]^\top \end{pmatrix} \mathrm{vec}\big(\widetilde{\boldsymbol{Z}}_t\big).$$

The raw score in Proposition 2.1 is scaled by $|\boldsymbol{I}_{mn} - \boldsymbol{G}_t|$, such that the step sizes reduce as the spectral radius of $\boldsymbol{G}_t$ approaches one; compare the scaling in D'Innocenzo et al. (2024). Note that the expressions remain valid also for more general specifications of the model than the ones considered in this paper, e.g., for non-diagonal $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$. The scores $\boldsymbol{s}_t^r$ and $\boldsymbol{s}_t^c$ each consist of two terms that have an intuitive interpretation. The first terms arise from the matrix-valued linear regression model with time-varying coefficients, as described in (1). For instance, the first term in the first row of (9) adjusts $\boldsymbol{R}_{i,i,t}^r$ to be more in line with the new observation $(\boldsymbol{Y}_t^r\boldsymbol{B}^\top, \boldsymbol{Y}_t)$ conditional on the other parameters of the model. The second term then corrects the first step for the bias arising from the endogeneity due to the contemporaneous network structure. Finally, the weight matrix $\boldsymbol{\mathcal{W}}_t$ mitigates the effect of outliers in $\boldsymbol{Y}_t$ on the scores, thereby ensuring that the filter remains robust; compare Creal et al. (2013), Blasques et al. (2016), and Zheng et al. (2023). As $\boldsymbol{\mathcal{E}}_t$ diverges to infinity, the weight $\boldsymbol{\mathcal{W}}_t$ tends to zero for finite ν. Note that for $\nu \to \infty$, we have $\boldsymbol{\mathcal{W}}_t = \boldsymbol{I}_m$ for all t.

Note that also the score-driven dynamics for $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$ explicitly exploit the matrix structure of the data as opposed to modeling the data in vectorized form. This induces a substantial reduction in the number of parameters. For instance, with regard to spillover effects we now have $m+n$ time-varying parameters rather than the $m \cdot n$ time-varying parameters we would have under a vectorized form. Moreover, we only have $m+n-1$ covariance related parameters in $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ rather than $m \cdot n$ of them. This considerably simplifies the estimation problem.

2.4 Maximum likelihood estimation

The static parameter vector $\boldsymbol{\theta}$, which includes *all* the static parameters in the model, is estimated via maximum likelihood estimation (MLE). We initialize the filter recursion at a fixed nonrandom value $\widehat{\boldsymbol{f}}_0 \in \mathbb{F}$, and denote the resulting initialized sequence by $\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}) = \widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}, \widehat{\boldsymbol{f}}_0)$. For each t, the log-likelihood contribution is given by

$$
\begin{aligned}
\ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big) &= \log p_Y\big(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big) \\
&= \log \frac{\Gamma_m\big((\nu+m+n-1)/2\big)}{\Gamma_m\big((\nu+m-1)/2\big)\big((\nu-2)\pi\big)^{mn/2}} - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{m}{2}\log|\boldsymbol{\Omega}| + \log\big|\boldsymbol{Z}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\big| \\
&- \frac{(\nu+m+n-1)}{2}\log\Big|\boldsymbol{I}_m + (\nu-2)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)^\top\Big|.
\end{aligned}
$$

The MLE is then defined as

$$
\widehat{\boldsymbol{\theta}}_T = \widehat{\boldsymbol{\theta}}_T\big(\widehat{\boldsymbol{f}}_0\big) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, \widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, T^{-1}\sum_{t=1}^{T}\ell_t\big(\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big). \tag{10}
$$

Note that for a given value of $\boldsymbol{\theta}$, the evaluation of (10) is straightforward given the recursive nature of the data. The MLE can thus be found numerically by standard optimization techniques. Additionally, the estimation framework can be readily extended to incorporate conditional heteroskedasticity; further details are provided in Online Appendix F.

3 Asymptotic analysis

In this section we study some of the theoretical properties of the TVP-SMAR model, as well as the asymptotic behavior of the MLE defined in Section 2.4. We first show in Section 3.1 that the stochastic recurrence equation (SRE) in (7) admits a unique strictly stationary and ergodic (SE) solution, which ensures the existence of an SE process $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$. We then demonstrate that the initialized filter sequence $\big\{\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}) = \widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}, \widehat{\boldsymbol{f}}_0),\, t = 1, \ldots, T\big\}$, when applied to SE data, converges uniformly in $\boldsymbol{\theta}$ to an SE limit for any initial value $\widehat{\boldsymbol{f}}_0 \in \mathbb{F}$. Building on these results, Section 3.2 establishes the strong consistency and asymptotic normality of the MLE $\widehat{\boldsymbol{\theta}}_T$ in (10). Throughout, for a vector $\boldsymbol{a} = (a_j) \in \mathbb{R}^n$, its p-norm is denoted by $\|\boldsymbol{a}\|_p = (\sum_{j=1}^n |a_j|^p)^{1/p}$. For a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_p$ represents the induced p-norm (i.e., $\|\boldsymbol{A}\|_p = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p$), where the subscript is omitted whenever $p = 2$. For a square matrix $\boldsymbol{A}$, let $\varrho(\boldsymbol{A})$ be its spectral radius. Moreover, let $\log^+(x) = \max\{\log(x), 0\}$ for $x > 0$, and let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues, respectively.

3.1 Stationarity and filter invertibility

We begin by introducing the following assumptions.

Assumptions: *A1 There exists a complete and separable metric space* $(\mathbb{F}, \|\cdot\|)$ *a.s. such that* $\mathbb{F} \subset \big\{\boldsymbol{f} \in \mathbb{R}^{m+n} : \sup_{(t,\boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}} \varrho\big(\boldsymbol{G}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big) < 1\big\}$.

A2 $\forall \boldsymbol{f} \in \mathbb{F}$, $\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big\|\big[\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big]^{-1}\big\|\Big) < \infty$*;*

A3 The joint process $\big\{(\boldsymbol{E}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c),\, t \in \mathbb{Z}\big\}$ *is strictly stationary and ergodic. Moreover, for* $t \in \mathbb{Z}$*,*

(a) the errors $\boldsymbol{E}_t$ *are i.i.d. and follow a zero-mean matrix-valued* t *distribution with pdf given in* (6)*, degrees of freedom* $\nu > 2$*, and real symmetric spread matrices*

$\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ satisfying $0 < \lambda_L \leq \lambda_{\min}(\boldsymbol{\Sigma}) \leq \lambda_{\max}(\boldsymbol{\Sigma}) \leq \lambda_U < \infty$ and $0 < \lambda_L \leq \lambda_{\min}(\boldsymbol{\Omega}) \leq \lambda_{\max}(\boldsymbol{\Omega}) \leq \lambda_U < \infty$, for some constants λ_L and λ_U;

(b) *the exogenous regressors $\boldsymbol{X}_t$ satisfy $\mathbb{E}\big(\log^+ \|\boldsymbol{X}_t\|\big) < \infty$;*

(c) *the (observable) spatial weight matrices $\boldsymbol{W}_t^r$ and $\boldsymbol{W}_t^c$ have zero diagonal elements and satisfy the moment condition $\mathbb{E}\big(\log^+ \|\boldsymbol{W}_t^r\|\big) + \mathbb{E}\big(\log^+ \|\boldsymbol{W}_t^c\|\big) < \infty$.*

Assumption A1 requires the existence of a Polish space in which the filters reside, ensuring both the stability of $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$ and the invertibility of $\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) = \boldsymbol{I}_{mn} - \boldsymbol{G}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)$ for all $\boldsymbol{f} \in \mathbb{F}$. This assumption is typically not hard to verify. For instance, note that $\sup_{(t,\boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}} \varrho\big(\boldsymbol{G}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big) \leq \|\boldsymbol{f}\|_\infty \sup_{(t,\boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}} \big(\|\boldsymbol{B}\|\, \|\boldsymbol{W}_t^r\| + \|\boldsymbol{A}\|\, \|\boldsymbol{W}_t^c\|\big)$. If there exists a constant $C_0 \in (0,1)$ such that $\sup_{(t,\boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}} \big(\|\boldsymbol{B}\|\, \|\boldsymbol{W}_t^r\| + \|\boldsymbol{A}\|\, \|\boldsymbol{W}_t^c\|\big) \leq C_0$, we can take $\mathbb{F} = \{\boldsymbol{f} \in \mathbb{R}^{m+n} : \|\boldsymbol{f}\|_\infty \leq C_0^{-1}\}$, which is a closed subset of the finite dimensional Euclidean space $\mathbb{R}^{m+n}$ and therefore (under the norm $\|\cdot\|$) complete and separable.

Assumption A2 is required to guarantee the existence of a logarithmic moment for the process $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$. Note that Assumption A1 alone does not imply the moment condition in Assumption A2. Essentially, for $\boldsymbol{f} \in \mathbb{F}$, Assumption A1 ensures strict positiveness and boundedness of the absolute eigenvalues of $\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)$ uniformly in $(t, \boldsymbol{\theta})$. However, since $\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)$ is not symmetric, this does not necessarily provide control over the individual elements of its inverse (and also its adjoint).

Assumption A3 requires that the random components in (1) are jointly SE. It can be shown that $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$ is SE in our context with some additional arguments. Assumptions A3(a)-A3(c) further impose (moment) conditions on each random component, all of which are rather mild.

The following result establishes that the score-driven dynamics in (7) admit an SE solution as a potential data generating process (DGP).

Proposition 3.1 (Existence of a DGP)**.** *Consider the stochastic recurrence equation (SRE) in Eq. (7), with the updating step provided in Proposition 2.1. Let the SRE be evaluated at* $\widehat{\boldsymbol{f}}_t^{\star} = \big(\widehat{\boldsymbol{f}}_t^{\star r\top}, \widehat{\boldsymbol{f}}_t^{\star c\top}\big)^\top \in \mathbb{F}$ *as* $\widehat{\boldsymbol{f}}_{t+1}^{\star} = \boldsymbol{\omega} + \boldsymbol{\Phi}\widehat{\boldsymbol{f}}_t^{\star} + \boldsymbol{K}\boldsymbol{s}_t\big(\boldsymbol{Y}_t^{\star}, \widehat{\boldsymbol{f}}_t^{\star}, \boldsymbol{\theta}_s\big)$, *with nonrandom initialization* $\widehat{\boldsymbol{f}}_0^{\star} \in \mathbb{F}$, *where* $\boldsymbol{s}_t(\cdot)$ *is given in (9), and* $\boldsymbol{Y}_t^{\star} = \hat{\boldsymbol{R}}_t^{*r}\boldsymbol{W}_t^{r}\boldsymbol{Y}_t^{\star}\boldsymbol{B}^\top + \boldsymbol{A}\boldsymbol{Y}_t^{\star}\boldsymbol{W}_t^{c\top}\hat{\boldsymbol{R}}_t^{*c} + \boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top + \boldsymbol{E}_t$ *for* $\hat{\boldsymbol{R}}_t^{*r} = \mathrm{diag}(\widehat{\boldsymbol{f}}_t^{\star r})$ *and* $\hat{\boldsymbol{R}}_t^{*c} = \mathrm{diag}(\widehat{\boldsymbol{f}}_t^{\star c})$. *Then, for* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\widehat{\boldsymbol{f}}_t^{\star}$ *follows the SRE given by* $\widehat{\boldsymbol{f}}_{t+1}^{\star} = \boldsymbol{\psi}_t^{\star}\big(\widehat{\boldsymbol{f}}_t^{\star}, \boldsymbol{\theta}\big)$ *a.s., where*

$$\boldsymbol{\psi}_t^{\star}\big(\boldsymbol{f}, \boldsymbol{\theta}\big) = \boldsymbol{\omega} + \boldsymbol{\Phi}\boldsymbol{f} + \boldsymbol{K}\boldsymbol{s}_t^{\star}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big), \qquad \boldsymbol{s}_t^{\star}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big) = \begin{pmatrix} \boldsymbol{s}_t^{\star r}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big) \\ \boldsymbol{s}_t^{\star c}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big) \end{pmatrix}, \tag{11}$$

where $\boldsymbol{s}_t^{\star r}\big(\cdot, \boldsymbol{\theta}_s\big)$ *and* $\boldsymbol{s}_t^{\star c}\big(\cdot, \boldsymbol{\theta}_s\big)$ *and their derivatives are defined explicitly in the proof of this proposition and in Lemma E.1 .*

If Assumptions A1 - A3 hold and if, moreover, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

SE1 $\mathbb{E}\bigg(\log^{+}\sup_{\boldsymbol{f}\in\mathbb{F}}\bigg\|\boldsymbol{\Phi} + \boldsymbol{K}\,\dfrac{\partial}{\partial \boldsymbol{f}^\top}\boldsymbol{s}_t^{\star}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big)\bigg\|\bigg) < \infty$;

SE2 $\mathbb{E}\bigg(\log\sup_{\boldsymbol{f}\in\mathbb{F}}\bigg\|\dfrac{\partial}{\partial \boldsymbol{f}^\top}\boldsymbol{\psi}_t^{\star(r)}(\boldsymbol{f}, \boldsymbol{\theta})\bigg\|\bigg) < 0$ *for some integer* $r \geq 1$, *where* $\boldsymbol{\psi}_t^{\star(r)}(\,\cdot\,, \boldsymbol{\theta})$ *denotes the r-fold convolution of* $\boldsymbol{\psi}_t^{\star}\big(\,\cdot\,, \boldsymbol{\theta}\big)$ *as* $\boldsymbol{\psi}_t^{\star(r)}(\,\cdot\,, \boldsymbol{\theta}) = \boldsymbol{\psi}_t^{\star}\big(\,\cdot\,, \boldsymbol{\theta}\big) \circ \boldsymbol{\psi}_{t-1}^{\star}\big(\,\cdot\,, \boldsymbol{\theta}\big) \circ \ldots \circ \boldsymbol{\psi}_{t-r+1}^{\star}\big(\,\cdot\,, \boldsymbol{\theta}\big)$;

then, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *there exists a unique strictly stationary and ergodic solution* $\big\{\boldsymbol{f}_t^{\star}(\boldsymbol{\theta}),\, t \in \mathbb{Z}\big\}$ *with values in* $\mathbb{F}$, *as* $t \to \infty$. *Moreover, each* $\boldsymbol{f}_t^{\star}(\boldsymbol{\theta})$ *is measurable with respect to the* σ*-field generated by* $\big\{\boldsymbol{\psi}_{t-k}^{\star}\big(\,\cdot\,, \boldsymbol{\theta}\big),\, k \geq 1\big\}$.

The proof of Proposition 3.1 builds on fundamental results by Bougerol (1993) and Straumann and Mikosch (2006). Given the inherent complexity of our matrix-valued spatial model, as seen for instance by the construction of $\boldsymbol{s}_t^{\star}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big)$ in Lemma E.1, these assumptions are not simple to check analytically. Nevertheless, Assumptions SE1 - SE2 echo conditions found in the seminal work of Blasques et al. (2022) within the score-driven literature. Notably, the r-fold convolution in Assumption SE2 serves to enlarge the contraction region. When $r = 1$, this assumption reduces to those in D'Innocenzo et al. (2024) for the much simpler vector-valued setting.

Proposition 3.2 (Invertibility). *Let $\{\boldsymbol{Y}_t,\, t \in \mathbb{Z}\}$ be generated by (1), with dynamics specified in Proposition 2.1, for some "true" parameter $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Suppose the assumptions in Proposition 3.1 are satisfied and that $\boldsymbol{\Theta}$ is compact. Moreover, assume that*

IV1 $\mathbb{E}\left(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \boldsymbol{\Phi} + \boldsymbol{K} \dfrac{\partial}{\partial \boldsymbol{f}^\top} \boldsymbol{s}_t\big(\mathbf{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big) \right\| \right) < \infty;$

IV2 $\mathbb{E}\left(\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \dfrac{\partial}{\partial \boldsymbol{f}^\top} \boldsymbol{\psi}_t^{(r)}(\boldsymbol{f}, \boldsymbol{\theta}) \right\| \right) < 0$ *for some integer $r \geq 1$, with* $\boldsymbol{\psi}_t^{(r)}(\,\cdot\,, \boldsymbol{\theta}) = \boldsymbol{\psi}_t\big(\,\cdot\,, \boldsymbol{\theta}\big) \circ \boldsymbol{\psi}_{t-1}\big(\,\cdot\,, \boldsymbol{\theta}\big) \circ \ldots \circ \boldsymbol{\psi}_{t-r+1}\big(\,\cdot\,, \boldsymbol{\theta}\big)$ *the r-fold convolution of* $\boldsymbol{\psi}_t\big(\,\cdot\,, \boldsymbol{\theta}\big)$.

Then, the sequence $\big\{\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}),\, t \in \mathbb{Z}^+\big\}$, initialized at any nonrandom starting value $\widehat{\boldsymbol{f}}_0 \in \mathbb{F}$, converges exponentially fast almost surely (e.a.s.) to a unique SE sequence $\big\{\boldsymbol{f}_t(\boldsymbol{\theta}),\, t \in \mathbb{Z}\big\}$ with values in $\mathbb{F}$, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, as $t \to \infty$; that is, there exists a $\rho > 1$ such that $\rho^i \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\big\| \xrightarrow{a.s.} 0$ as $t \to \infty$.

Proposition 3.2 establishes that the initialized filter process converges uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ to a unique SE limiting process irrespective of the initial value. This result forms the basis for the consistency of the MLE and is known as the invertibility of the filter.

3.2 Strong consistency and asymptotic normality of the MLE

To establish the strong consistency of $\widehat{\boldsymbol{\theta}}_T$, some additional assumptions are required.

Assumptions: *SC1* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\Phi}\| < 1$ *and* $\sup_{(t,\boldsymbol{f},\boldsymbol{\theta})\in\mathbb{Z}\times\mathbb{F}\times\boldsymbol{\Theta}} \varrho\big(\boldsymbol{G}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big) < 1.$

SC2 For $0 < q_2 \leq q_1/2 \leq 1$, the following moment condition holds:

$$\mathbb{E}\left(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \left\| \big[\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big]^{-1} \right\|^{q_1} \right) + \mathbb{E}\|\boldsymbol{X}_t\|^{q_1} + \mathbb{E}\big\|\boldsymbol{W}_t^r\big\|^{q_2} + \mathbb{E}\big\|\boldsymbol{W}_t^c\big\|^{q_2} < \infty. \tag{12}$$

SC3 For all $t \in \mathbb{Z}$, $p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big) = p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}^\dagger, \boldsymbol{\theta}_s^\dagger\big)$ for almost every $\boldsymbol{Y} \in \mathbb{R}^{m \times n}$ with respect to the Lebesgue measure if and only if $\boldsymbol{f} = \boldsymbol{f}^\dagger$ and $\boldsymbol{\theta}_s = \boldsymbol{\theta}_s^\dagger$. Moreover, $\det(\boldsymbol{K}_0) \neq 0$, where $\boldsymbol{K}_0$ denotes the true parameter value of $\boldsymbol{K}$ in the SRE (7).

Assumption SC1 is needed to ensure the existence of a moment $\mathbb{E}\Big(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big\|\boldsymbol{f}_t(\boldsymbol{\theta})\big\|^p\Big) < \infty$ for some $p > 0$. Note that the condition $\sup_{(t,\boldsymbol{f},\boldsymbol{\theta}) \in \mathbb{Z} \times \mathbb{F} \times \boldsymbol{\Theta}} \varrho\big(\boldsymbol{G}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big) < 1$ in Assumption SC1 is stronger than Assumption A1. It essentially requires that there exists some $\varrho_U \in [0, 1)$ such that $\mathbb{F} \subset \big\{\boldsymbol{f} \in \mathbb{R}^{m+n} : \sup_{(t,\boldsymbol{\theta}) \in \mathbb{Z} \times \boldsymbol{\Theta}} \varrho\big(\boldsymbol{G}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big) \leq \varrho_U < 1\big\}$. This condition ensures that the eigenvalues of $\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)$ are uniformly bounded away from zero over all $(\boldsymbol{\theta}, \boldsymbol{f}) \in \boldsymbol{\Theta} \times \mathbb{F}$. The condition (12) in Assumption SC2 is stronger than those stated in Assumptions A2-A3, which together imply the existence of the $q_1 \in (0, 2]$ moment of $\boldsymbol{Y}_t$. Finally, Assumption SC3 imposes a generic condition for the identifiable uniqueness of the true static parameters. There are several common approaches to ensure identification. As discussed before, in this paper we achieve identifiability by normalizing $\boldsymbol{A}_{1,1} = \boldsymbol{B}_{1,1} = \boldsymbol{C}_{1,1} = \boldsymbol{\Omega}_{1,1} = 1$, with both $\boldsymbol{A}$ and $\boldsymbol{B}$ assumed to be diagonal matrices. Under this normalization, Assumption SC3 is satisfied, provided that there exists a $t \geq 1$ such that both $\boldsymbol{W}_t^r$ and $\boldsymbol{W}_t^c$ contain no zero rows. While neither $\boldsymbol{\Sigma}$ nor $\boldsymbol{\Omega}$ is required to be diagonal for identification, it is worth noting that we impose diagonality to facilitate a structural interpretation on the model's error terms $\boldsymbol{E}_t$.

The following lemma establishes that the empirical maximum likelihood objective function defined in (10), based on a nonrandom initial value $\widehat{\boldsymbol{f}}_0 \in \mathbb{F}$, converges uniformly and a.s. to a continuous objective as $T \to \infty$.

Lemma 1. *Recall $\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T \ell_t\big(\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$. Define $\mathcal{L}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T \ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$ and $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\big(\ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)\big)$. Under the assumptions of Proposition 3.2 and Assumptions*

SC1 - SC2, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0, \qquad T \to \infty, \tag{13}$$

where $\mathcal{L}(\cdot)$ is continuous on $\boldsymbol{\Theta}$.

Together, Lemma 1 and the identification condition in Assumption SC3 provide the key elements for establishing the strong consistency of $\widehat{\boldsymbol{\theta}}_T$.

Theorem 3.1 (Strong Consistency)**.** *Suppose the assumptions of Proposition 3.2 and Assumptions SC1 - SC3 hold. Then, for any filter initialization $\widehat{\boldsymbol{f}}_0 \in \mathbb{F}$, the maximum likelihood estimator defined in (10) is strongly consistent; that is, $\widehat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $T \to \infty$.*

In what follows, we introduce an additional set of assumptions to establish the asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$.

Assumptions: *AN1 The compact parameter space $\boldsymbol{\Theta}$ coincides with the closure of its open interior.*

AN2 There exists a constant $\varepsilon > 0$ and functions $g_0, g_1 \geq 0$ such that the following holds a.s.: for all $t \in \mathbb{Z}$, all $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$, and all $\boldsymbol{\theta} \in \mathbb{B}_\varepsilon(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \varepsilon\}$,

$$\left| p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big) - p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0}\big)\right| \leq g_0(\boldsymbol{Y}, \boldsymbol{\mathcal{X}}_t)\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|, \tag{14}$$

$$\left\| \frac{\partial\, p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{\theta}} - \frac{\partial\, p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0}\big)}{\partial \boldsymbol{\theta}} \right\| \leq g_1(\boldsymbol{Y}, \boldsymbol{\mathcal{X}}_t)\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|, \tag{15}$$

where $\int g_1(\boldsymbol{Y}, \boldsymbol{\mathcal{X}}_t)\, \mathrm{d}\boldsymbol{Y} < \infty$ and $\int g_2(\boldsymbol{Y}, \boldsymbol{\mathcal{X}}_t)\, \mathrm{d}\boldsymbol{Y} < \infty$.

AN3 $\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{f}\in\mathbb{F}} \Big\|\frac{\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_{s0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{f}^\top}\Big\|\Big) < \infty$ and $\mathbb{E}\Big(\Big\|\frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})}{\partial\boldsymbol{\theta}}\Big\|^2\Big) < \infty$.

AN4 Suppose that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \mathbb{E}\Big(\frac{\partial^2 \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big)\Big\| \xrightarrow{a.s.} 0$ as $T \to \infty$.

Assumption AN1 ensures that partial (high-order) derivatives with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ are well defined by ruling out "unusual" cases, such as when $\boldsymbol{\Theta}$ has an empty interior or consists of isolated points; see also Lin and Lucas (2025). Assumption AN2 guarantees that, for all $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$, integration and differentiation with respect to $\boldsymbol{\theta}$ can be interchanged a.s. for the maps $\boldsymbol{\theta} \mapsto p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$ and $\boldsymbol{\theta} \mapsto \partial\, p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)/\partial\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Note that, as usual, the integral of a scalar function of a matrix argument is interpreted as an iterated integral over each entry of the matrix; cf. Gupta and Nagar (2018, Chapter 1.4). Assumption AN3 imposes moment conditions needed to establish asymptotic normality, while Assumption AN4 requires that a uniform law of large numbers can be applied to the empirical Hessian matrix of the likelihood objective function. These two conditions can, in principle, be derived from more primitive assumptions. However, the associated proofs are extremely tedious in our matrix-valued setting, as they involve verifying the invertibility of the first- and second-order derivative processes of $\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, ultimately requiring computation of third-order derivatives of the scaled score term defined in (9). The expressions of these derivatives become increasingly long and opaque. Since these technicalities contribute little to the overall understanding of the asymptotic theory, we impose these conditions directly as assumptions. We expect no particular difficulties, however, given results for scalar and vector-valued models with score-driven dynamics; see for instance Blasques et al. (2016) and D'Innocenzo et al. (2024).

Theorem 3.2 below provides the asymptotic distribution of the estimator of the static parameters.

Theorem 3.2. *Suppose $\boldsymbol{\theta}_0 \in \operatorname{int}(\boldsymbol{\Theta})$, where $\operatorname{int}(\boldsymbol{\Theta})$ denotes the interior of $\boldsymbol{\Theta}$. Under the assumptions of Theorem 3.1 and Assumptions AN1 - AN4, we have $\sqrt{T}\big(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\big) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\mathcal{I}}_0^{-1}\big)$ as $T \to \infty$, provided that $\boldsymbol{\mathcal{I}}_0 = \mathbb{E}\Big(\frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})}{\partial \boldsymbol{\theta}^\top}\Big)$ is invertible.*

4 Simulation study

This section consists of two main parts. First, we examine whether the asymptotic results derived in Section 3 hold up in finite samples under correct model specification. Second, we assess the finite-sample performance of the proposed score-driven filters in terms of mean squared error.

Correct model specification. We set $(m, n) = (4, 3)$ and omit the exogenous variables. Let $\text{diag}(\boldsymbol{A}) = (1, 0.9, 0.1, 0.8)^\top$, $\text{diag}(\boldsymbol{B}) = (1, 0.8, 1.2)^\top$, $\boldsymbol{\Sigma} = 5 \cdot \boldsymbol{I}_m$, $\text{diag}(\boldsymbol{\Omega}) = (1, 0.8, 0.8)^\top$, $\boldsymbol{E}_t \sim \boldsymbol{T}_{4,3}(5, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$, where $\boldsymbol{\omega}$, $\boldsymbol{K}$, and $\boldsymbol{\Phi}$ in (7) are specified given by

$$\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\omega}_r \\ \boldsymbol{\omega}_c \end{bmatrix}, \qquad \boldsymbol{K} = \begin{bmatrix} k^r \cdot \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & k^c \cdot \boldsymbol{I}_n \end{bmatrix}, \qquad \boldsymbol{\Phi} = \begin{bmatrix} \phi^r \cdot \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & \phi^c \cdot \boldsymbol{I}_n \end{bmatrix},$$

with $\boldsymbol{\omega}_r = (-0.045, -0.015, 0.015, 0.045)^\top$, $\boldsymbol{\omega}_c = (-0.651, 0, 0.651)^\top$, $k_r = 0.05$, $k_c = 0.07$, $\phi_r = 0.95$, $\phi_c = 0.8$. We simulate a 4×4 row connection network $\boldsymbol{W}_t^r \equiv \boldsymbol{W}^r$, where $\text{vec}(\boldsymbol{W}^r)$ is simulated with $\boldsymbol{W}_{i,j}^r \overset{\text{i.i.d.}}{\sim} N(0, 1)$. We then remove the diagonal elements of $\boldsymbol{W}^r$ and perform column normalization. A similar procedure is followed for the 3×3 column connection network $\boldsymbol{W}_t^c \equiv \boldsymbol{W}^c$. Next, we fix $\boldsymbol{W}^r$ and $\boldsymbol{W}^c$, simulate 500 random data sets for each of the three sample sizes $T = 500, 1000, 2000$, and estimate the model for each simulated data set using the MLE defined in (10).

Figure 2 shows the kernel density estimates of a subset of the model parameters. Results are similar for the remaining parameters and are therefore omitted. We see that as the sample size T increases, the kernel density estimates center more and more tightly around the true value (indicated by the red vertical line). Also the shape of the distribution appears to converge to the normal distribution. This suggests that the asymptotic results of

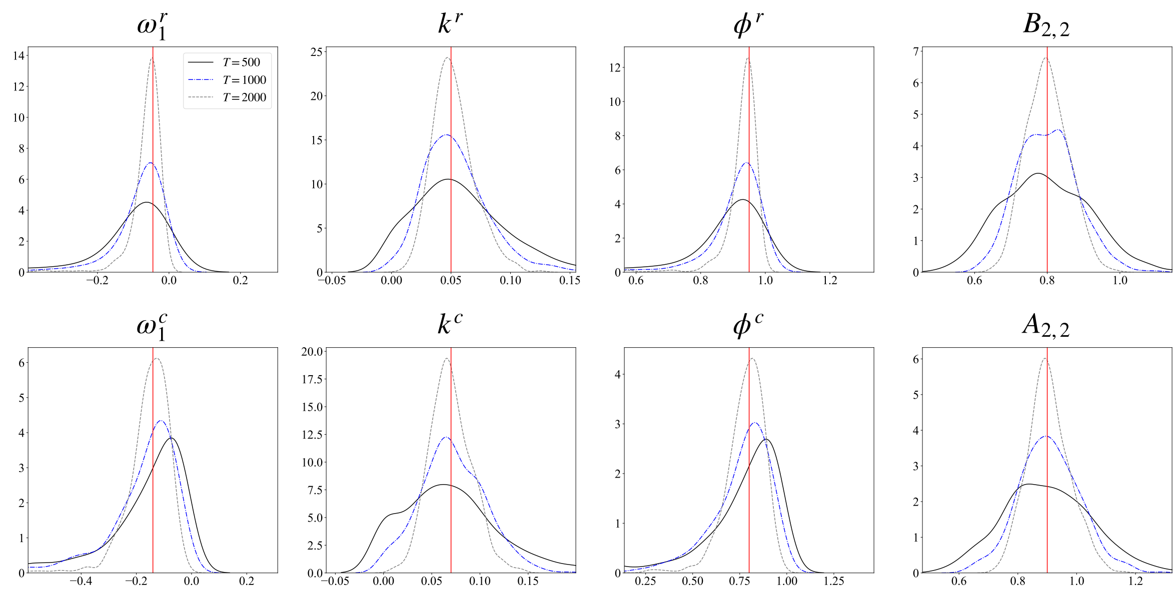


Figure 2: Kernel density estimates of the estimated parameters from 500 simulations for three sample sizes ($T = 500, 1000, 2000$) are presented, with vertical red lines indicating the true parameter values.

Theorem 3.2 provide a reasonable approximation and, thus, an adequate basis for inference in finite samples.

Model misspecification. To investigate whether our new model can successfully filter the spatial spillover dynamics if the model is misspecified, we consider a setting with $(m, n) = (2, 2)$, $\text{diag}(\boldsymbol{A}) = \text{diag}(\boldsymbol{B}) = (1, 1)^\top$, $\boldsymbol{\Sigma} = 5 \cdot \boldsymbol{I}_m$, and $\text{diag}(\boldsymbol{\Omega}) = (0.8, 0.8)^\top$. The connection matrices $\boldsymbol{W}^r$ and $\boldsymbol{W}^c$ follows the same procedure as before. We assume $\boldsymbol{E}_t \sim \boldsymbol{T}_{2,2}(3, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$. Given an initial value $\boldsymbol{R}_0^r$, we consider five types of dynamics for $\boldsymbol{R}_t^r = \text{diag}(\boldsymbol{R}_{1,1,t}^r, \boldsymbol{R}_{2,2,t}^r)$: (1) Const: $\boldsymbol{R}_{i,i,t}^r = \boldsymbol{R}_{i,i,0}^r$; (2) FastS: $\boldsymbol{R}_{i,i,t}^r = \boldsymbol{R}_{i,i,0}^r\big(1+0.6\cdot\sin(2\pi t/200)\big)$; (3) SlowS: $\boldsymbol{R}_{i,i,t}^r = \boldsymbol{R}_{i,i,0}^r\big(1+0.6\cdot\sin(2\pi t/20)\big)$; (4) Break: $\boldsymbol{R}_{i,i,t}^r = \boldsymbol{R}_{i,i,0}^r\big(1+0.9\mathbb{1}\{t > T/2\}\big)$; (5) AR: $\boldsymbol{R}_{i,i,t}^r = 0.01\boldsymbol{R}_{i,i,0}^r + 0.99\boldsymbol{R}_{i,i,t-1}^r + 0.01e_{i,t}$, where $e_{i,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ for $i = 1, 2$. We then generate $\boldsymbol{Y}_t$ according to (1), fixing the initial values as $\boldsymbol{R}_0^r = \boldsymbol{R}_0^c = \text{diag}(0.5, 0.5)$, and considering $T \in \{250, 500, 1000\}$. We set $\boldsymbol{R}_t^r = \boldsymbol{R}_t^c$ and let the two diagonal elements of $\boldsymbol{R}_t^r$ follow one of the (possibly) different dynamic specifications formulated above. For

Table 1: Empirical MSE of time-varying spatial spillover effects.

Dynamics	T	Static-SMAR-t	TVP-SMAR-t	Static-SMAR-$\mathcal{N}$	TVP-SMAR-$\mathcal{N}$
Const + Break	250	0.205	**0.040**	0.213	0.048
	500	0.201	**0.032**	0.209	0.040
	1000	0.201	**0.026**	0.208	0.035
SlowS + Break	250	0.229	**0.068**	0.243	0.075
	500	0.230	**0.048**	0.245	0.060
	1000	0.234	**0.051**	0.250	0.063
FastS + Break	250	0.236	**0.069**	0.253	0.081
	500	0.234	**0.063**	0.251	0.074
	1000	0.235	**0.059**	0.251	0.068
Break + AR	250	0.127	**0.042**	0.139	0.051
	500	0.127	**0.033**	0.138	0.041
	1000	0.124	**0.026**	0.133	0.033

each case, we repeat the simulation $M = 200$ times. We then compute the mean squared error (MSE) as $(4MT)^{-1}\sum_{t=1}^{T}\sum_{k=1}^{M}\left(\|\hat{\boldsymbol{R}}_t^{(k),r} - \boldsymbol{R}_t^{(k),r}\|^2 + \|\hat{\boldsymbol{R}}_t^{(k),c} - \boldsymbol{R}_t^{(k),c}\|^2\right)$, where $\hat{\boldsymbol{R}}_t^{(k),r}$ and $\boldsymbol{R}_t^{(k),r}$ denote the estimated and true paths of $\boldsymbol{R}_t^r$ in the kth replication, respectively, and similarly for $\hat{\boldsymbol{R}}_t^{(k),c}$ and $\boldsymbol{R}_t^{(k),c}$.

Table 1 displays a subset of the results. The full set of results can be found in Appendix G.1. Three key findings emerge. First, the TVP-SMAR-t model consistently achieves the lowest MSE across all scenarios and sample sizes, confirming its robustness and superior ability to capture heterogeneous time-varying spillover effects, even if the model is dynamically misspecified. Second, models based on the matrix Student's t distribution generally outperform their Gaussian counterparts, particularly in nonstationary environments such as Break and FastS. Third, estimation accuracy improves with larger sample sizes, despite the misspecification of the dynamics of $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$. We conclude that the new model can be estimated using typical empirical samples sizes and that the asymptotic theory provides a good framework for inference and for the convergence of the filtered paths.

5 Two empirical applications

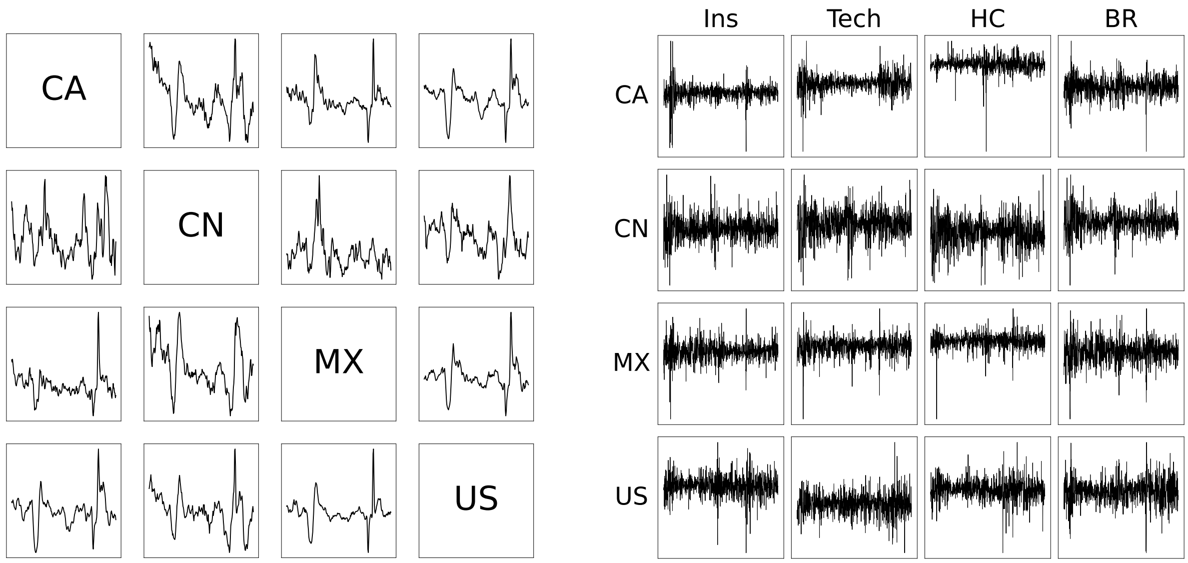
We study two applications. Section 5.1 examines an international bilateral import-export trade network, while Section 5.2 analyzes weekly sector-level global stock returns. The variables of interest in both applications are subject to incidental shocks.

5.1 Application 1: dynamic transport network

We study the monthly bilateral trade data between Canada (CA), Mainland China (CN), Mexico (MX), and United States (US) using the IMF-DOTS database and constructing a matrix valued import-export variable as in Cai et al. (2025). To account for discrepancies in reported import and export values between countries, we use the CIF (Cost, Insurance, and Freight) import data to represent trade flows between countries and regions. For example, the value of China's exports to the United States in January 2022 is based on the import figures reported by the United States for the same month. We compute the year-on-year growth rate of monthly imports.[2]

Figure 3(a) displays the monthly trade growth rate time series for the four economies, covering 241 time points from October 2004 to October 2024. For example, the time series in the first row and second column represents the monthly export from China (CN) to Canada (CA). The figure clearly shows that the import-export time series often exhibit outliers, particularly around the COVID-19 pandemic, where some countries' trade figures experienced significant jumps. This underlines the importance of employing a model that can account for such large fluctuations. For the connection network, we define $\boldsymbol{W}_t^r = \boldsymbol{W}_t^c = \boldsymbol{W}$, where $\boldsymbol{W}$ is a 4×4 matrix representing the inverse of geographical distances between countries, with distance measurements following the methodology described in Mayer and Zignago

[2]The year-on-year growth rate r_t is calculated by $r_t = (y_t - y_{t-12})/y_{t-12} \times 100$, where y_t denotes the monthly import level.



(a) Monthly import-export time series data (b) Weekly country-sector time series data

Figure 3: Data used in Section 5.1 (left) and Section 5.2 (right).

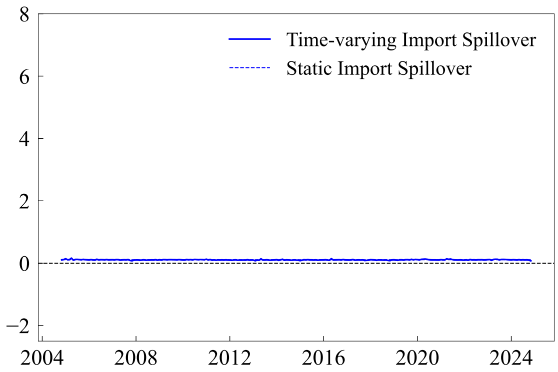
Table 2: Estimation results for global trade network dynamics.

	static scalar	dynamic scalar	static diagonal	dynamic diagonal
Loglik	-3805	-3792	-2988	**-2879**
AIC	7672	7650	6041	**5820**

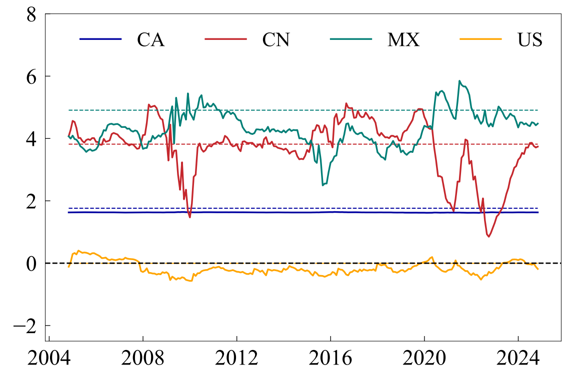
(2011). We normalize $\boldsymbol{W}$ by its largest eigenvalue. It is important to note that while the row-wise and column-wise connection matrices are identical, the economic interpretations of row-wise and column-wise spatial spillover effects differ. Specifically, $\boldsymbol{R}_t^r$ captures the time-varying spatial spillover effects from the perspective of imports, reflecting how a country's import growth rate is influenced by changes in the import growth rates of other countries. By contrast, $\boldsymbol{R}_t^c$ examines the export spillover effects, describing how a country's export growth rate is affected by fluctuations in other countries' export growth rates.

Table 2 presents the full-sample log-likelihood and AIC results for four different model specifications. The first two columns correspond to homogeneous models: the first assumes $\boldsymbol{R}_t^r = \rho^r \boldsymbol{I}_4$ and $\boldsymbol{R}_t^c = \rho^c \boldsymbol{I}_4$, while the second one allows for time-varying (scalar) spillover parameters, $\boldsymbol{R}_t^r = \rho_t^r \boldsymbol{I}_4$ and $\boldsymbol{R}_t^c = \rho_t^c \boldsymbol{I}_4$. The last two columns represent heterogeneous model specifications with diagonal $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$. To maintain parsimony, we assume that $\boldsymbol{K}$ is diagonal and that $\boldsymbol{\Phi} = \mathrm{diag}(\phi^r \boldsymbol{I}_m, \phi^c \boldsymbol{I}_n)$. Despite this pooling of some of the static parameters, the individual time-varying parameters in $\boldsymbol{R}_t^r$ and $\boldsymbol{R}_t^c$ are still allowed to be all different. Overall, compared to the static homogeneous model (column 1), incorporating heterogeneity (column 3) results in an increase of 817 points in the log-likelihood, while introducing time-varying parameters further improves the log-likelihood by an additional 109 points. A comparison of the first and second columns, as well as the third and fourth columns, reveals that dynamic models typically yield higher log-likelihoods and lower AIC values than their static counterparts. This highlights the importance of incorporating dynamic spillover effects in matrix network dynamics. Furthermore, a comparison of the first and third columns, as well as the second and fourth columns, shows that accounting for heterogeneity significantly improves model fit, thereby validating the benefits of incorporating individual heterogeneity at both the row and column levels from an in-sample fitting perspective. Both time-variation as well as row and column heterogeneity thus appear to be important stylized facts.

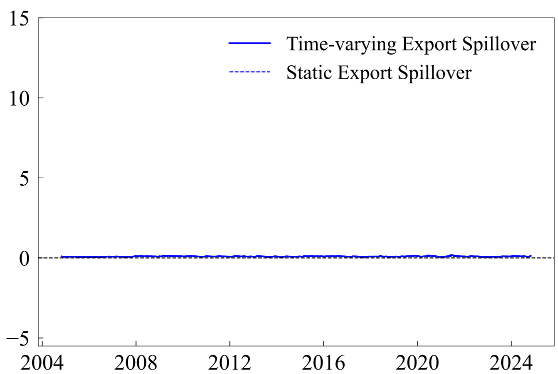
Figure 4 presents the estimated time-varying spatial spillover effects from the import and export perspectives under both homogeneous and heterogeneous scenarios. Comparing the left and right panels, the homogeneous models are too restrictive: the heterogeneous spatial spillovers exhibit much greater variability and substantial differences across countries. It is thus important to consider network dynamics within a heterogeneous rather than homogeneous matrix spatial autoregressive model setting. Comparing the import and export spillovers in Figures 4(b) and 4(d), we also see that the export spatial spillover effects are more dispersed than their import counterparts. In particular, from an imports



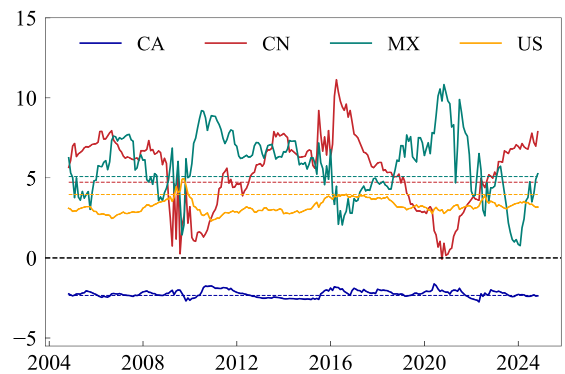
(a) Dynamic import spillover



(b) Country-specific import spillover



(c) Dynamic export spillover



(d) Country-specific export spillover

Figure 4: Filtered spatial autoregressive parameters for the import and export trade flow. The left and right panels are for the homogeneous and heterogeneous models, respectively. The sample ranges from October 2004 to October 2024.

perspective, Mexico (MX) experiences stronger positive spillover effects, while from an exports perspective, Canada (CA) experiences more pronounced negative spillover effects. We also see that there is strong time-variation in several of the individual spillover parameters (such as the US, Mexico, China), and somewhat less in others (like Canada).

Also note that the spillover parameters can exceed one or go below zero. The spectral radius $\varrho(\boldsymbol{G}_t)$, however, remains well below 1 in all cases, illustrating that the model remains spatially stable. We also see that pooling all the different types of spillover into a homogeneous (scalar) specification as in Figures 4(a) and 4(c) masks the wide cross-sectional variation in spillover strengths. In addition, it also reduces the time-variation, as some spatial spillovers increase at a time where others decrease. Due to the pooling, these off-setting effects cause the scalar spillover parameter to hardly move or not move at all. For instance, during 2015-2017, Mexico's and China's export spillover parameters show substantial variation and go through a trough and a peak period, respectively. The scalar model in Figure 4(c), however, hardly moves at all because it pools these offsetting parameters and their dynamics, and because a scalar parameter needs to remain between -1 and $+1$ for spatial stability.

5.2 Application 2: global equity market network

In our second application we highlight the usefulness of the restricted TVP-SMAR specification in Eq. (3) and consider weekly stock (log) return data from four sectors: Insurance (Ins), Technology (Tech), Health Care (HC), and Basic Resources (BR), across the same four countries examined in Section 5.1. The data come from the London Stock Exchange Group (LSEG) Datastream database and span the period from March 08, 2009, to March 10, 2024, comprising 784 weekly observations. Figure 3(b) shows the data.

Since column-wise network information is not available, we focus exclusively on row-wise spatial spillover effects, corresponding to country-level interactions. The row-wise time-varying connection matrix $\boldsymbol{W}_t^r$ is constructed using the time-varying trade figures between countries using the IMF-DOTS database. We normalize $\boldsymbol{W}_t^r$ by its largest eigenvalue at each time point.

Table 3 compares the in-sample performance of the four alternative model specifications. In line with our earlier findings, incorporating time-variation enhances the in-sample fit. We also see that the network heterogeneity adds less in this case. This outcome may be attributed to the heterogeneity that is already captured by the heterogeneity in the row-wise network connection matrix $\boldsymbol{W}_t^r$.

Table 3: Estimation results for global equity markets.

	static scalar	dynamic scalar	static diagonal	dynamic diagonal
Loglik	-16075.09	-15969.04	-16068.98	**-15946.24**
AIC	32240	32028	32228	**31994**

To investigate the out-of-sample performance of the different models, we split the data into a training set consisting of the first 680 observations, and a test set comprising the last 104 observations (about two years). For each week in the forecasting period, each model provides a forecast of next week's covariance matrix of the returns. We use these forecasts to construct a minimum variance portfolio (MVP) for each model and each week by solving the optimization problem:

$$\min_{\boldsymbol{w}_t \in \mathbb{R}^{16}} \boldsymbol{w}_t^\top \boldsymbol{Z}_{t+1}(\boldsymbol{\Omega}_{t+1} \otimes \boldsymbol{\Sigma}_{t+1})\boldsymbol{Z}_{t+1}^\top \boldsymbol{w}_t, \qquad \boldsymbol{Z}_t = \left(\boldsymbol{I}_{mn} - (\boldsymbol{B} \otimes \boldsymbol{R}_t^r \boldsymbol{W}_t^r)\right)^{-1},$$

where $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Sigma}_t$ are indexed by t in case we allow for conditional heteroskedasticity of the structural shocks; see Appendix F. We do so with and without short-sale constraints ($\boldsymbol{w}_t \geq 0$). We compare our model with the recent matrix factor GARCH (MF-GARCH) model of Yu et al. (2024), which shows better out-of-sample performance than traditional multivariate GARCH-type models.

Table 4 summarizes the out-of-sample annualized average returns (AV), their standard deviation (SD), the information ratio (IR = AV/SD), and the p-values of the model confidence set (MCS) test for each model. Details on the computation of these numbers are provided in Appendix G.2. For each measure, the model with the best performance is highlighted in bold. The results clearly show that accounting for both time-variation in spatial spillovers and heterogeneity delivers superior out-of-sample performance, yielding the highest AV, the lowest SD, the highest IR, and the highest p-value among all models

Table 4: Performance comparison of unconstrained and constrained MVP strategies

	Unconstrained MVP				**Constrained MVP**			
Model	**AV**	**SD**	**IR**	**MCS**	**AV**	**SD**	**IR**	**MCS**
dynamic diagonal	**3.65**	**12.09**	**2.18**	**1.00**	**3.88**	**12.01**	**2.33**	**1.00**
static diagonal	0.98	12.12	0.58	0.93	0.98	12.12	0.59	0.92
dynamic scalar	-1.90	12.75	-0.51	0.32	0.97	12.25	0.57	0.31
static scalar	0.92	12.18	0.54	0.32	0.92	12.18	0.54	0.31
MF-GARCH	-1.92	14.97	-0.92	0.01	-2.55	14.53	-1.12	0.01
equal weights	0.24	13.65	0.12	0.02	0.24	13.65	0.13	0.02

considered, including the MF-GARCH model.

6 Conclusion

In this paper we introduced the time-varying parameter spatial matrix autoregressive (TVP-SMAR) model. The model integrated matrix-valued time series analysis, time-varying spatial dependence, and spatial heterogeneity into one unified modeling framework. By explicitly exploiting the two-way spatial spillover effects along rows and columns, the model addresses the inherent complexities of matrix-valued time series data in a flexible, yet parsimonious way. The new model also leveraged the matrix-valued Student's t-distribution to provide robust score-driven dynamics for the time-varying spillovers and encompassed a wide range of earlier spatial models. In particular, the use of a matrix Student's t distribution sets the model apart from spatial autoregressions applied to vectorized matrix-valued observations, as it also exploits the differences between the row and column dimensions of the data.

Despite its flexibility, the model could still be estimated by straightforward maximum likelihood procedures, for which we proved the required consistency and asymptotic normality result. We also showed that the model was invertible, such that the estimated time-varying spillover paths can be recovered from the data and these paths converge exponentially fast almost surely to their true, limiting counterparts.

Empirically, we showcased the model's usefulness through two different applications. The first application based on international trade data demonstrated the importance of incorporating both dynamic and heterogeneous spatial spillover effects when analyzing trade relationships at the country level. It also revealed that the spatial spillovers can differ substantially across the row and column dimensions of the matrix-valued data. In our second application we examined global stock market data and applied a restricted version of the TVP-SMAR to portfolio selection problems. We showed that the TVP-SMAR with time-varying and heterogeneous spillovers outperformed various alternative specifications, including a recent matrix-valued factor GARCH model, in an out-of-sample setting. Both real-world applications underscore the model's effectiveness in parsimoniously capturing complex, dynamic interactions in large-dimensional contexts.

References

Anselin, L. (2022). Spatial Econometrics. *Handbook of Spatial Analysis in the Social Sciences*, 101–122.

Anselin, L., I. Syabri, and Y. Kho (2009). GeoDa: an introduction to spatial data analysis. In *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, pp. 73–89. Springer.

Aquaro, M., N. Bailey, and M. H. Pesaran (2021). Estimation and inference for spatial models with heterogeneous coefficients: an application to us house prices. *Journal of Applied Econometrics 36*(1), 18–44.

Asgharian, H., W. Hess, and L. Liu (2013). A spatial analysis of international stock market linkages. *Journal of Banking & Finance 37*(12), 4738–4754.

Beutner, E., Y. Lin, and A. Lucas (2023). Consistency, distributional convergence, and optimality of score-driven filters. Technical Report TI 2023-051/III, Tinbergen Institute Discussion Paper.

Blasques, F., P. Gorgi, S. J. Koopman, and J. Sampi (2023). Does trade integration imply growth in Latin America? Evidence from a dynamic spatial spillover model. Technical Report 2023-007/III, Tinbergen Institute Discussion Paper.

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

Blasques, F., S. J. Koopman, A. Lucas, and J. Schaumburg (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics 195*(2), 211–223.

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

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

Cai, X., X. Kong, X. Wu, and P. Zhao (2025). Matrix-factor-augmented regression. *Journal of Business & Economic Statistics*. In press.

Catania, L. and A. G. Billé (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics 32*(6), 1178–1196.

Chang, J., J. He, L. Yang, and Q. Yao (2023). Modelling matrix time series via a tensor CP-decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(1), 127–148.

Chen, B., E. Y. Chen, S. Bolivar, and R. Chen (2024). Time-varying matrix factor models. *arXiv preprint arXiv:2404.01546*.

Chen, E. Y. and R. Chen (2022). Modeling dynamic transport network with matrix factor models: an application to international trade flow. *Journal of Data Science 21*(3), 490–507.

Chen, R., H. Xiao, and D. Yang (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics 222*(1), 539–560.

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., S. J. Koopman, A. Lucas, and M. Zamojski (2024). Observation-driven filtering of time-varying parameters using moment conditions. *Journal of Econometrics 238*(2), 105635.

de Punder, R., T. Dimitriadis, and R.-J. Lange (2024). Kullback-Leibler-based characterizations of score-driven updates. *arXiv preprint arXiv:2408.02391*.

Denbee, E., C. Julliard, Y. Li, and K. Yuan (2021). Network risk and key players: A structural analysis of interbank liquidity. *Journal of Financial Economics 141*(3), 831–859.

D'Innocenzo, E., A. Lucas, A. Opschoor, and X. Zhang (2024). Heterogeneity and dynamics in network models. *Journal of Applied Econometrics 39*(1), 150–173.

Elhorst, J. P., M. Gross, and E. Tereanu (2021). Cross-sectional dependence and spillovers in space and time: Where spatial econometrics and global var models meet. *Journal of Economic Surveys 35*(1), 192–226.

Gasperoni, F., A. Luati, L. Paci, and E. D'Innocenzo (2023). Score-driven modeling of spatio-temporal data. *Journal of the American Statistical Association 118*(542), 1066–1077.

Gorgi, P., C. Lauria, and A. Luati (2023). On the optimality of score-driven models. *Biometrika*, asad067.

Gupta, A. K. and D. K. Nagar (2018). *Matrix Variate Distributions.* Chapman and Hall/CRC.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series.* Number 52. Cambridge University Press.

He, Y., X. Kong, L. Trapani, and L. Yu (2024). Online change-point detection for matrix-valued time series with latent two-way factor structure. *Annals of Statistics 52*(4), 1646–1670.

Heil, T. L., F. J. Peter, and P. Prange (2022). Measuring 25 years of global equity market co-movement using a time-varying spatial model. *Journal of International Money and Finance 128*, 102708.

Jiang, H., B. Shen, Y. Li, and Z. Gao (2024). Regularized estimation of high-dimensional matrix-variate autoregressive models. *arXiv preprint arXiv:2410.11320*.

Kou, S., X. Peng, and H. Zhong (2018). Asset pricing with spatial interaction. *Management Science 64*(5), 2083–2101.

LeSage, J. and R. K. Pace (2009). *Introduction to Spatial Econometrics.* Chapman and Hall/CRC.

Li, Z. and H. Xiao (2024). Cointegrated matrix autoregression models. *arXiv preprint arXiv:2409.10860*.

Lin, Y. and A. Lucas (2025). Functional location-scale models with robust observation-driven dynamics. Technical Report TI 2025-027/III, Tinbergen Institute Discussion Paper.

Mayer, T. and S. Zignago (2011). Notes on CEPII's distances measures: The GeoDist database.

Pu, D., K. Fang, W. Lan, J. Yu, and Q. Zhang (2024). Multivariate spatiotemporal models with low rank coefficient matrix. *Journal of Econometrics 246*(1-2), 105897.

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

Yang, K. and L.-f. Lee (2021). Estimation of dynamic panel spatial vector autoregression: Stability and spatial multivariate cointegration. *Journal of Econometrics 221*(2), 337–367.

Yu, C., D. Li, F. Jiang, and K. Zhu (2024). Matrix GARCH model: Inference and application. *Journal of the American Statistical Association*, 1–31.

Yu, C., D. Li, X. Zhang, and H. Tong (2024). Two-way threshold matrix autoregression. *arXiv preprint arXiv:2407.10272*.

Yu, H., S. Ye, and T. Zheng (2025). Time-varying parameter MAR model with an application to multi-country economic network. *working paper*.

Zhang, H.-F. (2024). Additive autoregressive models for matrix valued time series. *Journal of Time Series Analysis 45*(3), 398–420.

Zhang, X. and K.-S. Chan (2024). Adaptive matrix change point detection: Leveraging structured mean shifts. *arXiv preprint arXiv:2401.17473*.

Zheng, T., S. Ye, and Y. Hong (2023). Fast estimation of a large TVP-VAR model with score-driven volatilities. *Journal of Economic Dynamics and Control 157*, 104762.

Supplementary Materials for "Matrix-Valued Spatial Autoregressions with Dynamic and Robust Heterogeneous Spillovers"

Yicong Lin[a], *Andre Lucas*[a], *Shiqi Ye*[b]

[a]: *Vrije Universiteit Amsterdam* and *Tinbergen Institute*

[b]: *AMSS Center for Forecasting Science, Chinese Academy of Sciences*

This supplement includes all technical proofs and additional results.

Contents

Further Notation: Recall that a vector-valued sequence $\{\boldsymbol{a}_t,\, t \in \mathbb{Z}\}$ is said to converge to zero exponentially fast almost surely (e.a.s.) as $t \to \infty$, denoted by $\boldsymbol{a}_t \xrightarrow{e.a.s.} \mathbf{0}$, if there exists some $\gamma > 1$ such that $\gamma^t \|\boldsymbol{a}_t\| \xrightarrow{a.s.} 0$, where $\xrightarrow{a.s.}$ denotes almost sure (a.s.) convergence. To facilitate the proofs in the sections that follow, we introduce some additional notation: (i) For a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_F$ denotes the Frobenius norm (i.e., $\|\boldsymbol{A}\|_F = \sqrt{\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{A})}$). (ii) The symbol C denotes a generic positive constant, which may differ from one occurrence to another. (iii) Consider a complete separable metric space denoted by (E, d_E). Following Bougerol (1993, Section 3), we define the Lipschitz coefficient ρ associated with a random map $\varphi : E \to E$ as:

$$\rho(\varphi) = \sup_{x,y \in E,\, x \neq y} \left\{ \frac{d_E\big(\varphi(x), \varphi(y)\big)}{d_E(x, y)} \right\}. \tag{N1}$$

It is important to note that if φ is measurable, then so is $\rho(\varphi)$ (Bougerol, 1993, p. 955). This definition will be used repeatedly in the subsequent sections, with the specific space E being identified in each instance. (iv) We adopt the "good" notation for matrix derivatives as suggested by Magnus and Neudecker (2019), and define:

$$\mathrm{D}_{\boldsymbol{X}} \boldsymbol{F}(\boldsymbol{X}) = \frac{\partial \mathrm{vec}(\boldsymbol{F}(\boldsymbol{X}))}{\partial \mathrm{vec}(\boldsymbol{X})^\top}. \tag{N2}$$

A Proof of Proposition 2.1

In what follows, we use $|\det(\boldsymbol{Q})|$ to denote the absolute value of the determinant of a matrix $\boldsymbol{Q}$, and $|\boldsymbol{Q}|$ to denote the determinant itself. Recall $\boldsymbol{G}_t = \boldsymbol{B} \otimes \operatorname{diag}(\boldsymbol{f}_t^r)\boldsymbol{W}_t^r + \operatorname{diag}(\boldsymbol{f}_t^c)\boldsymbol{W}_t^c \otimes \boldsymbol{A}$ and $\boldsymbol{\mathcal{E}}_t = \boldsymbol{Y}_t - \operatorname{diag}(\boldsymbol{f}_t^r)\boldsymbol{W}_t^r\boldsymbol{Y}_t\boldsymbol{B}^\top - \boldsymbol{A}\boldsymbol{Y}_t\boldsymbol{W}_t^{c\top}\operatorname{diag}(\boldsymbol{f}_t^c) - \boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top$. Define $\boldsymbol{\Xi}_t = \boldsymbol{\mathcal{E}}_t\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t^\top$. Note that $\boldsymbol{Y}_t$ is a linear transformation of $\boldsymbol{E}_t$ as defined in (1). Write $\boldsymbol{f}_t^r = (f_{1,t}^r, \ldots, f_{m,t}^r)^\top$ and $\boldsymbol{f}_t^c = (f_{1,t}^c, \ldots, f_{n,t}^c)^\top$. By Gupta and Nagar (2018, Chapter 1.3), one has $p_Y\big(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) = p_E\big(\boldsymbol{\mathcal{E}}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big)\Big|\det\left(\mathrm{D}_{\boldsymbol{Y}_t}\boldsymbol{\mathcal{E}}_t\right)\Big| = p_E\big(\boldsymbol{\mathcal{E}}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big)\big|\det(\boldsymbol{I}_{mn} - \boldsymbol{G}_t)\big|$. Given that $\varrho(\boldsymbol{G}_t) < 1$, the determinant of $\boldsymbol{I}_{mn} - \boldsymbol{G}_t$ is positive. Then, the log conditional density of $\boldsymbol{Y}_t$ is given by:

$$\log p_Y\big(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) = \log \frac{\Gamma_m\big((\nu+m+n-1)/2\big)}{\Gamma_m\big((\nu+m-1)/2\big)\big((\nu-2)\pi\big)^{mn/2}} - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{m}{2}\log|\boldsymbol{\Omega}|$$
$$+ \log|\boldsymbol{I}_{mn} - \boldsymbol{G}_t| - \frac{(\nu+m+n-1)}{2}\log\Big|\boldsymbol{I}_m + (\nu-2)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}_t\Big|. \quad \text{(A.1)}$$

Note that $\mathrm{d}\log|\boldsymbol{X}| = \operatorname{tr}(\boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X})$ for any matrix $\boldsymbol{X}$. This implies

$$\mathrm{d}\log p_Y\big(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) = -\frac{1}{2}\operatorname{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{W}}_t\,\mathrm{d}\boldsymbol{\Xi}_t\big) - \operatorname{tr}\big(\boldsymbol{Z}_t^{-1}\mathrm{d}\boldsymbol{G}_t\big) =: P_{1,t} - P_{2,t}, \quad \text{(A.2)}$$

where $\boldsymbol{\mathcal{W}}_t = \frac{\nu+m+n-1}{\nu-2}\big(\boldsymbol{I}_m + (\nu-2)^{-1}\boldsymbol{\Xi}_t\boldsymbol{\Sigma}^{-1}\big)^{-1}$, and $\boldsymbol{Z}_t = \boldsymbol{I}_{mn} - \boldsymbol{G}_t$ (also defined in the proposition).

For $P_{1,t}$, note that $-\mathrm{d}\boldsymbol{\mathcal{E}}_t = \sum_{j=1}^m \boldsymbol{\mathcal{P}}_m^{(j)}\boldsymbol{W}_t^r\boldsymbol{Y}_t\boldsymbol{B}^\top\mathrm{d}f_{j,t}^r + \sum_{j=1}^n \boldsymbol{A}\boldsymbol{Y}_t\boldsymbol{W}_t^{c\top}\boldsymbol{\mathcal{P}}_n^{(j)}\mathrm{d}f_{j,t}^c$, where $\boldsymbol{\mathcal{P}}_K^{(j)}$ denotes $K \times K$ matrix with zeros everywhere except for a single 1 at the j_{th} diagonal position for any $K \in \mathbb{Z}^+$ and $j = 1, \ldots, K$. Then, we have

$$-\mathrm{d}\boldsymbol{\Xi}_t = (-\mathrm{d}\boldsymbol{\mathcal{E}}_t)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t^\top + \boldsymbol{\mathcal{E}}_t\boldsymbol{\Omega}^{-1}(-\mathrm{d}\boldsymbol{\mathcal{E}}_t)^\top$$

$$
\begin{aligned}
&= \sum_{j=1}^{m} \Big(\boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top + \boldsymbol{\mathcal{E}}_t \boldsymbol{\Omega}^{-1} \boldsymbol{B} \boldsymbol{Y}_t^\top \boldsymbol{W}_t^{r\top} \boldsymbol{\mathcal{P}}_m^{(j)} \Big) \mathrm{d} f_{j,t}^r \\
&\quad + \sum_{j=1}^{n} \Big(\boldsymbol{A} \boldsymbol{Y}_t \boldsymbol{W}_t^{c\top} \boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top + \boldsymbol{\mathcal{E}}_t \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{W}_t^c \boldsymbol{Y}_t^\top \boldsymbol{A}^\top \Big) \mathrm{d} f_{j,t}^c .
\end{aligned}
$$

Since $\boldsymbol{\Sigma}$ is symmetric, $\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t = \frac{\nu+m+n-1}{\nu-2} \big(\boldsymbol{\Sigma} + (\nu-2)^{-1} \boldsymbol{\Xi}_t \big)^{-1}$ is also symmetric. It implies that $\operatorname{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t (\boldsymbol{Q} + \boldsymbol{Q}^\top) \big) = \operatorname{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \boldsymbol{Q} \big) + \operatorname{tr}\big((\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t)^\top \boldsymbol{Q})^\top \big) = 2\operatorname{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \boldsymbol{Q} \big)$ for any $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$. Using this identity and given the symmetry of $\boldsymbol{\Omega}$, it follows that

$$
\begin{aligned}
P_{1,t} &= \frac{1}{2} \sum_{j=1}^{m} \operatorname{tr}\Big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \big(\boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top + \boldsymbol{\mathcal{E}}_t \boldsymbol{\Omega}^{-1} \boldsymbol{B} \boldsymbol{Y}_t^\top \boldsymbol{W}_t^{r\top} \boldsymbol{\mathcal{P}}_m^{(j)} \big) \Big) \mathrm{d} f_{j,t}^r \\
&\quad + \frac{1}{2} \sum_{j=1}^{n} \operatorname{tr}\Big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \big(\boldsymbol{A} \boldsymbol{Y}_t \boldsymbol{W}_t^{c\top} \boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top + \boldsymbol{\mathcal{E}}_t \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{W}_t^c \boldsymbol{Y}_t^\top \boldsymbol{A}^\top \big) \Big) \mathrm{d} f_{j,t}^c \\
&= \sum_{j=1}^{m} \operatorname{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top \big) \mathrm{d} f_{j,t}^r + \sum_{j=1}^{n} \operatorname{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \boldsymbol{A} \boldsymbol{Y}_t \boldsymbol{W}_t^{c\top} \boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top \big) \mathrm{d} f_{j,t}^c \\
&= \sum_{j=1}^{m} \big[\boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \big]_{j,j} \mathrm{d} f_{j,t}^r + \sum_{j=1}^{n} \big[\boldsymbol{\Omega}^{-1} \boldsymbol{\mathcal{E}}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{W}}_t \boldsymbol{A} \boldsymbol{Y}_t \boldsymbol{W}_t^{c\top} \big]_{j,j} \mathrm{d} f_{j,t}^c ,
\end{aligned}
$$

where $[\boldsymbol{Q}]_{k,\ell}$ denotes the $(k,\ell)_{\mathrm{th}}$ element of the matrix $\boldsymbol{Q}$.

Similarly, we obtain

$$
\begin{aligned}
P_{2,t} &= \sum_{j=1}^{m} \operatorname{tr}\Big(\boldsymbol{Z}_t^{-1} \big(\boldsymbol{B} \otimes \boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{W}_t^r \big) \Big) \mathrm{d} f_{j,t}^r + \sum_{j=1}^{n} \operatorname{tr}\Big(\boldsymbol{Z}_t^{-1} \big(\boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{W}_t^c \otimes \boldsymbol{A} \big) \Big) \mathrm{d} f_{j,t}^c \\
&= \sum_{j=1}^{m} \big[\operatorname{vec}\big(\boldsymbol{B}^\top \otimes \boldsymbol{W}_t^{r\top} \boldsymbol{\mathcal{P}}_m^{(j)} \big) \big]^\top \operatorname{vec}\big(\boldsymbol{Z}_t^{-1} \big) \mathrm{d} f_{j,t}^r + \sum_{j=1}^{n} \big[\operatorname{vec}\big(\boldsymbol{W}_t^{c\top} \boldsymbol{\mathcal{P}}_n^{(j)} \otimes \boldsymbol{A}^\top \big) \big]^\top \operatorname{vec}\big(\boldsymbol{Z}_t^{-1} \big) \mathrm{d} f_{j,t}^c ,
\end{aligned}
$$

where the second step follows from the identity $\operatorname{tr}(\boldsymbol{Q}_1 \boldsymbol{Q}_2) = \big[\operatorname{vec}(\boldsymbol{Q}_1^\top)\big]^\top \operatorname{vec}(\boldsymbol{Q}_2)$ for any compatible matrices $\boldsymbol{Q}_1, \boldsymbol{Q}_2$. Combining these results, we obtain that $\mathrm{d} \log p_Y\big(\boldsymbol{Y}_t \,\big|\, \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s \big)$ equals to

$$\sum_{j=1}^{m}\Big(\big[\boldsymbol{W}_t^r\boldsymbol{Y}_t\boldsymbol{B}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{W}}_t\big]_{j,j} - \big[\mathrm{vec}\big(\boldsymbol{B}^\top\otimes\boldsymbol{W}_t^{r\top}\boldsymbol{\mathcal{P}}_m^{(j)}\big)\big]^\top\mathrm{vec}\big(\boldsymbol{Z}_t^{-1}\big)\Big)\mathrm{d}f_{j,t}^r$$
$$+\sum_{j=1}^{n}\Big(\big[\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{W}}_t\boldsymbol{A}\boldsymbol{Y}_t\boldsymbol{W}_t^{c\top}\big]_{j,j} - \big[\mathrm{vec}\big(\boldsymbol{W}_t^{c\top}\boldsymbol{\mathcal{P}}_n^{(j)}\otimes\boldsymbol{A}^\top\big)\big]^\top\mathrm{vec}\big(\boldsymbol{Z}_t^{-1}\big)\Big)\mathrm{d}f_{j,t}^c.$$

Note that the coefficient associated with $\mathrm{d}f_{j,t}^r$ yields $\frac{\partial \log p_Y(\boldsymbol{Y}_t|\boldsymbol{\mathcal{X}}_t,\boldsymbol{f}_t,\boldsymbol{\theta}_s)}{\partial f_{j,t}^r}$, and similarly for $\mathrm{d}f_{j,t}^c$. These observations lead to (9). □

B Proofs of filter properties

Proof of Proposition 3.1. We treat the sequence $\{\widehat{\boldsymbol{f}}_{t+1}^{\star},\, i \in \mathbb{Z}^{+}\}$ as a random process that takes values in the Polish space $(E, d_E) = \big(\mathbb{F}, \|\cdot\|\big)$ (see the discussion above (N1)), and then apply Theorem 3.1 of Bougerol (1993) to obtain the proposition. First, note that under Assumption A1 it holds a.s. that

$$\big|\boldsymbol{Z}_t\big(\widehat{\boldsymbol{f}}_t^{\star}, \boldsymbol{\theta}_s\big)\big| \operatorname{vec}(\boldsymbol{Y}_t^{\star}) = \widetilde{\boldsymbol{Z}}_t^{\star} \operatorname{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^{\top} + \boldsymbol{E}_t\big), \tag{B.1}$$

where $\widetilde{\boldsymbol{Z}}_t^{\star} = \widetilde{\boldsymbol{Z}}_t\big(\widehat{\boldsymbol{f}}_t^{\star}, \boldsymbol{\theta}_s\big)$ with $\widetilde{\boldsymbol{Z}}_t\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big) = \big|\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big|\, \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)^{-1}$. Define $\boldsymbol{\mathcal{W}}_t^{\star} = \frac{\nu+m+n-1}{\nu-2}\big(\boldsymbol{\Sigma} + (\nu-2)^{-1}\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^{\top}\big)^{-1}$, and let $\boldsymbol{S}_{k\times k^2}$ be a selection matrix such that $\operatorname{diag}(\boldsymbol{Q}) = \boldsymbol{S}_{k\times k^2}\operatorname{vec}(\boldsymbol{Q})$ for any $k \times k$ matrix $\boldsymbol{Q}$, where $k \in \mathbb{Z}^{+}$. This, for instance, allows us to write

$$\begin{aligned}
\operatorname{diag}\big(\boldsymbol{W}_t^{r}\boldsymbol{Y}_t^{\star}\boldsymbol{B}^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^{\top}\boldsymbol{\mathcal{W}}_t^{\star}\big) &= \boldsymbol{S}_{m\times m^2}\operatorname{vec}\big(\boldsymbol{W}_t^{r}\boldsymbol{Y}_t^{\star}\boldsymbol{B}^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^{\top}\boldsymbol{\mathcal{W}}_t^{\star}\big)\\
&= \boldsymbol{S}_{m\times m^2}\Big(\big(\boldsymbol{\mathcal{W}}_t^{\star}\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\boldsymbol{B}\big) \otimes \boldsymbol{W}_t^{r}\Big)\operatorname{vec}(\boldsymbol{Y}_t^{\star}).
\end{aligned}$$

Then, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, using (B.1), the SRE can equivalently be reformulated as $\widehat{\boldsymbol{f}}_{t+1}^{\star} = \boldsymbol{\phi}_t^{\star}\big(\widehat{\boldsymbol{f}}_t^{\star}\big) = \boldsymbol{\psi}_t^{\star}\big(\widehat{\boldsymbol{f}}_t^{\star}, \boldsymbol{\theta}\big)$, where $\boldsymbol{\psi}_t^{\star}\big(\cdot, \boldsymbol{\theta}\big)$ is given in (11) with

$$\begin{aligned}
\boldsymbol{s}_t^{\star r}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big) &= \boldsymbol{S}_{m\times m^2}\Big(\big(\boldsymbol{\mathcal{W}}_t^{\star}\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\boldsymbol{B}\big) \otimes \boldsymbol{W}_t^{r}\Big)\\
&\qquad \times \widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\operatorname{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^{\top} + \boldsymbol{E}_t\big) - \boldsymbol{b}_t^{r}\big(\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s), \boldsymbol{B}\big), \qquad \text{(B.2)}\\
\boldsymbol{s}_t^{\star c}\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big) &= \boldsymbol{S}_{n\times n^2}\Big(\boldsymbol{W}_t^{c} \otimes \big(\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^{\top}\boldsymbol{\mathcal{W}}_t^{\star}\boldsymbol{A}\big)\Big)\\
&\qquad \times \widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\operatorname{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^{\top} + \boldsymbol{E}_t\big) - \boldsymbol{b}_t^{c}\big(\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s), \boldsymbol{A}\big). \qquad \text{(B.3)}
\end{aligned}$$

Since the joint process $\big\{(\boldsymbol{E}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^{r}, \boldsymbol{W}_t^{c}),\, t \in \mathbb{Z}\big\}$ is SE (Assumption A3), it follows that $\{\boldsymbol{\phi}_t^{\star}\}$ is also SE for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (see, e.g., White, 2001, Theorem 3.35). As in Lin and Lucas

(2025, Proposition 3), the following high-level (HL) conditions suffice to apply Theorem 3.1 in Bougerol (1993) or Straumann and Mikosch (2006, Theorem 2.8): For $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

HL1 $\mathbb{E}\Big(\log^+ \big\|\boldsymbol{\phi}_1^\star(\widehat{\boldsymbol{f}}_0^\star) - \widehat{\boldsymbol{f}}_0^\star\big\|\Big) < \infty$ for some $\widehat{\boldsymbol{f}}_0^\star \in \mathbb{F}$;

HL2 $\mathbb{E}\Big(\log^+ \rho(\boldsymbol{\phi}_1^\star)\Big) < \infty$;

HL3 $\mathbb{E} \log \rho\big(\boldsymbol{\phi}_t^{\star(r)}\big) < 0$ for some $r \geq 1$, where $\boldsymbol{\phi}_t^{\star(r)} = \boldsymbol{\phi}_t^\star \circ \boldsymbol{\phi}_{t-1}^\star \circ \cdots \circ \boldsymbol{\phi}_{t-r+1}^\star$ is the r-fold convolution of $\boldsymbol{\phi}_t^\star(\cdot)$.

Verification of Condition HL1: For completeness, the following inequalities are adopted from Lin and Lucas (2025, (B.3) - (B.4)): For any matrices $\boldsymbol{X}_i$ with compatible dimensions, where $i = 1, \ldots, K$ for $K \in \mathbb{Z}^+$, we have

$$\log^+ \left\| \prod_{i=1}^K \boldsymbol{X}_i \right\| \leq \log^+ \left(\prod_{i=1}^K \|\boldsymbol{X}_i\| \right) \leq \sum_{i=1}^K \log^+ \|\boldsymbol{X}_i\|, \tag{B.4}$$

$$\log^+ \left\| \sum_{i=1}^K \boldsymbol{X}_i \right\| \leq \log^+ \left(\sum_{i=1}^K \|\boldsymbol{X}_i\| \right) \leq \log(K) + \sum_{i=1}^K \log^+ \|\boldsymbol{X}_i\|. \tag{B.5}$$

Using (B.4) - (B.5), we obtain

$$\begin{aligned} \mathbb{E}\Big(\log^+ \big\|\boldsymbol{\phi}_1^\star(\widehat{\boldsymbol{f}}_0^\star) - \widehat{\boldsymbol{f}}_0^\star\big\|\Big) &\leq \log(3) + \log^+ \|\boldsymbol{\omega}\| \\ &+ \log^+ \big\|\big(\boldsymbol{\Phi} - \boldsymbol{I}_{m+n}\big)\widehat{\boldsymbol{f}}_0^\star\big\| + \log^+ \|\boldsymbol{K}\| + \mathbb{E}\Big(\log^+ \big\|\boldsymbol{s}_1^\star\big(\widehat{\boldsymbol{f}}_0^\star, \boldsymbol{\theta}_s\big)\big\|\Big), \end{aligned} \tag{B.6}$$

for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Since $\big\{\boldsymbol{s}_t^\star\big(\cdot, \boldsymbol{\theta}_s\big), \, t \in \mathbb{Z}\big\}$ is strictly stationary, it suffices to show that $\mathbb{E}\Big(\log^+ \big\|\boldsymbol{s}_t^\star\big(\widehat{\boldsymbol{f}}_0^\star, \boldsymbol{\theta}_s\big)\big\|\Big) < \infty$. By the norm equivalence in finite dimensional spaces, it suffices to consider each block in $\boldsymbol{s}_t^\star\big(\widehat{\boldsymbol{f}}_0^\star, \boldsymbol{\theta}_s\big)$ separately. It is well known $\|\boldsymbol{X}_1 \otimes \boldsymbol{X}_2\| = \|\boldsymbol{X}_1\| \, \|\boldsymbol{X}_2\|$ for any matrices $\boldsymbol{X}_1, \boldsymbol{X}_2$. Using (B.4) - (B.5), we have, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\mathbb{E}\Big(\log^+\|\boldsymbol{s}_t^{\star r}(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\|\Big)\le C+\mathbb{E}\Big(\log^+\big\|\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\big\|+\log^+\big\|\boldsymbol{W}_t^r\big\|+\log^+\big\|\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\big\|$$
$$+\log^+\big\|\mathrm{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top+\boldsymbol{E}_t\big)\big\|+\log^+\big\|\boldsymbol{b}_t^r\big(\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s),\boldsymbol{B}\big)\big\|\Big). \quad \text{(B.7)}$$

To proceed, we will repeatedly use the following property: for any random matrix $\boldsymbol{Q}$, if there exists an $p>0$ such that $\mathbb{E}\|\boldsymbol{Q}\|^p<\infty$, then $\mathbb{E}\big(\log^+\|\boldsymbol{Q}\|\big)<\infty$. First, the real matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are positive definite (Assumption A3(a)). It follows that $\mathbb{E}\big(\log^+\big\|\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\big\|\big)<\infty$ because $\sup_{t\in\mathbb{Z}}\big\|\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\big\|\le C\big\|\boldsymbol{\Sigma}^{-1/2}\big\|\Big(\sup_{t\in\mathbb{Z}}\big\|\boldsymbol{\mathcal{W}}_t^{\star 1/2}\boldsymbol{E}_t\boldsymbol{\Omega}^{-1/2}\big\|\Big)\big\|\boldsymbol{\Omega}^{-1/2}\big\|\le C$. Second, $\mathbb{E}\big(\log^+\big\|\boldsymbol{W}_t^r\big\|\big)+\mathbb{E}\big(\log^+\big\|\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\big\|\big)<\infty$ follows directly from Assumptions A1, A2, and A3(c). Third, given Assumption A3(b) and the fact that, in finite dimensions, $\mathbb{E}\big\|\mathrm{vec}(\boldsymbol{E}_t)\big\|^q<\infty$ for any $q\in(0,\nu)$, we obtain $\mathbb{E}\Big(\log^+\big\|\mathrm{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top+\boldsymbol{E}_t\big)\big\|\Big)<\infty$. Finally, it suffices to consider each element of $\boldsymbol{b}_t^r\big(\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s),\boldsymbol{B}\big)$. Note that $|\mathrm{tr}(\boldsymbol{Q})|\le mn\|\boldsymbol{Q}\|$ for any $\boldsymbol{Q}\in\mathbb{R}^{mn\times mn}$. For $j=1,\ldots,m$, we have

$$\mathbb{E}\bigg(\log^+\Big|\mathrm{tr}\Big(\big(\boldsymbol{B}\otimes\boldsymbol{\mathcal{P}}_m^{(j)}\boldsymbol{W}_t^r\big)\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\Big)\Big|\bigg)$$
$$\le\mathbb{E}\Big(\log^+(mn)+\log^+\|\boldsymbol{B}\|+\log^+\|\boldsymbol{W}_t^r\|+\log^+\big\|\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\big\|\Big)<\infty.$$

This implies $\mathbb{E}\Big(\log^+\big\|\boldsymbol{b}_t^r\big(\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s),\boldsymbol{B}\big)\big\|\Big)<\infty$. Then, $\mathbb{E}\Big(\log^+\|\boldsymbol{s}_t^{\star r}(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\|\Big)<\infty$ follows from (B.7). By similar arguments, one can show that the second block $\boldsymbol{s}_t^{\star c}(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)$ in $\boldsymbol{s}_t^\star(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)$ satisfies $\mathbb{E}\Big(\log^+\|\boldsymbol{s}_t^{\star c}(\widehat{\boldsymbol{f}}_0^\star,\boldsymbol{\theta}_s)\|\Big)<\infty$.

Verification of Condition HL2: By applying the mean value theorem for vector-valued functions in Rudin (1976, Theorem 9.19), we have

$$\rho(\boldsymbol{\phi}_1^\star)=\sup_{\|\boldsymbol{f}_1-\boldsymbol{f}_2\|>0}\frac{\|\boldsymbol{\phi}_1^\star(\boldsymbol{f}_1)-\boldsymbol{\phi}_1^\star(\boldsymbol{f}_2)\|}{\|\boldsymbol{f}_1-\boldsymbol{f}_2\|}\le\sup_{\boldsymbol{f}\in\mathbb{F}}\bigg\|\boldsymbol{\Phi}+\boldsymbol{K}\frac{\partial}{\partial\boldsymbol{f}^\top}\boldsymbol{s}_t^\star(\boldsymbol{f},\boldsymbol{\theta}_s)\bigg\|. \quad \text{(B.8)}$$

Condition HL2 is then fulfilled under Assumption SE1.

Verification of Condition HL3: Similarly, by the mean value theorem above, we have

$$\rho\big(\boldsymbol{\phi}_t^{\star(r)}\big) = \sup_{\|\boldsymbol{f}_1-\boldsymbol{f}_2\|>0} \frac{\big\|\boldsymbol{\phi}_t^{\star(r)}(\boldsymbol{f}_1) - \boldsymbol{\phi}_t^{\star(r)}(\boldsymbol{f}_2)\big\|}{\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|} \leq \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \frac{\partial}{\partial \boldsymbol{f}^\top} \boldsymbol{\phi}_t^{\star(r)}(\boldsymbol{f}) \right\|. \tag{B.9}$$

Condition HL3 is also satisfied under Assumption SE2.

Proposition 3.1 follows directly from Theorem 3.1 of Bougerol (1993), together with the fact that, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{f}_t^\star(\boldsymbol{\theta})$ admits an almost sure representation in terms of $\big\{\boldsymbol{\psi}_{t-k}^\star\big(\cdot, \boldsymbol{\theta}\big),\, k \geq 1\big\}$ and is therefore measurable with respect to the σ-field it generates; see also Straumann and Mikosch (2006, Theorem 2.8). □

Next, we establish the invertibility of the filter. To that end, additional notation is introduced. Let $\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big)$ denote the space of continuous $\mathbb{F}$-valued functions on $\boldsymbol{\Theta}$, equipped with the supremum norm $\|\cdot\|_{\boldsymbol{\Theta}}$, defined by $\|\boldsymbol{f}\|_{\boldsymbol{\Theta}} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{f}(\boldsymbol{\theta})\|$ for $\boldsymbol{f} \in \mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big)$. Given that $\dim(\boldsymbol{f}) < \infty$, $\big(\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big), \|\cdot\|_{\boldsymbol{\Theta}}\big)$ is a complete and separable (and thus Polish) space, provided that $\boldsymbol{\Theta}$ is compact (Kechris, 2012, Theorem 4.19).

Proof of Proposition 3.2. In contrast to Proposition 3.1, the result here requires convergence to be uniform in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. To invoke Bougerol (1993, Theorem 3.1), we treat the sequence $\big\{\widehat{\boldsymbol{f}}_t(\cdot),\, t \in \mathbb{Z}^+\big\}$, initialized at some $\widehat{\boldsymbol{f}}_0 \in \mathbb{F}$, as a sequence of random elements in the Polish space $(E, d_E) = \big(\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big), \|\cdot\|_{\boldsymbol{\Theta}}\big)$; see Eq. (N1). Moreover, one can write $\widehat{\boldsymbol{f}}_{t+1} = \boldsymbol{\phi}_t\big(\widehat{\boldsymbol{f}}_t\big)$, where the random maps $\boldsymbol{\phi}_t : \big(\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big), \|\cdot\|_{\boldsymbol{\Theta}}\big) \to \big(\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big), \|\cdot\|_{\boldsymbol{\Theta}}\big)$ are given by $\big[\boldsymbol{\phi}_t(\boldsymbol{f})\big](\boldsymbol{\theta}) = \boldsymbol{\psi}_t\big(\boldsymbol{f}(\boldsymbol{\theta}), \boldsymbol{\theta}\big)$ with $\boldsymbol{\psi}_t\big(\cdot, \boldsymbol{\theta}\big)$ provided in (7).

Note that $\big\{\boldsymbol{\phi}_t,\, t \in \mathbb{Z}\big\}$ is SE. This follows from the following arguments. If $\{\boldsymbol{U}_t,\, t \in \mathbb{Z}\}$ is an SE sequence and $\boldsymbol{V}_t = \boldsymbol{F}(\boldsymbol{U}_t, \boldsymbol{U}_{t-1}, \ldots)$ for some measurable function $\boldsymbol{F}$ that is independent of t, then $\{\boldsymbol{V}_t,\, t \in \mathbb{Z}\}$ is also SE; see, for example, White (2001, Theorem 3.35). It is then straightforward to verify that the joint process $\big\{(\boldsymbol{U}_t, \boldsymbol{V}_t),\, t \in \mathbb{Z}\big\}$ is SE as well. Since $\boldsymbol{f}_t^{\star}(\boldsymbol{\theta}_0)$ is measurable with respect to the σ-field generated by $\big\{(\boldsymbol{E}_s, \boldsymbol{X}_s, \boldsymbol{W}_s^r, \boldsymbol{W}_s^c),\, s \leq t-1\big\}$ (Proposition 3.1), the process $\big\{(\boldsymbol{f}_t^{\star}(\boldsymbol{\theta}_0), \boldsymbol{E}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c),\, t \in \mathbb{Z}\big\}$ is SE under Assumption A3. It then follows that the process $\big\{(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c),\, t \in \mathbb{Z}\big\}$ is jointly SE, which in turn implies that $\big\{\boldsymbol{\phi}_t,\, t \in \mathbb{Z}\big\}$ is SE.

As in Proposition 3.1, the following high-level conditions (abbreviated as $\widetilde{\mathrm{HL}}$) are sufficient to apply Theorem 3.1 of Bougerol (1993) and Theorem 2.8 of Straumann and Mikosch (2006):

$\widetilde{\mathrm{HL}}1$ $\mathbb{E}\Big(\log^+ \big\|\boldsymbol{\phi}_1(\widehat{\boldsymbol{f}}_0) - \widehat{\boldsymbol{f}}_0\big\|_{\boldsymbol{\Theta}}\Big) < \infty$, where $\widehat{\boldsymbol{f}}_0(\boldsymbol{\theta}) = \widehat{\boldsymbol{f}}_0 \in \mathbb{F}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$;

$\widetilde{\mathrm{HL}}2$ $\mathbb{E}\Big(\log^+ \rho(\boldsymbol{\phi}_1)\Big) < \infty$;

$\widetilde{\mathrm{HL}}3$ $\mathbb{E}\Big(\log \rho\big(\boldsymbol{\phi}_t^{(r)}\big)\Big) < 0$ for some integer $r \geq 1$, where $\boldsymbol{\phi}_t^{(r)} = \boldsymbol{\phi}_t \circ \boldsymbol{\phi}_{t-1} \circ \cdots \circ \boldsymbol{\phi}_{t-r+1}$ is the r-fold convolution of $\boldsymbol{\phi}_t$.

Verification of Condition $\widetilde{\mathrm{HL}}1$: By repeatedly applying (B.4)- (B.5), and using the strict stationarity of $\{\boldsymbol{\phi}_t\}$, we obtain

$$\begin{aligned}\mathbb{E}\Big(\log^+ \big\|\boldsymbol{\phi}_1(\widehat{\boldsymbol{f}}_0) - \widehat{\boldsymbol{f}}_0\big\|_{\boldsymbol{\Theta}}\Big) \leq \log(3) &+ \log^+ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{\omega}\| + \log^+ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big\|\big(\boldsymbol{\Phi} - \boldsymbol{I}_{m+n}\big)\widehat{\boldsymbol{f}}_0^{\star}\big\| \\ &+ \log^+ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{K}\| + \mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big\|\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s\big)\big\|\Big). \quad \text{(B.10)}\end{aligned}$$

Since $\boldsymbol{\Theta}$ is compact, it remains to verify that the last quantity is finite. By the norm equivalence in finite dimensional spaces, it suffices to consider each block in $\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s\big)$ separately. Note that $\|\mathrm{diag}(\boldsymbol{Q})\| \leq \sqrt{k}\,\|\boldsymbol{Q}\|$ for any $k \times k$ matrix $\boldsymbol{Q}$ with $k \in \mathbb{Z}^+$. For the upper block, we have

$$\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{s}_t^r\big(\boldsymbol{Y}_t, \widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s\big)\big\|\Big) \le C + \mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big|\det\big(\boldsymbol{Z}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s)\big)\Big| + \log^+ \big\|\boldsymbol{W}_t^r\big\| + \log^+ \big\|\boldsymbol{Y}_t\big\|$$
$$+ \log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s)^\top \widetilde{\boldsymbol{\mathcal{W}}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s)\Big\| + \log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{b}_t^r\big(\widetilde{\boldsymbol{Z}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s), \boldsymbol{B}\big)\big\|\Big), \quad \text{(B.11)}$$

where $\widetilde{\boldsymbol{\mathcal{W}}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s) = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{W}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s) = \frac{\nu+m+n-1}{\nu-2}\Big(\boldsymbol{\Sigma} + (\nu-2)^{-1}\boldsymbol{\mathcal{E}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s)^\top\Big)^{-1}$. If $\mathbb{E}\big(\log^+ \|\boldsymbol{Y}_t\|\big) < \infty$, then by arguments similar to those following (B.7), it follows that $\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{s}_t^r\big(\boldsymbol{Y}_t, \widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s\big)\big\|\Big) < \infty$. Similar to (B.1), $\mathrm{vec}(\boldsymbol{Y}_t)$ admits an almost sure representation: $\mathrm{vec}(\boldsymbol{Y}_t) = \big[\boldsymbol{Z}_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})\big]^{-1}\mathrm{vec}\big(\boldsymbol{C}_0\boldsymbol{X}_t\boldsymbol{D}_0^\top + \boldsymbol{E}_t\big)$, where $\boldsymbol{\theta}_{s0}$, $\boldsymbol{C}_0$, and $\boldsymbol{D}_0$ denote the true parameter values. Under Assumptions A2 and A3, we arrive at

$$\begin{aligned}\mathbb{E}\big(\log^+ \|\boldsymbol{Y}_t\|\big) &\le \mathbb{E}\big(\log^+ \|\mathrm{vec}(\boldsymbol{Y}_t)\|\big)\\ &\le \mathbb{E}\Big(\log^+ \big\|\big[\boldsymbol{Z}_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})\big]^{-1}\big\|\Big) + \mathbb{E}\Big(\log^+ \big\|\mathrm{vec}\big(\boldsymbol{C}_0\boldsymbol{X}_t\boldsymbol{D}_0^\top + \boldsymbol{E}_t\big)\big\|\Big) < \infty.\end{aligned} \quad \text{(B.12)}$$

Similarly, for the bottom block, we also have $\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{s}_t^c\big(\boldsymbol{Y}_t, \widehat{\boldsymbol{f}}_0, \boldsymbol{\theta}_s\big)\big\|\Big) < \infty$. It then follows from (B.10) that Condition $\widetilde{\text{HL}}$1 is satisfied.

Verification of Condition $\widetilde{\text{HL}}$2: As in Lin and Lucas (2025, Proof of Proposition 3), one has

$$\begin{aligned}\rho(\boldsymbol{\phi}_1) &= \sup_{\|\boldsymbol{f}_1-\boldsymbol{f}_2\|_{\boldsymbol{\Theta}}>0} \frac{\|\boldsymbol{\phi}_1(\boldsymbol{f}_1) - \boldsymbol{\phi}_1(\boldsymbol{f}_2)\|_{\boldsymbol{\Theta}}}{\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_{\boldsymbol{\Theta}}}\\ &= \sup_{\|\boldsymbol{f}_1-\boldsymbol{f}_2\|_{\boldsymbol{\Theta}}>0} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{\frac{\big\|\boldsymbol{\psi}_1\big(\boldsymbol{f}_1(\boldsymbol{\theta}), \boldsymbol{\theta}\big) - \boldsymbol{\psi}_1\big(\boldsymbol{f}_2(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\big\|}{\|\boldsymbol{f}_1(\boldsymbol{\theta}) - \boldsymbol{f}_2(\boldsymbol{\theta})\|} \frac{\|\boldsymbol{f}_1(\boldsymbol{\theta}) - \boldsymbol{f}_2(\boldsymbol{\theta})\|}{\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_{\boldsymbol{\Theta}}}\right\}\\ &\le \sup_{\|\boldsymbol{f}_1-\boldsymbol{f}_2\|_{\boldsymbol{\Theta}}>0} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{\frac{\big\|\boldsymbol{\psi}_1\big(\boldsymbol{f}_1(\boldsymbol{\theta}), \boldsymbol{\theta}\big) - \boldsymbol{\psi}_1\big(\boldsymbol{f}_2(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\big\|}{\|\boldsymbol{f}_1(\boldsymbol{\theta}) - \boldsymbol{f}_2(\boldsymbol{\theta})\|}\right\}\end{aligned}$$

$$\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\substack{\bar{\boldsymbol{f}}_1,\bar{\boldsymbol{f}}_2\in\mathbb{F},\\ \|\bar{\boldsymbol{f}}_1-\bar{\boldsymbol{f}}_2\|>0}} \left\{ \frac{\left\|\boldsymbol{\psi}_1\big(\bar{\boldsymbol{f}}_1,\boldsymbol{\theta}\big)-\boldsymbol{\psi}_1\big(\bar{\boldsymbol{f}}_2,\boldsymbol{\theta}\big)\right\|}{\|\bar{\boldsymbol{f}}_1-\bar{\boldsymbol{f}}_2\|} \right\}$$

$$\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \boldsymbol{\Phi} + \boldsymbol{K}\frac{\partial}{\partial \boldsymbol{f}^\top}\boldsymbol{s}_1\big(\boldsymbol{Y}_1,\boldsymbol{f},\boldsymbol{\theta}_s\big) \right\|, \tag{B.13}$$

where we apply the mean value theorem in Rudin (1976, Theorem 9.19) to the vector-valued function $\boldsymbol{\psi}_1\big(\,\cdot\,,\boldsymbol{\theta}\big)$ in the last step. Hence, $\mathbb{E}\big(\log^+\rho(\boldsymbol{\phi}_1)\big)<\infty$ by Assumption IV1.

Verification of Condition $\widetilde{\text{HL}}$3: In a similar manner to (B.13), we obtain

$$\mathbb{E}\Big(\log\rho\big(\boldsymbol{\phi}_t^{(r)}\big)\Big) \leq \mathbb{E}\left(\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\|\frac{\partial}{\partial \boldsymbol{f}^\top}\boldsymbol{\psi}_t^{(r)}(\boldsymbol{f},\boldsymbol{\theta})\right\|\right). \tag{B.14}$$

Condition $\widetilde{\text{HL}}$3 then immediately follows from Assumption IV2.

Proposition 3.2 then follows from an application of Bougerol (1993, Theorem 3.1) or Straumann and Mikosch (2006, Theorem 2.8). □

C Strong consistency

Proof of Lemma 1. We show that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$, as $T \to \infty$. Then, (13) follows directly from the triangle inequality.

Note that there exists some $\rho > 1$ such that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\big\| \le C\rho^{-t}$ a.s. for all $t \in \mathbb{Z}^+$ by Proposition 3.2. Applying a mean value theorem of $\ell_t\big(\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$ around $\ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$, we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| &\le T^{-1}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big|\ell_t\big(\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big) - \ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)\Big| \\
&\le T^{-1}\sum_{t=1}^{T} \sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \left\|\frac{\partial \ell_t\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{f}}\right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\Big\| \\
&\le CT^{-1}\sum_{t=1}^{T} \left\{\rho^{-t} \sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \left\|\frac{\partial \ell_t\big(\boldsymbol{f}, \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{f}}\right\|\right\}, \qquad \text{(C.1)}
\end{aligned}$$

where $\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_s)/\partial \boldsymbol{f} = \big|\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big|^{-1} \boldsymbol{s}_t\big(\boldsymbol{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big)$. Since the determinant function is continuous, it follows that for every $(\boldsymbol{\theta}, \boldsymbol{f}) \in \boldsymbol{\Theta} \times \mathbb{F}$, the quantity $\big\|\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_s)/\partial \boldsymbol{f}\big\|$ is a continuous function of the jointly SE process $\big\{(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c),\ t \in \mathbb{Z}\big\}$ (see p. S10). Moreover, for fixed $(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c)$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the mapping $\boldsymbol{f} \mapsto \big\|\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_s)/\partial \boldsymbol{f}\big\|$ is continuous over $\mathbb{F}$. Given the compactness of $\boldsymbol{\Theta}$ and the separability of $\mathbb{F}$, the supremum $\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big\|\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_s)/\partial \boldsymbol{f}\big\|$ is measurable with respect to $\big\{(\boldsymbol{Y}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c),\ t \in \mathbb{Z}\big\}$ and hence is SE by White (2001, Theorem 3.35). Furthermore, using Assumption SC2 and arguments similar to those for (B.11), we have $\mathbb{E}\Big(\log^+ \sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big\|\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_s)/\partial \boldsymbol{f}\big\|\Big)$ is bounded by

$$\mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \Big|\det\big(\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big)^{-1}\Big|\right) + \mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \Big\|\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big)\Big\|\right) < \infty. \qquad \text{(C.2)}$$

Hence, by Lemma 2.2 of Berkes et al. (2003), $\sum_{t=1}^{T}\Big\{\rho^{-t} \sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big\|\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_s)/\partial \boldsymbol{f}\big\|\Big\} <$

∞. As a result, we have $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0$ as $T \to \infty$.

It remains to show that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0$, which follows from the uniform law of large numbers in White (1996, Theorem A.2.2), applied to the SE sequence $\{\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s), t \in \mathbb{Z}\}$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, provided that $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\right|\Big) < \infty$. Note that $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log\det\left(\boldsymbol{Z}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\right)\right|\Big) < \infty$ given Assumption A1. Under Assumption A3(a), to prove $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\right|\Big) < \infty$, it suffices to establish that

$$\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log\det\left(\boldsymbol{\Sigma} + (\nu-2)^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)^\top\right)\right|\right) < \infty. \tag{C.3}$$

For convenience, define $\boldsymbol{\mathcal{S}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s) = \boldsymbol{\Sigma} + (\nu-2)^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)^\top$, and denote by $\lambda_i(\cdot)$ the ith eigenvalue of a matrix. Clearly, one has $\left|\log\det\left(\boldsymbol{\mathcal{S}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\right)\right| \leq \sum_{i=1}^m \left|\log\lambda_i\big(\boldsymbol{\mathcal{S}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\big)\right|$. By applying Weyl's inequality and Assumption A3(a), it is not hard to see

$$0 < \lambda_L \leq \lambda_i\big(\boldsymbol{\mathcal{S}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\big) \leq \lambda_U + \frac{1}{(\nu-2)\lambda_L}\left\|\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\right\|^2, \quad i = 1, \ldots, m. \tag{C.4}$$

Before continuing, note that for any $x \in [x_L, x_U] \subset (0, \infty)$, it follows that

$$|\log(x)| = \log^+(x) + \log^-(x) \leq \log^+(x_U) + \log^-(x_L), \tag{C.5}$$

where $\log^-(\cdot) = \max\{-\log(\cdot), 0\}$. Using (C.4), in conjunction with (C.5) and (B.4)–(B.5), we obtain

$$\begin{aligned}
&\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log\det\left(\boldsymbol{\Sigma} + (\nu-2)^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)^\top\right)\right|\right) \\
&\leq \sum_{i=1}^m \mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log\lambda_i\Big(\boldsymbol{\Sigma} + (\nu-2)^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)\boldsymbol{\Omega}^{-1}\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)^\top\Big)\right|\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq m\,\mathbb{E}\Bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\log^+\Big(\lambda_U+\frac{1}{(\nu-2)\lambda_L}\big\|\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\theta}_s)\big\|^2\Big)+\log^-(\lambda_L)\Bigg)\\
&\leq C+C\mathbb{E}\bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\log^+\big\|\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\theta}_s)\big\|\bigg). \qquad \text{(C.6)}
\end{aligned}
$$

Note that, given the compactness of $\boldsymbol{\Theta}$,

$$
\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\log^+\big\|\boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\theta}_s)\big\| \leq C+3\log^+\big\|\boldsymbol{Y}_t\big\|+\log^+\big\|\boldsymbol{X}_t\big\|\\
+\log^+\big\|\boldsymbol{W}_t^r\big\|+\log^+\big\|\boldsymbol{W}_t^c\big\|+2\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{f}_t(\boldsymbol{\theta})\big\|.
\end{aligned}
$$

By Eq. (C.6), the moment result of $\boldsymbol{Y}_t$ in (B.12) and Assumptions A3(b)–(c), establishing (C.3) reduces to verifying that $\mathbb{E}\big(\log^+\|\boldsymbol{f}_t\|_{\boldsymbol{\Theta}}\big)<\infty$. We establish that there exists some $p\in(0,1]$ such that $\mathbb{E}\big(\|\boldsymbol{f}_t\|_{\boldsymbol{\Theta}}^p\big)<\infty$. We iterate the SRE (7) backward $k\in\mathbb{Z}^+$ steps and apply the c_r inequality for $p\in(0,1]$, yielding

$$
\begin{aligned}
\mathbb{E}\Big(\|\boldsymbol{f}_{t+1}\|_{\boldsymbol{\Theta}}^p\Big) &= \mathbb{E}\Bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Bigg\|\boldsymbol{\Phi}^k\boldsymbol{f}_{t-k+1}(\boldsymbol{\theta})+\sum_{j=0}^{k-1}\boldsymbol{\Phi}^j\Big(\boldsymbol{\omega}+\boldsymbol{K}\,\boldsymbol{s}_{t-j}\big(\boldsymbol{Y}_{t-j},\boldsymbol{f}_{t-j}(\boldsymbol{\theta}),\boldsymbol{\theta}_s\big)\Big)\Bigg\|^p\Bigg)\\
&\leq \Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\Phi}\|^p\Big)^k\mathbb{E}\Big(\|\boldsymbol{f}_{t-k+1}\|_{\boldsymbol{\Theta}}^p\Big)+\sum_{j=0}^{k-1}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\Phi}\|^p\Big)^j\mathbb{E}\bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|\boldsymbol{\omega}+\boldsymbol{K}\,\boldsymbol{s}_{t-j}\big(\boldsymbol{Y}_{t-j},\boldsymbol{f}_{t-j}(\boldsymbol{\theta}),\boldsymbol{\theta}_s\big)\Big\|^p\bigg).
\end{aligned}
$$

If $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\omega}+\boldsymbol{K}\,\boldsymbol{s}_{t-j}\big(\boldsymbol{Y}_{t-j},\boldsymbol{f}_{t-j}(\boldsymbol{\theta}),\boldsymbol{\theta}_s\big)\big\|^p\Big)\leq C$, using Assumption SC1 and the SE property of $\big\{\boldsymbol{f}_t(\boldsymbol{\theta}),\,t\in\mathbb{Z}\big\}$ (Proposition 3.2), we immediately obtain

$$
\mathbb{E}\big(\|\boldsymbol{f}_t\|_{\boldsymbol{\Theta}}^p\big)\leq C\frac{1}{1-\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\Phi}\|^p}<\infty \qquad \text{(C.7)}
$$

by letting $k\to\infty$. By the compactness of $\boldsymbol{\Theta}$ and the c_r inequality for $p\in(0,1]$ again, we have $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\omega}+\boldsymbol{K}\,\boldsymbol{s}_t\big(\boldsymbol{Y}_t,\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\theta}_s\big)\big\|^p\Big)\leq C+C\mathbb{E}\Big(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}}\big\|\boldsymbol{s}_t\big(\boldsymbol{Y}_t,\boldsymbol{f},\boldsymbol{\theta}_s\big)\big\|^p\Big)$. Thus, we are left to prove that $\mathbb{E}\Big(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}}\big\|\boldsymbol{s}_t\big(\boldsymbol{Y}_t,\boldsymbol{f},\boldsymbol{\theta}_s\big)\big\|^p\Big)\leq C$.

Similar to the proof of Proposition 3.2, by the norm equivalence in finite dimensional spaces, it suffices to consider each block in $\boldsymbol{s}_t\big(\boldsymbol{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big)$ separately. Recall that $\widetilde{\boldsymbol{\mathcal{W}}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{W}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)$. It is simple to obtain that $\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big|\det\big(\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big)\big|^p \le C$ given Assumption A1. Since $\|\mathrm{diag}(\boldsymbol{Q})\| \le \|\boldsymbol{Q}\|_F \le \sqrt{k}\|\boldsymbol{Q}\|$ for any $k \times k$ matrix $\boldsymbol{Q}$, we arrive at

$$
\begin{aligned}
&\mathbb{E}\bigg(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big\|\boldsymbol{s}_t^r\big(\boldsymbol{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big)\big\|^p\bigg) \\
&\le C\mathbb{E}\bigg\{\bigg(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \Big|\det\big(\boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\big)\Big|^p\bigg)\big\|\boldsymbol{W}_t^r\big\|^p\big\|\boldsymbol{Y}_t\big\|^p\bigg\} + C\mathbb{E}\bigg(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \Big\|\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\Big\|^p\bigg) \\
&\le C\mathbb{E}\Big(\big\|\boldsymbol{W}_t^r\big\|^p\big\|\boldsymbol{Y}_t\big\|^p\Big) + C\mathbb{E}\bigg(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \Big\|\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)\Big\|^p\bigg) \\
&\le C\sqrt{\mathbb{E}\big\|\boldsymbol{W}_t^r\big\|^{2p}\mathbb{E}\big\|\boldsymbol{Y}_t\big\|^{2p}} + C,
\end{aligned}
\tag{C.8}
$$

where the last step follows from the Cauchy-Schwarz inequality and Assumptions A1 and SC2. Note that, by the Cauchy-Schwarz inequality again, we have

$$
\mathbb{E}\big\|\boldsymbol{Y}_t\big\|^{2p} \le \bigg(\mathbb{E}\Big\|\big[\boldsymbol{Z}_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})\big]^{-1}\Big\|^{4p}\bigg)^{1/2}\bigg(\mathbb{E}\Big\|\mathrm{vec}\big(\boldsymbol{C}_0\boldsymbol{X}_t\boldsymbol{D}_0^\top + \boldsymbol{E}_t\big)\Big\|^{4p}\bigg)^{1/2}. \tag{C.9}
$$

Take $p \le q_1/4 \in (0, 1/2]$ above, where q_1 is specified in Assumption SC2. By Assumption SC2, we have $\mathbb{E}\|\boldsymbol{Y}_t\|^{2p} \le C$. Therefore, by (C.8) and Assumption SC2, we conclude that $\mathbb{E}\Big(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big\|\boldsymbol{s}_t^r\big(\boldsymbol{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big)\big\|^p\Big) \le C$. Similarly, one can also show that $\mathbb{E}\Big(\sup_{(\boldsymbol{\theta},\boldsymbol{f})\in\boldsymbol{\Theta}\times\mathbb{F}} \big\|\boldsymbol{s}_t^c\big(\boldsymbol{Y}_t, \boldsymbol{f}, \boldsymbol{\theta}_s\big)\big\|^p\Big) \le C$.

Finally, the continuity of $\boldsymbol{\theta} \mapsto \mathcal{L}(\boldsymbol{\theta})$ follows from Theorem A.2.2 of White (1996). This completes the proof. □

Proof of Theorem 3.1. The proof follows the standard consistency arguments presented in Blasques et al. (2022, Theorem 4.6) and Lin and Lucas (2025, Theorem 1), which are grounded in classical results such as Theorem 3.4 of White (1996) and Theorem 3.3 of Gallant and White (1988). Recall that Lemma 1 establishes the uniform convergence of the empirical criterion function $\boldsymbol{\theta} \mapsto \widehat{\mathcal{L}}_T(\boldsymbol{\theta})$ to the continuous limiting function $\boldsymbol{\theta} \mapsto \mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\big(\ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)\big)$ over the compact parameter space $\boldsymbol{\Theta}$. It then remains to establish the the identifiable uniqueness of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

Given the continuity of $\boldsymbol{\theta} \mapsto \mathcal{L}(\boldsymbol{\theta})$ and the compactness of $\boldsymbol{\Theta}$, by Definition 3.3 of White (1996), it only requires to prove $\mathcal{L}(\boldsymbol{\theta}_0) > \mathcal{L}(\boldsymbol{\theta})$ for every $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Note that, by the law of total expectation and Gibbs' inequality for the Kullback-Leibler divergence (see, e.g., White, 1996, Theorem 2.3), it follows that $\mathcal{L}(\boldsymbol{\theta}_0) \geq \mathcal{L}(\boldsymbol{\theta})$, with equality if and only if $p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0}\big) = p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$ for almost every $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$ with respect to the pdf $p_Y\big(\cdot \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0}\big)$. By Assumption SC3, one has, $\forall t \in \mathbb{Z}$, $\boldsymbol{f}_t(\boldsymbol{\theta}) = \boldsymbol{f}_t(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}_s = \boldsymbol{\theta}_{s0}$. Given these conditions, together with the assumption that $\det(\boldsymbol{K}_0) \neq 0$, it is straightforward to obtain $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (i.e., uniqueness) by following the steps outlined in Lin and Lucas (2025, Proof of Theorem 1), and the details are therefore omitted. $\square$

D Asymptotic normality

Lemma D.1. *Under the assumptions of Proposition 3.2 and Assumptions AN1 - AN3, we have* $\sqrt{T}\,\frac{\partial \widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\mathcal{I}}_0\big)$ *as* $T \to \infty$*, where* $\boldsymbol{\mathcal{I}}_0 = \mathbb{E}\Big(\frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0})}{\partial \boldsymbol{\theta}^\top}\Big)$ *as defined in Theorem 3.2.*

Proof of Lemma D.1. We first argue that $\sqrt{T}\,\Big\|\frac{\partial \widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big\| \xrightarrow{a.s.} 0$ as $T \to \infty$. By Proposition 3.2, and using arguments similar to those in (C.1), the result follows under the assumption that $\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{f} \in \mathbb{F}} \Big\|\frac{\partial \ell_t(\boldsymbol{f}, \boldsymbol{\theta}_{s0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{f}^\top}\Big\|\Big) < \infty$.

Next, we apply the Cramér-Wold device to derive the asymptotic distribution $\sqrt{T}\,\frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$. As such, let $\boldsymbol{a} \in \mathbb{R}^{\dim(\boldsymbol{\theta}_0)}$ be any unit vector. The next step is to consider the limiting distribution of

$$\sqrt{T}\left(\boldsymbol{a}^\top \frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) = \sum_{t=1}^{T} X_{T,t}(\boldsymbol{\theta}_0), \qquad X_{T,t}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}}\, \boldsymbol{a}^\top \frac{\partial \ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \tag{D.1}$$

as $T \to \infty$. As in Lin and Lucas (2025, Lemma E.1), we apply the CLT for a martingale difference array, as stated in Davidson (1994, Theorem 24.3), by verifying the following conditions:

CLT1 $\big\{X_{T,t}(\boldsymbol{\theta}_0), \mathcal{F}_{T,t}\big\}$ is a martingale difference array, where $\mathcal{F}_{T,t} = \sigma\big((\boldsymbol{Y}_s, \boldsymbol{\mathcal{X}}_s),\ s \le t\big)$ for all $t \le T$;

CLT2 $\sum_{t=1}^{T} X_{T,t}^2(\boldsymbol{\theta}_0) \xrightarrow{p} \boldsymbol{a}^\top \boldsymbol{\mathcal{I}}_0 \boldsymbol{a}$;

CLT3 $\max_{1 \le t \le T} |X_{T,t}(\boldsymbol{\theta}_0)| \xrightarrow{p} 0$.

Consider CLT1 first. Let $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot \mid \mathcal{F}_{T,t})$. Note that both the multivariate gamma function and the determinant function are differentiable. It is then straightforward to see that $p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$ is differentiable with respect to $\boldsymbol{\theta}$ for every $\boldsymbol{Y}$. Under

Assumption AN2, one can interchange the integration and differentiation of $\boldsymbol{\theta} \mapsto p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; see, for example, basic textbooks such as Schilling (2017, Theorem 12.5) or Casella and Berger (2024, Theorem 2.4.3). It is worth noting that the integral of a scalar function of a matrix argument is interpreted in the usual way, as an iterated integral over each entry of the matrix; see Gupta and Nagar (2018, Chapter 1.4). We arrive at

$$\begin{aligned}
\mathbb{E}_{t-1}\left(\frac{\partial \ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) &= \int \frac{\partial \log p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_{s0}\big) \mathrm{d}\boldsymbol{Y} \\
&= \int \frac{\partial\, p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathrm{d}\boldsymbol{Y} \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} \int p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)\, \mathrm{d}\boldsymbol{Y} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= \mathbf{0},
\end{aligned}$$

where the last step follows from the fact that $\int p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)\, \mathrm{d}\boldsymbol{Y} = 1$.

It remains to verify CLT2 and CLT3. As discussed in Chapter 24.3 of Davidson (1994, p. 385), if the sequence $\Big\{ \boldsymbol{a}^\top \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},\, t \in \mathbb{Z} \Big\}$ is SE with finite variance, then Condition CLT2 follows from Theorem 13.12 of Davidson (1994). Moreover, Condition CLT3 is implied by Theorem 23.16 of Davidson (1994), which ensures the Lindeberg condition holds for the triangular array $\{X_{T,t}\}$; see also Billingsley (1999, Theorem 18.3). By the chain rule, it is not hard to see that $\frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is a measurable function of the tuple $\big(\boldsymbol{E}_t, \boldsymbol{X}_t, \boldsymbol{W}_t^r, \boldsymbol{W}_t^c, \boldsymbol{f}_t(\boldsymbol{\theta}_0), \boldsymbol{f}_t^{(1)}(\boldsymbol{\theta}_0)\big)$, where $\boldsymbol{f}_t^{(1)}(\boldsymbol{\theta}_0) = \partial \boldsymbol{f}_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}^\top$, and is thus measurable with respect to the σ-field generated by $\big\{(\boldsymbol{E}_s, \boldsymbol{X}_s, \boldsymbol{W}_s^r, \boldsymbol{W}_s^c),\, s \le t\big\}$. Under Assumption A3, and using arguments similar to those in the proof of Proposition 3.2, it follows that the sequence $\Big\{ \boldsymbol{a}^\top \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},\, t \in \mathbb{Z} \Big\}$ is SE. Finally, we have $\mathbb{E}\Big\| \boldsymbol{a}^\top \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \Big\|^2 \le$

$\mathbb{E}\left\|\frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|^2 < \infty$ by Assumption AN3. This completes the proof. $\square$

Proof of Theorem 3.2. The proof follows standard arguments used in establishing the asymptotic normality of M-estimators; see, for example, White (1996, Theorem 6.2) and Hayashi (2000, Proposition 7.8). First, Theorem 3.1 and Assumption AN1 imply that $\widehat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \mathrm{int}(\boldsymbol{\Theta}) \neq \emptyset$, as $T \to \infty$. Second, Lemma D.1 shows $\sqrt{T}\,\frac{\partial \widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\mathcal{I}}_0\big)$, as $T \to \infty$. Third, the construction of $\widehat{\mathcal{L}}_T$ and Assumption AN4 imply that $\widehat{\mathcal{L}}_T$ is a.s. twice continuous differentible on $\boldsymbol{\Theta}$, and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} - \boldsymbol{\mathcal{J}}(\boldsymbol{\theta})\right\| \xrightarrow{e.a.s.} 0$, as $T \to \infty$, where $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}) = \mathbb{E}\Big(\frac{\partial^2 \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\theta}_s)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Big)$. Finally, note that, given Assumption AN2, the integration and differentiation of $\partial\, p_Y\big(\boldsymbol{Y} \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \boldsymbol{\theta}_s\big)/\partial \boldsymbol{\theta}$ with respect to $\boldsymbol{\theta}$ can be interchanged at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ a.s. for all $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$. Using standard textbook arguments together with the law of iterated expectation, it is not hard to obtain that $\boldsymbol{\mathcal{I}}_0 = -\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_0)$. It then follows that Theorem 3.2 holds. $\square$

E Auxiliary theoretical results

Lemma E.1 (Jacobian matrices of scores in the DGP). *Recall* $\boldsymbol{s}_t^{\star r}\big(\,\cdot\,,\boldsymbol{\theta}_s\big)$ *and* $\boldsymbol{s}_t^{\star c}\big(\,\cdot\,,\boldsymbol{\theta}_s\big)$ *from* (B.2) *and* (B.3)*, respectively. We obtain*

$$\mathrm{D}_{\boldsymbol{f}}\boldsymbol{s}_t^{\star r}\big(\boldsymbol{f},\boldsymbol{\theta}_s\big)=\Big(\big[\mathrm{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top+\boldsymbol{E}_t\big)\big]^\top\otimes\big[\boldsymbol{S}_{m\times m^2}\big(\big(\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\boldsymbol{B}\big)\otimes\boldsymbol{W}_t^r\big)\big]-\boldsymbol{\mathcal{B}}_t^r(\boldsymbol{B})\Big)$$
$$\times\Big(\mathrm{vec}\big(\boldsymbol{D}_{t,1}^r(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\ \cdots\ \mathrm{vec}\big(\boldsymbol{D}_{t,m}^r(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\ \mathrm{vec}\big(\boldsymbol{D}_{t,1}^c(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\ \cdots\ \mathrm{vec}\big(\boldsymbol{D}_{t,n}^c(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\Big), \tag{E.1}$$

$$\mathrm{D}_{\boldsymbol{f}}\boldsymbol{s}_t^{\star c}\big(\boldsymbol{f},\boldsymbol{\theta}_s\big)=\Big(\big[\mathrm{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top+\boldsymbol{E}_t\big)\big]^\top\otimes\big[\boldsymbol{S}_{n\times n^2}\big(\boldsymbol{W}_t^c\otimes\big(\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^\top\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{A}\big)\big)\big]-\boldsymbol{\mathcal{B}}_t^c(\boldsymbol{A})\Big)$$
$$\times\Big(\mathrm{vec}\big(\boldsymbol{D}_{t,1}^r(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\ \cdots\ \mathrm{vec}\big(\boldsymbol{D}_{t,m}^r(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\ \mathrm{vec}\big(\boldsymbol{D}_{t,1}^c(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\ \cdots\ \mathrm{vec}\big(\boldsymbol{D}_{t,n}^c(\boldsymbol{f},\boldsymbol{\theta}_s)\big)\Big), \tag{E.2}$$

where $\boldsymbol{\mathcal{B}}_t^r(\boldsymbol{B})=\begin{pmatrix}\big[\mathrm{vec}(\boldsymbol{B}^\top\otimes\boldsymbol{W}_t^{r\top}\boldsymbol{\mathcal{P}}_m^{(1)})\big]^\top\\ \vdots\\ \big[\mathrm{vec}(\boldsymbol{B}^\top\otimes\boldsymbol{W}_t^{r\top}\boldsymbol{\mathcal{P}}_m^{(m)})\big]^\top\end{pmatrix}$, $\boldsymbol{\mathcal{B}}_t^c(\boldsymbol{A})=\begin{pmatrix}\big[\mathrm{vec}(\boldsymbol{W}_t^{c\top}\boldsymbol{\mathcal{P}}_n^{(1)}\otimes\boldsymbol{A}^\top)\big]^\top\\ \vdots\\ \big[\mathrm{vec}(\boldsymbol{W}_t^{c\top}\boldsymbol{\mathcal{P}}_n^{(n)}\otimes\boldsymbol{A}^\top)\big]^\top\end{pmatrix}$, *and*

$$\boldsymbol{D}_{t,j}^r(\boldsymbol{f},\boldsymbol{\theta}_s)=\Big[\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)\big(\boldsymbol{B}\otimes\boldsymbol{\mathcal{P}}_m^{(j)}\boldsymbol{W}_t^r\big)-\mathrm{tr}\Big(\big(\boldsymbol{B}\otimes\boldsymbol{\mathcal{P}}_m^{(j)}\boldsymbol{W}_t^r\big)\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)\Big)\Big]\boldsymbol{Z}_t(\boldsymbol{f},\boldsymbol{\theta}_s)^{-1},$$
$$\boldsymbol{D}_{t,j}^c(\boldsymbol{f},\boldsymbol{\theta}_s)=\Big[\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)\big(\boldsymbol{\mathcal{P}}_n^{(j)}\boldsymbol{W}_t^c\otimes\boldsymbol{A}\big)-\mathrm{tr}\Big(\big(\boldsymbol{\mathcal{P}}_n^{(j)}\boldsymbol{W}_t^c\otimes\boldsymbol{A}\big)\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)\Big)\Big]\boldsymbol{Z}_t(\boldsymbol{f},\boldsymbol{\theta}_s)^{-1}.$$

Proof of Lemma E.1. Let $\boldsymbol{f}=(\boldsymbol{f}^{r\top},\boldsymbol{f}^{c\top})^\top$, where $\boldsymbol{f}^r=(f_1^r,\ldots,f_m^r)^\top$ and $\boldsymbol{f}^c=(f_1^c,\ldots,f_n^c)^\top$. One can equivalently write

$$\boldsymbol{s}_t^{\star r}\big(\boldsymbol{f},\boldsymbol{\theta}_s\big)=\Big(\big[\mathrm{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top+\boldsymbol{E}_t\big)\big]^\top\otimes\big[\boldsymbol{S}_{m\times m^2}\big(\big(\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{E}_t\boldsymbol{\Omega}^{-1}\boldsymbol{B}\big)\otimes\boldsymbol{W}_t^r\big)\big]-\boldsymbol{\mathcal{B}}_t^r(\boldsymbol{B})\Big)\mathrm{vec}\Big(\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)\Big),$$
$$\boldsymbol{s}_t^{\star c}\big(\boldsymbol{f},\boldsymbol{\theta}_s\big)=\Big(\big[\mathrm{vec}\big(\boldsymbol{C}\boldsymbol{X}_t\boldsymbol{D}^\top+\boldsymbol{E}_t\big)\big]^\top\otimes\big[\boldsymbol{S}_{n\times n^2}\big(\boldsymbol{W}_t^c\otimes\big(\boldsymbol{\Omega}^{-1}\boldsymbol{E}_t^\top\boldsymbol{\mathcal{W}}_t^\star\boldsymbol{A}\big)\big)\big]-\boldsymbol{\mathcal{B}}_t^c(\boldsymbol{A})\Big)\mathrm{vec}\Big(\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)\Big).$$

To compute $\mathrm{D}_{\boldsymbol{f}}\boldsymbol{s}_t^\star\big(\boldsymbol{f},\boldsymbol{\theta}_s\big)$, the key is to obtain $\mathrm{D}_{\boldsymbol{f}}\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f},\boldsymbol{\theta}_s)$ by the chain rule. Note that

$$\mathrm{d}\boldsymbol{Z}_t(\boldsymbol{f},\boldsymbol{\theta}_s)=-\boldsymbol{B}\otimes\Big(\mathrm{d}\,\mathrm{diag}(\boldsymbol{f}^r)\boldsymbol{W}_t^r\Big)-\Big(\mathrm{d}\,\mathrm{diag}(\boldsymbol{f}^c)\boldsymbol{W}_t^c\Big)\otimes\boldsymbol{A}$$

$$= -\boldsymbol{B} \otimes \left(\sum_{j=1}^{m} \boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{W}_t^r \mathrm{d} f_j^r \right) - \left(\sum_{j=1}^{n} \boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{W}_t^c \mathrm{d} f_j^c \right) \otimes \boldsymbol{A}$$

$$= -\sum_{j=1}^{m} \left(\boldsymbol{B} \otimes \boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{W}_t^r \right) \mathrm{d} f_j^r - \sum_{j=1}^{n} \left(\boldsymbol{\mathcal{P}}_n^{(j)} \boldsymbol{W}_t^c \otimes \boldsymbol{A} \right) \mathrm{d} f_j^c.$$

By using the identities $\mathrm{d}|\boldsymbol{X}| = |\boldsymbol{X}| \operatorname{tr}\left(\boldsymbol{X}^{-1} \mathrm{d}\boldsymbol{X}\right)$ and $\mathrm{d}\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{-1}$, we obtain

$$\begin{aligned}
\mathrm{d}\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) &= \mathrm{d}\left(\left| \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right| \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)^{-1} \right) \\
&= \left(\mathrm{d} \left| \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right| \right) \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)^{-1} + \left| \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right| \left(\mathrm{d} \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)^{-1} \right) \\
&= \operatorname{tr}\left(\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \mathrm{d} \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)^{-1} - \widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \left(\mathrm{d} \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \boldsymbol{Z}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)^{-1} \\
&=: \sum_{j=1}^{m} \boldsymbol{D}_{t,j}^r(\boldsymbol{f}, \boldsymbol{\theta}_s) \, \mathrm{d} f_j^r + \sum_{j=1}^{n} \boldsymbol{D}_{t,j}^c(\boldsymbol{f}, \boldsymbol{\theta}_s) \, \mathrm{d} f_j^c.
\end{aligned}$$

This gives $\mathrm{dvec}\left(\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) = \left(\operatorname{vec}\left(\boldsymbol{D}_{t,1}^r(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \cdots \operatorname{vec}\left(\boldsymbol{D}_{t,m}^r(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \operatorname{vec}\left(\boldsymbol{D}_{t,1}^c(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \cdots \operatorname{vec}\left(\boldsymbol{D}_{t,n}^c(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \right) \mathrm{d}\boldsymbol{f}$, and as a result,

$$\mathrm{D}_{\boldsymbol{f}} \widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) = \left(\operatorname{vec}\left(\boldsymbol{D}_{t,1}^r(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \cdots \operatorname{vec}\left(\boldsymbol{D}_{t,m}^r(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \operatorname{vec}\left(\boldsymbol{D}_{t,1}^c(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \cdots \operatorname{vec}\left(\boldsymbol{D}_{t,n}^c(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \right). \tag{E.3}$$

Using (E.3) and the chain rule, i.e., $\mathrm{D}_{\boldsymbol{f}} \boldsymbol{s}_t^{\star r}(\boldsymbol{f}, \boldsymbol{\theta}_s) = \left(\mathrm{D}_{\widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s)} \boldsymbol{s}_t^{\star r}(\boldsymbol{f}, \boldsymbol{\theta}_s) \right) \left(\mathrm{D}_{\boldsymbol{f}} \widetilde{\boldsymbol{Z}}_t(\boldsymbol{f}, \boldsymbol{\theta}_s) \right)$, we arrive at (E.1); the expression (E.2) follows analogously. $\square$

F Accommodating conditional heteroskedasticity

The model can be further extended to account for the conditional heteroskedasticity of the disturbances. Consider $\log(\boldsymbol{\Sigma}_t) = \mathrm{diag}(\boldsymbol{f}_t^{\sigma})$ and $\log(\boldsymbol{\Omega}_t) = \mathrm{diag}\big((1, \boldsymbol{f}_t^{\omega\top})\big)$, where

$$\boldsymbol{f}_t^{\sigma} = \big(\log \boldsymbol{\Sigma}_{1,1,t}, \ldots, \log \boldsymbol{\Sigma}_{m,m,t}\big)^{\top}, \qquad \boldsymbol{f}_t^{\sigma} = \big(\log \boldsymbol{\Omega}_{2,2,t}, \ldots, \log \boldsymbol{\Omega}_{n,n,t}\big)^{\top}, \tag{F.1}$$

and we set the first diagonal element of $\boldsymbol{\Omega}_t$ to be 1 for identification as before. Without confusion, we now let $\boldsymbol{f}_t = \big(\boldsymbol{f}_t^{r\top}, \boldsymbol{f}_t^{c\top}, \boldsymbol{f}_t^{\sigma\top}, \boldsymbol{f}_t^{\omega\top}\big)^{\top}$ follow the updating scheme as in Eq. (7), with $\boldsymbol{s}_t(\cdot) = \big(\boldsymbol{s}_t^{r}(\cdot)^{\top}, \boldsymbol{s}_t^{c}(\cdot)^{\top}, \boldsymbol{s}_t^{\sigma}(\cdot)^{\top}, \boldsymbol{s}_t^{\omega}(\cdot)^{\top}\big)^{\top}$, where $\boldsymbol{s}_t^{r}\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big)$ and $\boldsymbol{s}_t^{c}\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big)$ are given in Proposition 2.1, with $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ replaced by $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Omega}_t$, respectively. The constructions of $\boldsymbol{s}_t^{\sigma}\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big)$ and $\boldsymbol{s}_t^{\omega}\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big)$ are provided below.

Proposition F.1. *Under the assumptions of Proposition 2.1, we have*

$$\begin{pmatrix} \boldsymbol{s}_t^{\sigma}\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) \\ \boldsymbol{s}_t^{\omega}\big(\boldsymbol{Y}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) \end{pmatrix} = \begin{pmatrix} \frac{\partial \log p_Y(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s)}{\partial \boldsymbol{f}_t^{\sigma}} \\ \frac{\partial \log p_Y(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s)}{\partial \boldsymbol{f}_t^{\omega}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\,\mathrm{diag}\Big(\boldsymbol{\mathcal{E}}_t \widetilde{\boldsymbol{\mathcal{E}}}_t^{\dagger} \boldsymbol{\mathcal{W}}_t^{\dagger}\Big) - \frac{n}{2}\,\boldsymbol{\iota}_m \\ \frac{1}{2}\,\mathrm{diag}\Big(\widetilde{\boldsymbol{\mathcal{E}}}_t^{\dagger} \boldsymbol{\mathcal{W}}_t^{\dagger} \boldsymbol{\mathcal{E}}_t\Big) - \frac{m}{2}\,\boldsymbol{\iota}_{n-1} \end{pmatrix}, \tag{F.2}$$

where, for $k \in \mathbb{Z}^{+}$, $\boldsymbol{\iota}_k$ denotes the k-dimensional vector of ones, $\widetilde{\boldsymbol{\mathcal{E}}}_t^{\dagger} = \boldsymbol{\Omega}_t^{-1} \boldsymbol{\mathcal{E}}_t^{\top} \boldsymbol{\Sigma}_t^{-1}$, and

$$\boldsymbol{\mathcal{W}}_t^{\dagger} = \boldsymbol{\mathcal{W}}_t^{\dagger}(\boldsymbol{f}_t, \boldsymbol{\theta}_s) = \frac{\nu + m + n - 1}{\nu - 2}\Big(\boldsymbol{I}_m + (\nu - 2)^{-1} \boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s) \boldsymbol{\Omega}_t^{-1} \boldsymbol{\mathcal{E}}_t(\boldsymbol{f}_t, \boldsymbol{\theta}_s)^{\top} \boldsymbol{\Sigma}_t^{-1}\Big)^{-1}.$$

All other notation is defined in Proposition 2.1.

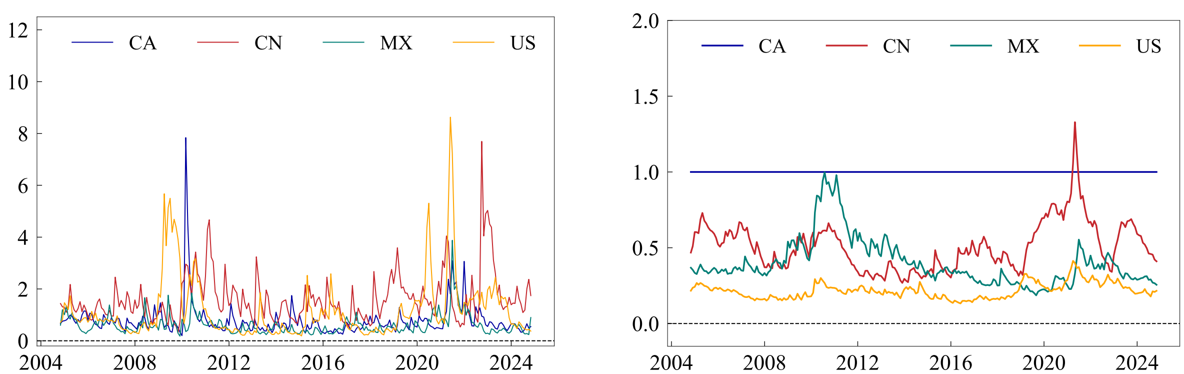
Proof of Proposition F.1. Similar to Eq. (A.2), we have

$$\begin{aligned} \mathrm{d} \log p_Y\big(\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{X}}_t, \boldsymbol{f}_t, \boldsymbol{\theta}_s\big) = {} & -\frac{1}{2}\,\mathrm{tr}\Big(\boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^{\dagger} \mathrm{d} \boldsymbol{\Xi}_t^{\dagger}\Big) - \mathrm{tr}\big(\boldsymbol{Z}_t^{-1} \mathrm{d} \boldsymbol{G}_t\big) + \frac{1}{2}\,\mathrm{tr}\Big(\boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^{\dagger} (\mathrm{d} \boldsymbol{\Sigma}_t) \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Xi}_t^{\dagger}\Big) \\ & - \frac{n}{2}\mathrm{tr}\big(\boldsymbol{\Sigma}_t^{-1} \mathrm{d} \boldsymbol{\Sigma}_t\big) - \frac{m}{2}\mathrm{tr}\big(\boldsymbol{\Omega}_t^{-1} \mathrm{d} \boldsymbol{\Omega}_t\big) =: P_{1,t}^{\dagger} - P_{2,t} + P_{3,t} - P_{4,t} - P_{5,t}, \end{aligned}$$

where $\boldsymbol{\Xi}_t^\dagger = \boldsymbol{\mathcal{E}}_t \boldsymbol{\Omega}_t^{-1} \boldsymbol{\mathcal{E}}_t^\top$. Write $\boldsymbol{f}_t^\sigma = (f_{1,t}^\sigma, f_{2,t}^\sigma, \ldots, f_{m,t}^\sigma)^\top$ and $\boldsymbol{f}_t^\omega = (f_{2,t}^\omega, f_{3,t}^\omega, \ldots, f_{n,t}^\omega)^\top$. By $\mathrm{d}\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{-1}$, and the identities $\boldsymbol{\Sigma}_t^{-1}\mathrm{d}\boldsymbol{\Sigma}_t = \sum_{j=1}^m \boldsymbol{\mathcal{P}}_m^{(j)} \mathrm{d}f_{j,t}^\sigma$ and $\boldsymbol{\Omega}_t^{-1}\mathrm{d}\boldsymbol{\Omega}_t = \sum_{j=2}^n \boldsymbol{\mathcal{P}}_n^{(j)} \mathrm{d}f_{j,t}^\omega$, using similar steps as in the proof of Proposition 2.1, we obtain

$$
\begin{aligned}
P_{1,t}^\dagger &= \sum_{j=1}^m \left[\boldsymbol{W}_t^r \boldsymbol{Y}_t \boldsymbol{B}^\top \boldsymbol{\Omega}_t^{-1} \boldsymbol{\mathcal{E}}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^\dagger\right]_{j,j} \mathrm{d}f_{j,t}^r + \sum_{j=1}^n \left[\boldsymbol{\Omega}_t^{-1} \boldsymbol{\mathcal{E}}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^\dagger \boldsymbol{A} \boldsymbol{Y}_t \boldsymbol{W}_t^{c\top}\right]_{j,j} \mathrm{d}f_{j,t}^c \\
&\quad + \frac{1}{2} \sum_{j=2}^n \left[\boldsymbol{\Omega}_t^{-1} \boldsymbol{\mathcal{E}}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^\dagger \boldsymbol{\mathcal{E}}_t\right]_{j,j} \mathrm{d}f_{j,t}^\omega, \\
P_{2,t} &= \sum_{j=1}^m \left[\mathrm{vec}\big(\boldsymbol{B}^\top \otimes \boldsymbol{W}_t^{r\top} \boldsymbol{\mathcal{P}}_m^{(j)}\big)\right]^\top \mathrm{vec}\big(\boldsymbol{Z}_t^{-1}\big) \mathrm{d}f_{j,t}^r \\
&\quad + \sum_{j=1}^n \left[\mathrm{vec}\big(\boldsymbol{W}_t^{c\top} \boldsymbol{\mathcal{P}}_n^{(j)} \otimes \boldsymbol{A}^\top\big)\right]^\top \mathrm{vec}\big(\boldsymbol{Z}_t^{-1}\big) \mathrm{d}f_{j,t}^c, \\
P_{3,t} &= \frac{1}{2} \sum_{j=1}^m \mathrm{tr}\Big(\boldsymbol{\mathcal{P}}_m^{(j)} \boldsymbol{\Xi}_t^\dagger \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^\dagger\Big) \mathrm{d}f_{j,t}^\sigma = \frac{1}{2} \sum_{j=1}^m \Big[\boldsymbol{\Xi}_t^\dagger \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mathcal{W}}_t^\dagger\Big]_{j,j} \mathrm{d}f_{j,t}^\sigma,
\end{aligned}
$$

and $P_{4,t} = -\dfrac{n}{2} \sum_{j=1}^m \mathrm{d}f_{j,t}^\sigma$, $P_{5,t} = -\dfrac{m}{2} \sum_{j=2}^n \mathrm{d}f_{j,t}^\omega$. By combining the preceding results, the proposition follows. □



(a) Dynamic import log volatility (b) Dynamic export log volatility

Figure F5: Filtered log volatility for the import and export trade flow. The left and right panels are for the import and export, respectively. The sample ranges from January 2000 to October 2024.

Figure F5 presents the time-varying log-volatilities for the import-export application from Section 5.1, more specifically the import ($\log(\boldsymbol{\Sigma}_t)$) and export ($\log(\boldsymbol{\Omega}_t)$) log volatility levels. The results indicate that, overall, the log-volatilities of imports and exports across countries remain relatively stable, with some exceptions during the aftermath of the financial crisis (2009–2010) and after the COVID-19 crises.

G Additional finite sample results

G.1 Additional simulation outputs

Table G.5: Empirical MSE of time-varying spatial spillover effects.

Dynamics	T	Static-SMAR-t	TVP-SMAR-t	Static-SMAR-$\mathcal{N}$	TVP-SMAR-$\mathcal{N}$
Const + SlowS	250	0.139	**0.046**	0.153	0.056
	500	0.136	**0.038**	0.145	0.044
	1000	0.125	**0.033**	0.134	0.039
Const + FastS	250	0.131	**0.047**	0.142	0.053
	500	0.127	**0.043**	0.136	0.047
	1000	0.124	**0.039**	0.134	0.043
Const + Break	250	0.205	**0.040**	0.213	0.048
	500	0.201	**0.032**	0.209	0.040
	1000	0.201	**0.026**	0.208	0.035
Const + AR	250	0.107	**0.033**	0.120	0.037
	500	0.106	**0.027**	0.114	0.028
	1000	0.102	**0.023**	0.110	0.024
SlowS + FastS	250	0.156	**0.069**	0.172	0.077
	500	0.154	**0.063**	0.167	0.071
	1000	0.154	**0.062**	0.167	0.068
SlowS + Break	250	0.229	**0.068**	0.243	0.075
	500	0.230	**0.048**	0.245	0.060
	1000	0.234	**0.051**	0.250	0.063
SlowS + AR	250	0.134	**0.057**	0.152	0.062
	500	0.132	**0.047**	0.141	0.051
	1000	0.131	**0.044**	0.140	0.049
FastS + Break	250	0.236	**0.069**	0.253	0.081
	500	0.234	**0.063**	0.251	0.074
	1000	0.235	**0.059**	0.251	0.068
FastS + AR	250	0.136	**0.059**	0.150	0.065
	500	0.133	**0.055**	0.144	0.058
	1000	0.130	**0.051**	0.139	0.054
Break + AR	250	0.127	**0.042**	0.139	0.051
	500	0.127	**0.033**	0.138	0.041
	1000	0.124	**0.026**	0.133	0.033

G.2 Additional details of Section 5.2

This section supplements Section 5.2 by providing additional details on the out-of-sample estimation procedure, portfolio selection methodology, specifications of the benchmark models used for comparison, and the computation of evaluation metrics for model performance.

First, for the out-of-sample estimation procedure. We split the data into a training set consisting of the first $T_{\text{train}} = 537$ observations and a test set comprising the last $T_{test} = 300$ observations. At each time point t_0, we fit the one-side TVP-MSAR model based on the latest T_{train} observations $\{\boldsymbol{Y}_t\}_{t=t_0-T_{\text{train}}}^{t_0-1}$, then use the fitted model to predict the conditional covariance matrix of $\text{vec}(\boldsymbol{Y}_{t_0})$ by $\hat{\boldsymbol{H}}_{t_0}$, where $\hat{\boldsymbol{H}}_{t_0}$ is computed by:

$$\hat{\boldsymbol{H}}_{t_0} = \left(\boldsymbol{I}_{mn} - \hat{\boldsymbol{G}}_{t_o}\right)^{-1} \left(\hat{\boldsymbol{\Omega}}_{t_0} \otimes \hat{\boldsymbol{\Sigma}}_{t_0}\right) \left(\left(\boldsymbol{I}_{mn} - \hat{\boldsymbol{G}}_{t_o}\right)^{-1}\right)^{\top}, \tag{G.1}$$

where $\hat{\boldsymbol{G}}_{t_0}(\hat{\boldsymbol{f}}_{t_0}, \hat{\boldsymbol{W}}_{t_0}^r)$, $\hat{\boldsymbol{\Omega}}_{t_0}(\hat{\boldsymbol{f}}_{t_0})$, and $\hat{\boldsymbol{\Sigma}}_{t_0}(\hat{\boldsymbol{f}}_{t_0})$, and $\hat{\boldsymbol{f}}_{t_0}$ are obtained by the score-driven update given the estimated model parameters and the training observations. For the spatial connection matrix $\hat{\boldsymbol{W}}_{t_0}^r$, we simply let $\hat{\boldsymbol{W}}_{t_0}^r = \boldsymbol{W}_{t_0-1}^r$.

Using $\hat{\boldsymbol{H}}_{t_0}$, we select the unconstrained minimum variance portfolio (MVP) from 16 sector indexes by choosing the weight vector:

$$\hat{\boldsymbol{w}}_{t_0}^u \coloneqq \underset{\boldsymbol{w}_{t_0}'\mathbf{1}_{mn}=1}{\text{argmin}}\ \boldsymbol{w}_{t_0}^{\top}\hat{\boldsymbol{H}}_{t_0}\boldsymbol{w}_{t_0} = \frac{\hat{\boldsymbol{H}}_{t_0}^{-1}\mathbf{1}_{mn}}{\mathbf{1}_{mn}^{\top}\hat{\boldsymbol{H}}_{t_0}^{-1}\mathbf{1}_{mn}}, \tag{G.2}$$

for $m = n = 4$. If the short sales are not allowed, one can also select the constrained MVP by choosing the weight vector:

$$\hat{\boldsymbol{w}}_{t_0}^u \coloneqq \underset{\boldsymbol{w}_{t_0}'\mathbf{1}_{mn}=1,\ \boldsymbol{w}_{t_0}\geq\mathbf{0}}{\text{argmin}}\ \boldsymbol{w}_{t_0}^{\top}\hat{\boldsymbol{H}}_{t_0}\boldsymbol{w}_{t_0}, \tag{G.3}$$

where (G.3) has no closed form solution and needs to be computed by numerical optimization methods.

For the benchmark matrix factor GARCH (MF-GARCH) model, we implement the R code provided by Yu et al. (2024) and select the factor dimensions $r_1 = r_2 = 2$ for all fitted models. The choice of factor dimensions is based on the eigenvalue ratio method (Yu et al., 2022), as recommended in Yu et al. (2024).

Finally, we construct out-of-sample averaged return (AV), standard deviation (SD), and information ratio (IR) to compare the out-of-sample portfolio selection performance of different models, where

$$\mathrm{AV} = \frac{1}{T_{test}} \sum_{t_0} \hat{\boldsymbol{w}}_{t_0}^\top \boldsymbol{R}_{t_0}, \qquad \mathrm{SD} = \sqrt{\frac{1}{T_{test} - 1} \sum_{t_0} \left(\hat{\boldsymbol{w}}_{t_0}^\top \boldsymbol{R}_{t_0} - \mathrm{AV} \right)^2},$$

and IR = AV/SD. We also perform model confidence set (MCS) tests for the squared returns according to Hansen et al. (2011).

References in Online Appendix

Berkes, I., L. Horváth, and P. Kokoszka (2003). GARCH processes: Structure and estimation. *Bernoulli*, 201–227.

Billingsley, P. (1999). *Convergence of Probability Measures.* John Wiley & Sons.

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

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

Casella, G. and R. Berger (2024). *Statistical Inference* (2nd ed.). CRC press.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

Gallant, A. R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Oxford: Basil Blackwell.

Gupta, A. K. and D. K. Nagar (2018). *Matrix Variate Distributions.* Chapman and Hall/CRC.

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

Hayashi, F. (2000). *Econometrics.* Princeton, N.J.: Princeton University Press.

Kechris, A. (2012). *Classical Descriptive Set Theory*, Volume 156. New York: Springer.

Lin, Y. and A. Lucas (2025). Functional location-scale models with robust observation-driven dynamics. Technical Report TI 2025-027/III, Tinbergen Institute Discussion Paper.

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

Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). New York: McGraw-Hill.

Schilling, R. L. (2017). *Measures, Integrals and Martingales* (2nd ed.). Cambridge: Cambridge University Press.

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

White, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians* (Rev. ed.). Bingley: Emerald.

Yu, C., D. Li, F. Jiang, and K. Zhu (2024). Matrix GARCH model: Inference and application. *Journal of the American Statistical Association*, 1–31.

Yu, L., Y. He, X. Kong, and X. Zhang (2022). Projected estimation for large-dimensional matrix factor models. *Journal of Econometrics 229*(1), 201–217.