TI 2025-027/III
Tinbergen Institute Discussion Paper

# Functional Location-Scale Models with Robust Observation-Driven Dynamics

*Yicong Lin*[1]
*Andre Lucas*[2]

1 Vrije Universiteit Amsterdam, Tinbergen Institute

2 Vrije Universiteit Amsterdam, Tinbergen Institute

# Functional Location-Scale Models with Robust Observation-Driven Dynamics

Yicong Lin[1] and Andre Lucas*[1]

[1]Vrije Universiteit Amsterdam and Tinbergen Institute

April, 2025

**Abstract**

We introduce a new class of location-scale models for dynamic functional data in arbitrary but fixed dimensions, where the location and scale functional parameters can evolve over time. A key feature of the parameter dynamics in these models is its observation-driven nature, where the one-step-ahead evolution is fully determined conditional on past observations, yet remains stochastic unconditionally. We estimate the model using a likelihood-based approach designed for sparsely observed data and establish the consistency and asymptotic normality of the underlying static parameters that govern the location-scale dynamics. The choice of objective function and the construction of the dynamics together shield the time-varying location and scale parameters from the potentially distorting effects of influential observations. Simulations reveal that our method can recover the unobserved location-scale dynamics from sparse data, even in the presence of model mis-specification and substantial outliers. We apply our framework to examine the intraday volatility dynamics of Pfizer stock returns during the COVID-19 pandemic, and $PM_{2.5}$ concentrations measured by low-cost sensors across Europe. The proposed model exhibits robust performance in capturing dynamics for both datasets despite the presence of many large shocks.

**Keywords:** time variation, location-scale, functional score-driven dynamics, sparse data, outlier robustness.

*Corresponding author: Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. E-mail address: a.lucas@vu.nl.

# 1 Introduction

Functional data analysis deals with observations that are (in principle) infinite-dimensional, such as curves and surfaces defined over a continuum; see, among many others, Hörmann and Kokoszka (2010), Hörmann et al. (2015). With longer recorded datasets being increasingly available for such objects, modeling the time-series dynamics of functional data has become a central topic in the field. As the number of time series observations grows, questions naturally arise regarding the stability versus the possible time-variation of the (functional) parameters that characterize the observed functional data, such as their location and scale properties.

The challenge of time-varying parameters has been extensively studied in conventional time series analysis of finite-dimensional objects. Popular approaches include nonparametric kernel smoothing techniques (Cai, 2007), parameter-driven (or state-space) models (Durbin and Koopman, 2012), and observation-driven models (Creal et al., 2013; Harvey, 2013). In contrast, time series models for functional observations are much less developed. Often, the functional parameters are assumed to be constant over time. An important exception is the class of functional GARCH (fGARCH) models (e.g., Hörmann et al., 2013; Aue et al., 2017; Cerovecki et al., 2019). The class of fGARCH models, however, is particular to volatility curves in one dimension and leaves untouched how to generalize the method to higher dimensions and other (or more) time-varying parameters. This lack of attention to the dynamics of functional parameters is the more surprising given the many reports of either structural or gradual changes in such parameters across a range of applications (Horváth et al., 2014; Bardsley et al., 2017; Boniece et al., 2023; Li et al., 2024; Bastian and Dette, 2025).

In this paper we therefore introduce a new class of functional time series models where both the functional location and scale parameters can vary over time simultaneously. Specifically, we consider functional objects $Y_i(\boldsymbol{t})$, with $i = 1, \ldots, T$ denoting the time-series index, and $\boldsymbol{t} \in \mathbb{R}^d$ the functional index for a fixed but arbitrary dimension $d \geq 1$. Building on fundamental results for stochastic recurrence equations, we provide a full study of the new models' asymptotic properties, including stationarity and ergodicity and filter invertibility for the functional location and scale parameter processes. We establish theoretically that the time-varying location and scale dynamics can be recovered from sparsely observed data and provide a strong consistency and asymptotic normality result for the static parameters that govern these dynamics.

The proposed framework enjoys two key features. First, our approach is observation-driven in the terminology of Cox (1981): the dynamics of the mean and scale parameters are stochastic unconditionally, but fully determined one-step-ahead conditional on past observations. Observation-driven dynamics are also used in a functional context in, for example, Hörmann et al. (2013), Aue et al. (2017), and Cerovecki et al. (2019) to capture functional conditional heteroskedasticity (fGARCH) for $d = 1$. We explain the intrinsic connection between our new approach and the special case of the fGARCH, but also show how our new framework is much more general and allows for joint functional location and scale dynamics in arbitrary dimensions $d$. A distinct advantage of the observation-driven approach is that it facilitates a simple estimation methodology, especially when extending the models to incorporate nonlinearities and heavy-tailed features. Observation-driven models can also easily be applied for forecasting due to their predictive structure (see, e.g., Aue et al., 2015; Shang and Hyndman, 2017; Paparoditis and Shang, 2023).

The dynamics of the new model are inspired by the extensive literature on score-driven models, which are a sub-class of observation-driven models; see Creal et al. (2013) and Harvey (2013) for a general introduction, and Harvey and Luati (2014) and Gasperoni et al. (2023) for univariate and multivariate score-driven location and scale models for finite-dimensional objects.[1] Score-driven updates are optimal in a Kullback-Leibler sense for finite-dimensional observations (Blasques et al., 2015) and provide consistent estimates of the time-varying parameter paths even if the model is severely mis-specified (Beutner et al., 2023). This paper is the first to extend this approach to functional data.

[1] One can also consult the overview of more than 300 papers on score-driven dynamics on https://www.gasmodel.com.

As a second key feature, our framework is set in the context of sparsely observed data, subject to possibly considerable noise: the number of observations per functional object at each time point in our context is finite, and we explicitly address the common encounter in sparse designs of incidental outlying observations. Some of the challenges when working with sparse designs have been discussed in Yao et al. (2005), Wang et al. (2016), Zhang and Wang (2016), and Zhu and Wang (2023). In most practical settings, functional data are only recorded discretely and are often contaminated by measurement errors. Many earlier studies, however, typically assume dense and regularly-spaced observations, allowing one to pre-smooth the data to reconstruct the underlying functions; see Ramsay and Silverman (2005) for a textbook treatment. Pre-smoothing helps to mitigate the effect of outliers arising from phenomena such as market crashes in economics and finance, or weather anomalies and sensor malfunctioning in environmental studies. Pre-smoothing, however, is generally not feasible under sparse designs. Accounting for potential incidental outliers is therefore a crucial step in our framework.

To obtain outlier-robust time-varying dynamics, we incorporate heavy-tailed features by using Student's $t$ processes for the measurement errors (Heyde and Leonenko, 2005; Shah et al., 2014). This differs from the common practice in the literature, where Gaussianity is often assumed, either explicitly or implicitly. When combined with the score-driven dynamics of Creal et al. (2013) and Harvey (2013) mentioned above, outlying functional observations are now automatically downweighted in the functional parameter updates, thus ensuring their robustness. Similar robustness features have been observed in finite-dimensional settings (Harvey and Luati, 2014; Gasperoni et al., 2023), and this paper extends them to the functional time series context. Our finite sample results clearly show that the new approach is indeed robust to incidental shocks, in contrast to competing methods, which turn out to be more sensitive and numerically less stable.

The score-driven updates proposed in this paper take an intuitive form and follow by projecting the function-valued parameters onto a finite set of basis functions with time-varying coefficients. The (robust) functional parameter updates follow naturally from the fat-tailed Student's $t$ process based objective function used for estimation. More specifically, our parameter updates depend on a projection of the most recent functional observation onto the above set of basis functions across all $d$ dimensions. The resulting step then adjusts the basis function loading coefficients in a steepest ascent direction given the criterion function. Also the functional GARCH model of Cerovecki et al. (2019) and the functional factor models of Hays et al. (2012) make use of projections on a finite basis. We differ substantially from either set-up, however. In particular, our model considers integrated location *and* scale dynamics, includes non-Gaussian features, introduces a *robust* updating mechanism for both location and scale, uses non-linear score-driven dynamics rather than the parameter-driven dynamics of Hays et al. (2012) or the more linear observation-driven dynamics of Cerovecki et al. (2019), and finally, provides a solution to the challenges faced when studying the asymptotic properties of the non-linear dynamics in the new model.

The new model performs well in both simulated and empirical settings. In a controlled setting, the model recovers complex time variation in both means and variances in both $d = 1$ and $d = 2$ dimensional settings. This holds despite the time-series dynamics being severely mis-specified and the data being plagued by fat-tailed measurement errors. The new model also performs well in two empirical applications. In the first application, we analyze intraday volatility patterns ($d = 1$). Here the new model outperforms the existing benchmark of Cerovecki et al. (2019). Part of this outperformance can be attributed to the robust dynamics, especially during periods of market stress. In the second example we study a $d = 2$ dimensional setting. Here, we show that the model is able to capture the temporal developments in spatial $PM_{2.5}$ levels across Europe, using sparse and rather noisy 'citizen science' sensor measurements. We conclude that the new model offers a valuable addition to the functional time series modeling toolkit and can be further extended in many interesting directions.

The rest of this paper is set up as follows. In Section 2 we introduce the modeling framework, including the use of score-driven dynamics for functional data and the estimation of the model's static parameters. Section 3 formulates the conditions for stationarity and ergodicity, filter invertibility, consistency, and asymptotic normality. Section 4 shows the model's performance in a variety of controlled settings. Section 5 illustrates the model in two empirical contexts. Section 6 concludes. Proofs and additional simulation results are provided in the Online Appendix.

We adopt the following notational conventions. Vectors and matrices are in bold, whereas scalars are non-bold. For a vector $\boldsymbol{x} = (x_j) \in \mathbb{R}^n$, its $p$-norm is denoted by $\|\boldsymbol{x}\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$. The matrix norm $\|\boldsymbol{A}\|_{p,q}$ of a matrix $\boldsymbol{A}$ is defined as $\|\boldsymbol{A}\|_{p,q} = \sup_{\boldsymbol{x}\neq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_q / \|\boldsymbol{x}\|_p$. When $p = q$, we denote it as $\|\cdot\|_p$. The subscripts are omitted whenever $p = q = 2$. A block-diagonal matrix with, for instance, two blocks $\boldsymbol{A}$ and $\boldsymbol{B}$ is written as $\operatorname{diag}(\boldsymbol{A}, \boldsymbol{B})$. All random elements are assumed to be defined on some common probability space. A vector-valued sequence $\{\boldsymbol{x}_i,\ i \in \mathbb{Z}\}$ is said to converge to zero exponentially fast almost surely (e.a.s.) as $i \to \infty$, denoted by $\boldsymbol{x}_i \xrightarrow{e.a.s.} \boldsymbol{0}$, if there exists some $\gamma > 1$ such that $\gamma^i \|\boldsymbol{x}_i\| \xrightarrow{a.s.} 0$, where $\xrightarrow{a.s.}$ denotes almost sure (a.s.) convergence.

# 2 Observation-driven dynamics for functional time-series

We begin by introducing our location-scale model with time-varying functional parameters in Section 2.1, followed by a motivating example in Section 2.2 that leads to our formulation of updating schemes in Section 2.3. Section 2.4 then presents an estimation procedure based on this updating formulation.

## 2.1 General location-scale set-up

For a univariate random element $Y_i(\boldsymbol{t})$, consider the functional location-scale model

$$Y_i(\boldsymbol{t}) = \mu_i(\boldsymbol{t}) + \sigma_i(\boldsymbol{t})\varepsilon_i(\boldsymbol{t}), \qquad i \in [\![ T ]\!], \qquad \boldsymbol{t} \in \boldsymbol{\mathcal{T}}, \tag{2.1}$$

where $[\![ T ]\!]$ denotes $\{1, \ldots, T\}$ with $T \in \mathbb{Z}^+$. In our current context, the quantity $T$ typically denotes the length of the time-series. Model (2.1) has been previously used by, for instance, Hyndman and Shahid Ullah (2007) and Hyndman and Booth (2008). We assume that both the location $\mu_i(\boldsymbol{t})$ and scale $\sigma_i(\boldsymbol{t})$ can be defined using link functions of an underlying variable $\boldsymbol{f}_i(\boldsymbol{t}) \in \mathbb{R}^{n_f}$ for $n_f \geq 1$, where the measurable link functions are $g_\mu(\cdot)$ and $g_\sigma(\cdot)$ are such that $\mu_i(\boldsymbol{t}) = g_\mu(\boldsymbol{f}_i(\boldsymbol{t}))$ and $\sigma_i(\boldsymbol{t}) = g_\sigma(\boldsymbol{f}_i(\boldsymbol{t})) > 0$, for every $\boldsymbol{t} \in \boldsymbol{\mathcal{T}}$. We focus on the case of a compact $\boldsymbol{\mathcal{T}} \subset \mathbb{R}^d$. This allows us to not only model univariate functional data ($d = 1$), but also time-varying surfaces ($d = 2$) and time-varying three-dimensional shapes ($d = 3$).

We assume the error terms $\varepsilon_i(\cdot)$ are zero-mean Student processes in the terminology of Shah et al. (2014), independent across $i$. This allows for functional data that are characterized by incidental outliers and influential observations and endows the model with desirable robustness features later on when we describe the dynamics of $\mu_i(\cdot)$ and $\sigma_i(\cdot)$ (for similar properties in the context of non-functional data, see, for instance, Harvey and Luati, 2014; Gasperoni et al., 2023). Note that this automatically embeds the case of a Gaussian process innovation $\varepsilon_i(\cdot)$.

## 2.2 A motivating example: functional GARCH

To motivate our approach for modeling the time-variation in location and scale, we first briefly review the functional GARCH(1,1) framework proposed by Cerovecki et al. (2019). A number of aspects of the new modeling approach can be anchored to this simpler setting, which helps the intuition for the new approach.

Consider a setting with $t \in \mathcal{T} = [0, 1]$, i.e., $d = 1$, and let $i \in [\![ T ]\!]$ with $Y_i(t) = \sigma_i(t)\varepsilon_i(t)$. We assume the volatility process $\sigma_i(\cdot)$ satisfies the functional GARCH specification

$$\sigma_{i+1}^2(t) = \omega(t) + \int \alpha(t,s) Y_i^2(s)\, \mathrm{d}s + \int \beta(t,s)\sigma_i^2(s)\, \mathrm{d}s, \tag{2.2}$$

with $\{\varepsilon_i(\cdot)\}$ a sequence of i.i.d. random elements, $\omega(t) > 0$, and $\alpha(t,s), \beta(t,s) \geq 0$ for all $t, s \in \mathcal{T}$ (Cerovecki et al., 2019, Section 3.1). Assume there exists a finite set of linearly independent basis functions in $L^2(\mathcal{T})$ (the Hilbert space of all real, square-integrable functions on

$\mathcal{T}$), $\{\phi_1(\cdot),\dots,\phi_K(\cdot)\}$, such that

$$\omega(t)=\sum_{k=1}^{K} w_k\,\phi_k(t),\qquad \alpha(t,s)=\sum_{k_1,k_2=1}^{K} a_{k_1,k_2}\phi_{k_1}(t)\phi_{k_2}(s),\qquad \beta(t,s)=\sum_{k_1,k_2=1}^{K} b_{k_1,k_2}\phi_{k_1}(t)\phi_{k_2}(s),$$

where $w_k$, $a_{k_1,k_2}$, and $b_{k_1,k_2}$ for $k,k_1,k_2\in[\![K]\!]$ are unknown static parameters that need to be estimated. Define $\boldsymbol{\phi}_K(\cdot)=\big(\phi_1(\cdot),\dots,\phi_K(\cdot)\big)^\top$. We then write $\omega(t)=\boldsymbol{\phi}_K(t)^\top\boldsymbol{\omega}$, $\alpha(t,s)=\boldsymbol{\phi}_K(t)^\top\boldsymbol{A}\,\boldsymbol{\phi}_K(s)$, $\beta(t,s)=\boldsymbol{\phi}_K(t)^\top\boldsymbol{B}\,\boldsymbol{\phi}_K(s)$, where $\boldsymbol{\omega}=\big(\omega_1,\dots,\omega_K\big)^\top$, $\boldsymbol{A}=\big(a_{k_1,k_2},\,k_1,k_2\in[\![K]\!]\big)\in\mathbb{R}^{K\times K}$, and $\boldsymbol{B}=\big(b_{k_1,k_2},\,k_1,k_2\in[\![K]\!]\big)\in\mathbb{R}^{K\times K}$. Using these definitions, we rewrite (2.2) as

$$\sigma_{i+1}^2(t)=\boldsymbol{\phi}_K(t)^\top\boldsymbol{\omega}+\boldsymbol{\phi}_K(t)^\top\boldsymbol{A}\int\boldsymbol{\phi}_K(s)Y_i^2(s)\,\mathrm{d}s+\boldsymbol{\phi}_K(t)^\top\boldsymbol{B}\int\boldsymbol{\phi}_K(s)\sigma_i^2(s)\,\mathrm{d}s. \tag{2.3}$$

Multiplying (2.3) on both sides by $\boldsymbol{\phi}_K(t)$ and integrating over $\mathcal{T}$ w.r.t. $t$, we obtain

$$\begin{aligned}\int\boldsymbol{\phi}_K(t)\sigma_{i+1}^2(t)\,\mathrm{d}t&=\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t\,\boldsymbol{\omega}\\&+\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t\,\boldsymbol{A}\int\boldsymbol{\phi}_K(s)Y_i^2(s)\,\mathrm{d}s+\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t\,\boldsymbol{B}\int\boldsymbol{\phi}_K(s)\sigma_i^2(s)\,\mathrm{d}s,\end{aligned} \tag{2.4}$$

where $\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t\in\mathbb{R}^{K\times K}$ is invertible. Define $\boldsymbol{\gamma}_i=\Big(\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t\Big)^{-1}\int\boldsymbol{\phi}_K(t)\sigma_i^2(t)\,\mathrm{d}t$. Multiplying both sides of (2.4) by the inverse of $\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t$, we obtain

$$\boldsymbol{\gamma}_{i+1}=\boldsymbol{\omega}+\bar{\boldsymbol{B}}\boldsymbol{\gamma}_i+\boldsymbol{A}\int\boldsymbol{\phi}_K(s)Y_i^2(s)\,\mathrm{d}s,\qquad \bar{\boldsymbol{B}}=\boldsymbol{B}\int\boldsymbol{\phi}_K(t)\boldsymbol{\phi}_K(t)^\top\,\mathrm{d}t. \tag{2.5}$$

Note that $\sigma_i^2(\cdot)$ and $\boldsymbol{\gamma}_i$ are related by $\sigma_i^2(\cdot)=\boldsymbol{\phi}_K(\cdot)^\top\boldsymbol{\gamma}_i$, which is a function-on-function projection, with $\boldsymbol{\gamma}_i\in\mathbb{R}^{n_\gamma}$ being the projection coefficient for $n_\gamma=K$. The functional GARCH model (2.2) can thus be re-cast into specification (2.5), where its $n_\gamma=K$ projections in $\boldsymbol{\gamma}_i$ allow for the estimation of the static parameters $\boldsymbol{\omega}$, $\boldsymbol{A}$, and $\bar{\boldsymbol{B}}$ that define the functional parameters $\omega(\cdot)$, $\alpha(\cdot,\cdot)$, and $\beta(\cdot,\cdot)$ of the original functional GARCH dynamics in (2.2). If in practical settings a finite small $K<\infty$

provides a good approximation to the original volatility function and the parameters describing its dynamics, this provides a considerable dimension reduction. It is this intuition that we exploit as well for our more general specification of observation-driven functional location-scale dynamics in general dimensions.

### 2.3 Functional dynamics for the general setting

For an arbitrary dimension $d \in \mathbb{Z}^+$ with $\boldsymbol{t} = (t_1, \ldots, t_d)^\top \in \mathbb{R}^d$, we follow the same intuition for the time-varying parameter $\boldsymbol{f}_i(\cdot) \in \mathbb{R}^{n_f}$ in the general location-scale model (2.1) as we have seen for the functional GARCH model of Cerovecki et al. (2019) in Section 2.2 for $d = 1$, and which we also encounter in for instance Li et al. (2021) and Zhang and Li (2022) for $d = 2$ and Berild and Fuglstad (2023) for $d = 3$. In particular, for the $j_{\text{th}}$ element of $\boldsymbol{f}_i(\cdot)$, we project it onto a *finite* set of product basis functions $\{\phi_1(\cdot), \ldots, \phi_{K_j}(\cdot)\}$, with $K_j \in \mathbb{N}$ for $j = 1, \ldots, d$:

$$\boldsymbol{f}_i(\boldsymbol{t}) = \sum_{k_1=1}^{K_1} \cdots \sum_{k_d=1}^{K_d} \bar{\boldsymbol{\gamma}}_{i,k_1,\ldots,k_d} \times \phi_{k_1}(t_1) \times \phi_{k_d}(t_d) = \sum_{\mathbf{1} \leq \boldsymbol{k} \leq \boldsymbol{K}} \bar{\boldsymbol{\gamma}}_{i,\boldsymbol{k}}\, \phi_{\boldsymbol{k}}(\boldsymbol{t}) = \boldsymbol{\Gamma}_i\, \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}), \tag{2.6}$$

where we use the short-hand multi-index notation $\phi_{\boldsymbol{k}}(\boldsymbol{t}) = \prod_{i=1}^{d} \phi_{k_i}(t_i) \in \mathbb{R}$ for vectors of indices $\boldsymbol{k} = (k_1, \ldots, k_d)^\top, \boldsymbol{K} = (K_1, \ldots, K_d)^\top \in \mathbb{N}^d$, as well as the multi-index notation $\bar{\boldsymbol{\gamma}}_{i,\boldsymbol{k}} = \bar{\boldsymbol{\gamma}}_{i,k_1,\ldots,k_d} \in \mathbb{R}^{n_f}$, and the multi-index summation $\sum_{\mathbf{1} \leq \boldsymbol{k} \leq \boldsymbol{K}} = \sum_{k_1=1}^{K_1} \cdots \sum_{k_d=1}^{K_d}$ with $\mathbf{1}$ a vector of ones of appropriate dimension. Here, "$\leq$" denotes coordinatewise ordering in the space $\mathbb{N}^d$. Specifically, for $\boldsymbol{m} = (m_1, \ldots, m_d)$ and $\boldsymbol{n} = (n_1, \ldots, n_d)$, we have $\boldsymbol{m} \leq \boldsymbol{n}$ if and only if $m_i \leq n_i$ for all $1 \leq i \leq d$. Moreover, the matrix $\boldsymbol{\Gamma}_i = (\bar{\boldsymbol{\gamma}}_{i,\mathbf{1}}, \ldots, \bar{\boldsymbol{\gamma}}_{i,\boldsymbol{K}}) \in \mathbb{R}^{n_f \times (K_1 \cdot K_2 \cdots K_d)}$ gathers the basis coefficients, while the vector $\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}) = \big(\phi_{\mathbf{1}}(\boldsymbol{t}), \ldots, \phi_{\boldsymbol{K}}(\boldsymbol{t})\big)^\top \in \mathbb{R}^{(K_1 \cdot K_2 \cdots K_d) \times 1}$ all the possible cross-products of basis functions related to the $d$ different dimensions of $\boldsymbol{t}$.

Multiplying (2.6) from the right by $\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top$, and integrating over $\boldsymbol{t}$, we obtain

$$\int \boldsymbol{f}_i(\boldsymbol{t})\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top \mathrm{d}\boldsymbol{t} = \boldsymbol{\Gamma}_i \int \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top \mathrm{d}\boldsymbol{t} \qquad \Longleftrightarrow$$

$$\boldsymbol{\gamma}_i = \mathrm{vec}\left(\boldsymbol{\Gamma}_i^\top\right) = \mathrm{vec}\left(\left(\int \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top \mathrm{d}\boldsymbol{t}\right)^{-1} \int \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\boldsymbol{f}_i(\boldsymbol{t})^\top \mathrm{d}\boldsymbol{t}\right), \tag{2.7}$$

where $\boldsymbol{\gamma}_i \in \mathbb{R}^{n_\gamma}$ for $n_\gamma = n_f \cdot \prod_{i=1}^d K_i$. This specification of $\boldsymbol{\gamma}_i$ is completely analogous to the specification of $\boldsymbol{\gamma}_i$ in the functional GARCH settings above (2.5), but now for general dimensions $d$ and for a general location-scale model. Note that we immediately obtain the dynamics over $i$ for the location and scale from the identities

$$\mu_i(\boldsymbol{t}) = g_\mu\big(\boldsymbol{\Gamma}_i\, \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big), \qquad \sigma_i(\boldsymbol{t}) = g_\sigma\big(\boldsymbol{\Gamma}_i\, \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big), \tag{2.8}$$

for link functions $g_\mu(\,\cdot\,)$ and $g_\sigma(\,\cdot\,)$.

The current set-up covers a large number of different models; see for instance Peng and Paul (2009), Aue et al. (2017), Cerovecki et al. (2019), Berild and Fuglstad (2023). If $n_f = 2$ and $\boldsymbol{f} = (f_1, f_2)^\top \in \mathbb{R}^2$ only contains a single location-related and a single scale-related element, examples of typical choices for the link functions include the identity function $g_\mu(\boldsymbol{f}) = f_1$ for the mean and an exponential link function $g_\sigma(\boldsymbol{f}) = \exp(f_2)$ for the standard deviation or variance. The latter ensures positivity of the scale by construction for all indices $i \in \mathbb{Z}$ and $\boldsymbol{t} \in \boldsymbol{\mathcal{T}}$. Other choices are of course also possible and covered by the same specification in (2.7)–(2.8). Such choices can even include a structural time-series decomposition of the location and scale dynamics in, for example, level, trend, seasonal and cyclical components; see Harvey (1990) for examples in the finite dimensional, non-functional data setting.

To construct an updating equation for $\boldsymbol{\gamma}_i$, we assume

$$\boldsymbol{\gamma}_{i+1} = \boldsymbol{\omega} + \boldsymbol{A}\ \boldsymbol{s}\left(Y_i, \boldsymbol{\gamma}_i\right) + \boldsymbol{B}\ \boldsymbol{\gamma}_i, \tag{2.9}$$

where $\boldsymbol{s}(\cdot,\cdot)$ is a function(al) of the previous observation $Y_i(\cdot)$ and the previous value of $\boldsymbol{\gamma}_i$, respectively. We define $\boldsymbol{s}(\cdot,\cdot)$ in more detail in the next section. Note that if $d = 1$ and $\boldsymbol{s}(Y_i, \boldsymbol{\gamma}_i) = \int \boldsymbol{\phi}_K(s) Y_i^2(s)\,\mathrm{d}s$, we recover the functional GARCH model from Eq. (2.5). The updating framework in (2.9) is, however, much more general and makes the specification observation-driven in the terminology of Cox (1981). As a result, the model is easily estimated using standard methods, as we show in the next subsection. Note that (2.9) also embeds a functional generalization of the score-driven framework of Creal et al. (2013) and Harvey (2013). In the finite dimensional, non-functional data setting, such steps in the time-varying parameter result in expected (local) improvements in the Kulback-Leibler information; see Blasques et al. (2015), Creal et al. (2024), and De Punder et al. (2024). In addition, such steps result in a consistent estimate of the time-varying parameter paths, even if the model is misspecified; see Beutner et al. (2023). Such score-driven steps are defined with respect to the estimation criterion for the static parameters of the model: $\boldsymbol{\omega}$, $\boldsymbol{A}$ and $\boldsymbol{B}$. Therefore, we first discuss how to formulate an appropriate estimation criterion for the model's static parameters in order to explicitly define the update function $\boldsymbol{s}(\cdot,\cdot)$.

## 2.4 Parameter estimation and score-driven dynamics

In this section, we discuss how to estimate the static parameters of the model, including the parameters $\boldsymbol{\omega}$, $\boldsymbol{A}$, and $\boldsymbol{B}$ that govern the dynamics of $\boldsymbol{\gamma}_i$ in (2.9). This also gives rise to our specific choice of $\boldsymbol{s}(\cdot,\cdot)$ in (2.9) that drives the functional dynamics.

Assume that for each $i \in [\![\, T \,]\!]$ we observe $Y_i(\cdot)$ over a finite grid $\mathbb{T} = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N\}$; see Yao et al. (2005) and Zhang and Li (2022). For simplicity, we assume the number $N$ of observation points is the same for all $i$; these can be allowed vary across dimensions at the expense of additional notational complexity. We particularly focus on settings with sparsely observed data ($N < \infty$); see also Paul and Peng (2009) and Zhang and Wang (2016). In such settings with sparsely recorded data, we do not perform any pre-smoothing steps to remove measurement errors, but instead explicitly introduce potentially heavy-tailed measurement errors into our modeling framework. By allowing the errors to have fat tails, we can accommodate situations where they can be incidentally large, a phenomenon common in real-world applications but often overlooked in the existing literature. The location and scale functional dynamics we introduce later on will be robust to such challenges.

In case of no confusion, we omit the explicit dependence on $N$ in our notation. We gather the observed data in $\{Y_i(\boldsymbol{t}),\, \boldsymbol{t} \in \mathbb{T}\}$. Given our assumed Student $t$ process for $\varepsilon_i(\cdot)$ in (2.1), we can write

$$\begin{aligned}
&\boldsymbol{Y}_i = \boldsymbol{\mu}(\boldsymbol{\gamma}_i) + \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\,\boldsymbol{\varepsilon}_i, \qquad \boldsymbol{Y}_i = \big(Y_i(\boldsymbol{t}_1), \dots, Y_i(\boldsymbol{t}_N)\big)^\top, \qquad \boldsymbol{\varepsilon}_i = \big(\varepsilon_i(\boldsymbol{t}_1), \dots, \varepsilon_i(\boldsymbol{t}_N)\big)^\top, \\
&\boldsymbol{\mu}(\boldsymbol{\gamma}_i) = \Big(g_\mu\big(\boldsymbol{\Gamma}_i\,\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_1)\big), \dots, g_\mu\big(\boldsymbol{\Gamma}_i\,\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_N)\big)\Big)^\top, \qquad (2.10)\\
&\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i) = \operatorname{diag}\Big(g_\sigma\big(\boldsymbol{\Gamma}_i\,\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_1)\big), \dots, g_\sigma\big(\boldsymbol{\Gamma}_i\,\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_N)\big)\Big),
\end{aligned}$$

with $\boldsymbol{\varepsilon}_i \overset{\text{i.i.d.}}{\sim} t_{\nu_1}\big(\boldsymbol{0}, \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)\big)$ for $\nu_1 > 0$, and where $\boldsymbol{\Lambda}(\boldsymbol{\nu}_2) = \big(C_{\boldsymbol{\nu}_2}(\boldsymbol{t}_k, \boldsymbol{t}_\ell),\, k, \ell \in [\![\, N \,]\!]\big)$ is a square matrix of order $N$ that possibly depends on a vector of static parameters $\boldsymbol{\nu}_2 = (\nu_2 \dots, \nu_{n_\nu})^\top \in \mathbb{R}^{n_\nu - 1}$ for some $n_\nu \geq 1$. Here, $C_{\boldsymbol{\nu}_2}(\boldsymbol{s}, \boldsymbol{t}) = \operatorname{cov}(\varepsilon_i(\boldsymbol{s}), \varepsilon_i(\boldsymbol{t}))$ is some distance measure that describes the covariance structure of the Student $t$ process. Without loss of generality, we assume $C_{\boldsymbol{\nu}_2}(\boldsymbol{t}, \boldsymbol{t}) = 1$. This immediately implies $\mathbb{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top) = \nu_1(\nu_1 - 2)^{-1}\,\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)$ for $\nu_1 > 2$.

Various examples of covariance kernels for $d = 1$ are available in Rasmussen and Williams (2006, Chapter 4). For instance, $C_{\nu_2}(s, t) = \exp\big(-\nu_2^{-1}\,|s - t|\big)$, $\nu_2 > 0$, leads to a Student $t$ Ornstein-Uhlenbeck process (Heyde and Leonenko, 2005, Theorem 3.2). For $d \geq 2$, the following

Matérn covariance is commonly used in the spatial modeling literature, namely

$$C_{\boldsymbol{\nu}_2}(\boldsymbol{s},\boldsymbol{t}) = \frac{\nu_2^2}{\Gamma(\nu_3)2^{\nu_3-1}}\left(\sqrt{2\nu_3}\,\frac{\|\boldsymbol{s}-\boldsymbol{t}\|}{\nu_4}\right)^{\nu_3} K_{\nu_3}\left(\sqrt{2\nu_3}\,\frac{\|\boldsymbol{s}-\boldsymbol{t}\|}{\nu_4}\right), \qquad \boldsymbol{\nu}_2 = \big(\nu_2,\nu_3,\nu_4\big)^\top, \tag{2.11}$$

where $\nu_2^2$ denotes the spatial variance parameter, $\nu_3$ the spatial smoothness parameter, and $\nu_4$ the spatial range parameter, and where $K_{\nu_3}$ is the modified Bessel function of the second kind. Further examples can be found in, e.g., Cressie (1993, Section 2.5).

Given the observation-driven nature of the dynamics for $\boldsymbol{\gamma}_i$ as explicated in Eq. (2.9), the conditional joint density function $p(\boldsymbol{Y}_i \mid \boldsymbol{\mathcal{F}}_{i-1};\boldsymbol{\nu})$ of $\boldsymbol{Y}_i$ given the past filtration $\boldsymbol{\mathcal{F}}_{i-1} = \{\boldsymbol{Y}_j\}_{j=-\infty}^{i-1}$ can directly be written as $p(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i;\boldsymbol{\nu})$, with

$$\begin{aligned} p\big(\boldsymbol{Y}_i \,\big|\, \boldsymbol{\gamma}_i;\boldsymbol{\nu}\big) &= \frac{\Gamma\big((\nu_1+N)/2\big)}{\Gamma\big(\nu_1/2\big)(\nu_1\pi)^{N/2}}\,\big|\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big|^{-1}\big|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)\big|^{-1/2} \\ &\quad\times\Big(1+\nu_1^{-1}\big(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\gamma}_i)\big)^\top\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1}\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1}\big(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\gamma}_i)\big)\Big)^{-(\nu_1+N)/2}, \end{aligned} \tag{2.12}$$

where $\boldsymbol{\nu} = (\nu_1,\boldsymbol{\nu}_2^\top)^\top \in \boldsymbol{\Theta}_\nu \subset \mathbb{R}^+ \times \mathbb{R}^{n_\nu-1}$ for a parameter space $\boldsymbol{\Theta}_\nu$. Based on this, we now define the scaled score function $\boldsymbol{s}(\cdot,\cdot)$ driving the dynamics of $\boldsymbol{\gamma}_i$ in line with the score-driven framework of Creal et al. (2013) and Harvey (2013) as

$$\boldsymbol{s}(\boldsymbol{Y}_i,\boldsymbol{\gamma}_i,\boldsymbol{\nu}) = \boldsymbol{S}(\boldsymbol{\gamma}_i,\boldsymbol{\nu}) \;\cdot\; \nabla(\boldsymbol{Y}_i,\boldsymbol{\gamma}_i,\boldsymbol{\nu}), \qquad \nabla(\boldsymbol{Y}_i,\boldsymbol{\gamma}_i,\boldsymbol{\nu}) = \frac{\partial \log p(\boldsymbol{Y}_i \mid \boldsymbol{\gamma};\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}}\bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i}, \tag{2.13}$$

where $\nabla(\boldsymbol{Y}_i,\boldsymbol{\gamma}_i,\boldsymbol{\nu})$ is called the score, and $\boldsymbol{S}(\boldsymbol{\gamma}_i,\boldsymbol{\nu})$ a scaling matrix for the score. By inserting this expression into (2.9), we adjust the coefficients $\boldsymbol{\gamma}_i$ of the product basis $\boldsymbol{\phi}_{\boldsymbol{K}}(\cdot)$ in a scaled steepest-ascent direction to improve the model fit as measured by (2.12); see Blasques et al. (2015), Creal et al. (2024), De Punder et al. (2024). The steps mimic local Gauss-Newton type of improvements in $\boldsymbol{\gamma}_i$ with step-size given in $\boldsymbol{A}$, and mean-reversion described by $\boldsymbol{B}$. The dynamics defined by (2.13) yield consistent estimates of the path of the time-varying parameter $\boldsymbol{\gamma}_i$, even in the presence of model misspecification; see Beutner et al. (2023).

To write down the proper expression for the score in (2.13), define the derivatives $\dot{\boldsymbol{g}}_\mu(\boldsymbol{f}) = \partial g_\mu(\boldsymbol{f})/\partial \boldsymbol{f} \in \mathbb{R}^{n_f}$ and $\dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}) = \partial g_\sigma(\boldsymbol{f})/\partial \boldsymbol{f} \in \mathbb{R}^{n_f}$ of the link functions $g_\mu(\cdot)$ and $g_\sigma(\cdot)$, respectively. We now get the following result.

**Proposition 1** (Score-driven dynamics)**.** *Using the score driven dynamics in* (2.13) *based on the Student's t process implied density* (2.12)*, we have*

$$\begin{aligned}
&\frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}_i} = \nabla(\boldsymbol{Y}_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}) = \nabla^\mu(\boldsymbol{Y}_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}) + \nabla^\sigma(\boldsymbol{Y}_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}), \qquad (2.14)\\
&\nabla^\mu(\boldsymbol{Y}_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}) = w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})\ \dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma}_i)^\top \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i),\\
&\nabla^\sigma(\boldsymbol{Y}_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}) = \dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma}_i)^\top \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \Big( w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})\ \big( (\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i)) \odot \boldsymbol{e}_i(\boldsymbol{\gamma}_i) \big) - \boldsymbol{\iota}_N \Big)\\
&\boldsymbol{e}_i(\boldsymbol{\gamma}_i) = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1}\ \left(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma}_i)\right), \qquad w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) = \frac{1 + \nu_1^{-1} N}{1 + \nu_1^{-1}\, \boldsymbol{e}_i(\boldsymbol{\gamma}_i)^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\ \boldsymbol{e}_i(\boldsymbol{\gamma}_i)},
\end{aligned}$$

*for an $N \times 1$ vector $\boldsymbol{\iota}_N$ filled with ones, $\odot$ denoting the (element-wise) Hadamard product of two vectors, and where $\dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma}_i) = \dot{\boldsymbol{G}}_\mu(\mathbb{T}, \boldsymbol{\gamma}_i) \in \mathbb{R}^{N \times n_\gamma}$ and $\dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma}_i) = \dot{\boldsymbol{G}}_\sigma(\mathbb{T}, \boldsymbol{\gamma}_i) \in \mathbb{R}^{N \times n_\gamma}$ given by*

$$\dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma}_i) = \begin{pmatrix} \dot{\boldsymbol{g}}_\mu(\boldsymbol{f}_i(\boldsymbol{t}_1))^\top \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_1)^\top \\ \vdots \\ \dot{\boldsymbol{g}}_\mu(\boldsymbol{f}_i(\boldsymbol{t}_N))^\top \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_N)^\top \end{pmatrix}, \qquad \dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma}_i) = \begin{pmatrix} \dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}_i(\boldsymbol{t}_1))^\top \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_1)^\top \\ \vdots \\ \dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}_i(\boldsymbol{t}_N))^\top \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_N)^\top \end{pmatrix}.$$

The result in Proposition 1 has an intuitive form. The proposed discretized stochastic recurrence equation (SRE) mirrors well-known $t$-GAS models and their robustness properties to outliers (Gasperoni et al., 2023; D'Innocenzo et al., 2023). The time-varying coefficients $\boldsymbol{\gamma}_i$ of the functional basis as summarized in all the cross-products in $\boldsymbol{\phi}_{\boldsymbol{K}}(\cdot)$ react to two sources of information, namely regarding the mean (the first term in (2.14)) and regarding the scale (the second term in (2.14)). The core of the first term is given by the weighted (scaled) error term $w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})\, \boldsymbol{e}_i(\boldsymbol{\gamma}_i)$. If the error term is positive, $\boldsymbol{Y}_i$ lies higher for some values of $\boldsymbol{t}_i \in \mathbb{T}$, while the converse is true for negative values. The adjustment to $\boldsymbol{\gamma}_i$ then tries to reduce the sizes of these (positive and negative) error terms. It does so by a projection of the error terms on the functional basis via the pre-multiplication by $\dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma}_i)^\top$. The step for each observation is weighted by $w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})$, which takes the value 1 if $\nu_1 \to \infty$. So in the case of a Gaussian process for $\boldsymbol{\varepsilon}_i(\cdot)$, all steps receive an equal weight. In case of a Student's $t$ process with $\nu_1 < \infty$, however, the time-varying parameter $\boldsymbol{\gamma}_i$ reacts less to an 'outlying' value of $\boldsymbol{Y}_i$ due the denominator of $w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})$ in (2.14). This gives the model a robustness property compared to the Gaussian case; compare Creal et al. (2013) for the univariate case or Gasperoni et al. (2023) for the finite-dimensional setting. In particular, this robustness thus not only comes via the objective function in (2.12), but also via the location (and scale) dynamics in (2.14).

The second term in (2.14) gives the reaction for the variance related part of $\boldsymbol{\gamma}_i$. In this case, the standardized quadratic error terms play a dominant role. If these are higher than their conditionally expected value, i.e., higher than one, the scales need to be adjusted upwards. The converse holds if the squared errors are below their conditional expectation. As before, the $\boldsymbol{\gamma}_i$s are adjusted to reduce these discrepancies by projection the errors on the functional basis by the pre-multiplication by $\dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma}_i)^\top$. We also again see that outlying observations are downweighted by the presence of the weight $w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})$ in the second term in (2.14). Only for the Gaussian process ($\nu_1 \to \infty$) this weight is always equal to one. Finally note that in the special case where $\boldsymbol{\gamma}_i$ can be split into $\boldsymbol{\gamma}_i^\mu$ and $\boldsymbol{\gamma}_i^\sigma$, such that $\boldsymbol{\mu}(\boldsymbol{\gamma}_i)$ only depends on $\boldsymbol{\gamma}_i^\mu$ and $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)$ only depends on $\boldsymbol{\gamma}_i^\sigma$, then the first term in (2.14) only affects the mean dynamics, whereas the second term only affects the scale dynamics.

We gather the static parameters of the model in the vector $\boldsymbol{\theta}^\top = \big(\boldsymbol{\nu}^\top, \boldsymbol{\omega}^\top, \mathrm{vec}(\boldsymbol{A})^\top, \mathrm{vec}(\boldsymbol{B})^\top\big)$. We also slightly change the notation and explicitly write $\boldsymbol{\gamma}_i(\boldsymbol{\theta})$ to stress that the time-varying parameters are a function of $\boldsymbol{\theta}$ via the recursion in (2.13). For estimation purposes in a finite sample, we also have to initialize the recursion by some nonrandom initial value $\widehat{\gamma}_0 \in \boldsymbol{\mathcal{G}}$. This initialized sequence is denoted by $\widehat{\gamma}_i(\boldsymbol{\theta}) = \widehat{\gamma}_i(\boldsymbol{\theta}, \widehat{\gamma}_0)$. For $N \in \mathbb{Z}^+$, we have the time $i$ average likelihood function

$$
\begin{aligned}
\ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big) &= N^{-1}\log p\big(\boldsymbol{Y}_i \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta});\boldsymbol{\nu}\big) \\
&= \frac{1}{N}\log\left(\frac{\Gamma\big((\nu_1+N)/2\big)}{\Gamma\big(\nu_1/2\big)(\nu_1\pi)^{N/2}}\right) - \frac{1}{N}\log\det\Big(\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)\Big) \\
&\quad - \frac{1}{2N}\log\det\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)\big) - \frac{\nu_1+N}{2N}\log\Big(1+\nu_1^{-1}\boldsymbol{e}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)\Big).
\end{aligned}
\tag{2.15}
$$

We now define our estimator for $\boldsymbol{\theta}$ as

$$
\widehat{\boldsymbol{\theta}}_T = \widehat{\boldsymbol{\theta}}_T\big(\widehat{\boldsymbol{\gamma}}_0\big) := \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} T^{-1}\sum_{i=1}^{T}\ell_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big), \tag{2.16}
$$

where $\boldsymbol{\Theta}\subset\mathbb{R}^{n_\theta}$ denotes the parameter space of dimension $n_\theta$. The next section establishes the asymptotic properties of the estimator.

# 3 Asymptotic theory

To derive the asymptotic properties of the new estimator, let $\boldsymbol{\theta}_0^\top = (\boldsymbol{\nu}_0^\top, \boldsymbol{\omega}_0^\top, \operatorname{vec}(\boldsymbol{A}_0)^\top, \operatorname{vec}(\boldsymbol{B}_0)^\top)$ denote the vector of true parameters. We assume a convex, separable Banach filter space $\boldsymbol{\mathcal{G}}\subset\mathbb{R}^{n_\gamma}$ equipped with the norm $\|\cdot\| = \|\cdot\|_2$, such that $\boldsymbol{\gamma}_i(\boldsymbol{\theta})\in\boldsymbol{\mathcal{G}}$ for every $i\in\mathbb{Z}$ and $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. Throughout, we require that $\sum_{j=1}^{d} K_j < \infty$. Finally, for convenience, let $\boldsymbol{\mathcal{X}}$ be an *open* subset of $\mathbb{R}^{n_\mathcal{X}}$, and let $\boldsymbol{\mathcal{Y}}$ be a subset of $\mathbb{R}^{n^r_\mathcal{Y}\times n^c_\mathcal{Y}}$, where $n_\mathcal{X}, n^r_\mathcal{Y}, n^c_\mathcal{Y}\in\mathbb{Z}^+$. For $k\geq 0$, we write $\boldsymbol{F}\in\mathcal{C}^k(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Y}})$ if $\boldsymbol{F}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ is a possibly matrix-valued mapping, and each of its components has continuous derivatives on $\boldsymbol{\mathcal{X}}$ up to order $k$.

## 3.1 Stationarity and invertibility

We start by considering the model as a data generating process and proving the properties of stationarity and ergodicity, as well as filter invertibility. We make the following mild assumptions.

**Assumptions:** *A1 Let $\mathbb{F}$, the parameter space of $\boldsymbol{f}_i$, be some open subset of $\mathbb{R}^{n_f}$. Assume that the link functions satisfy $g_\mu \in \mathcal{C}^1(\mathbb{F}, \mathbb{R})$ and $g_\sigma \in \mathcal{C}^1(\mathbb{F}, \mathbb{R}^+)$ and that for $N \in \mathbb{Z}^+$ fixed,*

$$\sup_{\boldsymbol{\gamma} \in \boldsymbol{\mathcal{G}}} \left\| \dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma})^\top \boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1} \right\| < \infty, \qquad \sup_{\boldsymbol{\gamma} \in \boldsymbol{\mathcal{G}}} \left\| \dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma})^\top \boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1} \right\| < \infty. \tag{3.1}$$

*A2* $\sup_{\boldsymbol{\nu} \in \boldsymbol{\Theta}_\nu} \left\| \boldsymbol{\Lambda}(\boldsymbol{\nu}_2) \right\| < \infty$, $\sup_{\boldsymbol{\nu} \in \boldsymbol{\Theta}_\nu} \left\| \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \right\| < \infty$.

An example illustrating these assumptions will be provided below the following proposition. When considering the model as a DGP, we replace $\boldsymbol{Y}_i$ in (2.14) by its value in the DGP, i.e., by $\boldsymbol{\mu}(\boldsymbol{\gamma}_i) + \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\boldsymbol{\varepsilon}_i$. To distinguish this from the original recursion in (2.14), we substitute $\boldsymbol{\gamma}_i$ by $\boldsymbol{\gamma}_i^\varepsilon$. Let $\widehat{\boldsymbol{\gamma}}_i^\varepsilon = \widehat{\boldsymbol{\gamma}}_i^\varepsilon(\boldsymbol{\theta}, \widehat{\boldsymbol{\gamma}}_0^\varepsilon)$ denote its initialized counterpart, initialized at $i = 0$ by a nonrandom $\widehat{\boldsymbol{\gamma}}_0^\varepsilon$. We now obtain the following result.

**Proposition 2** (Existence of a DGP)**.** *Consider the stochastic recurrence equation (SRE) of Proposition 1, evaluated at $\widehat{\boldsymbol{\gamma}}_i^\varepsilon$ and at $\boldsymbol{Y}_i = \boldsymbol{\mu}(\widehat{\boldsymbol{\gamma}}_i^\varepsilon) + \boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}}_i^\varepsilon)\boldsymbol{\varepsilon}_i$, i.e., the SRE:*

$$\widehat{\boldsymbol{\gamma}}_{i+1}^\varepsilon = \boldsymbol{\omega} + \boldsymbol{B}\, \widehat{\boldsymbol{\gamma}}_i^\varepsilon + \boldsymbol{A}\, \boldsymbol{S}(\widehat{\boldsymbol{\gamma}}_i^\varepsilon, \boldsymbol{\nu}) \nabla^\varepsilon\big(\boldsymbol{\varepsilon}_i, \widehat{\boldsymbol{\gamma}}_i^\varepsilon, \boldsymbol{\nu}\big), \quad \nabla^\varepsilon\big(\boldsymbol{\varepsilon}_i, \widehat{\boldsymbol{\gamma}}_i^\varepsilon, \boldsymbol{\nu}\big) = \nabla\Big(\boldsymbol{\mu}(\widehat{\boldsymbol{\gamma}}_i^\varepsilon) + \boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}}_i^\varepsilon)\boldsymbol{\varepsilon}_i, \widehat{\boldsymbol{\gamma}}_i^\varepsilon, \boldsymbol{\nu}\Big). \tag{3.2}$$

*Suppose Assumptions A1 - A2 hold and that for $N \in \mathbb{Z}^+$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,*

*SE1* $\left\| \boldsymbol{S}(\widehat{\boldsymbol{\gamma}}_0^\varepsilon, \boldsymbol{\nu}) \right\| < \infty$ *for some constant vector* $\widehat{\boldsymbol{\gamma}}_0^\varepsilon \in \boldsymbol{\mathcal{G}}$*;*

*SE2* $\mathbb{E}\left( \log \sup_{\boldsymbol{\gamma} \in \boldsymbol{\mathcal{G}}} \left\| \boldsymbol{B} + \boldsymbol{A} \dfrac{\partial}{\partial \boldsymbol{\gamma}^\top} \Big( \boldsymbol{S}(\boldsymbol{\gamma}, \boldsymbol{\nu}) \nabla^\varepsilon(\boldsymbol{\varepsilon}_i, \boldsymbol{\gamma}, \boldsymbol{\nu}) \Big) \right\| \right) < 0.$

*Then, for a fixed $N \in \mathbb{Z}^+$, there exists a unique strictly stationary and ergodic (SE) solution $\big\{\boldsymbol{\gamma}_i^\varepsilon(\boldsymbol{\theta}),\, i \in \mathbb{Z}\big\}$ to (3.2) for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, as $i \to \infty$.*

Since $\boldsymbol{Y}_i$ is a measurable function of $\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\varepsilon}_i\big)$ for $i \in \mathbb{Z}$, the existence of a unique SE solution to (3.2) also implies that $\{\boldsymbol{Y}_i,\, i \in \mathbb{Z}\}$ is an SE sequence (White, 2001, Theorem 3.35). The conditions in Proposition 2 are mild and only require that the SRE for $\boldsymbol{\gamma}_i$ is contracting and that its derivative with respect to $\boldsymbol{\gamma}_i$ has a log moment.

The following example illustrates how the assumptions of Proposition 2, and in particular Assumption SE2, can be verified.

**Example (Volatility curves).** We consider an exponential version of the functional volatility curve model (2.5) from Section 2.2 by setting $g_\mu(f) \equiv 0$ and $g_\sigma(f) = \exp(f/2)$. We use so-called unit score scaling in the sense of Creal et al. (2013) and obtain $\boldsymbol{S}(\cdot,\cdot) \equiv \boldsymbol{I}_{n_\gamma}$ and

$$\boldsymbol{\gamma}_{i+1} = \boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{\gamma}_i + \frac{1}{2}\boldsymbol{A}\Big(\boldsymbol{\phi}_K(t_1)\cdots\boldsymbol{\phi}_K(t_N)\Big)\left(\frac{\big(1+\nu_1^{-1}N\big)\big((\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma}_i))\odot\boldsymbol{e}_i(\boldsymbol{\gamma}_i)\big)}{1+\nu_1^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma}_i)^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma}_i)} - \boldsymbol{\iota}_N\right). \tag{3.3}$$

As a result, $\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\big\|\dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma})\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\big\| = 0$. Also, using the fact that $\dot{g}_\sigma(f)/g_\sigma(f) = 1/2$, we have $\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\big\|\dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma})\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\big\| = \frac{1}{2}\Big(\sum_{j=1}^N \boldsymbol{\phi}_K(t_j)\boldsymbol{\phi}_K(t_j)^\top\Big)^{1/2} < \infty$, given square-integrable basis functions $\phi_1,\ldots,\phi_K \in L^2[0,1]$. Assumption A1 is thus easily fulfilled. For A2, we have to restrict the covariance structure of the process. In case of the autoregressive structure above (2.11), it suffices to assume $0 < \nu_2 < \infty$. Also for more general cases like banded Toeplitz covariance structures, there are easy-to-check sufficient ensuring the eigenvalues to be bounded from below and above (see Lemma 4.1 Gray, 2006). Assumption SE1 is trivially satisfied as we have unit scaling. Finally, to check SE2, note that

$$\nabla_i^\varepsilon(\boldsymbol{\varepsilon},\boldsymbol{\gamma},\boldsymbol{\nu}) = \frac{1}{2}\Big(\boldsymbol{\phi}_K(t_1)\cdots\boldsymbol{\phi}_K(t_N)\Big)\left(\frac{\big(1+\nu_1^{-1}N\big)\big((\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i)\odot\boldsymbol{\varepsilon}_i\big)}{1+\nu_1^{-1}\boldsymbol{\varepsilon}_i^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i} - \boldsymbol{\iota}_N\right),$$

which does not depend on $\boldsymbol{\gamma}_i$, such that Assumption SE2 reduces to $\|\boldsymbol{B}\| < 1$. □

Next, we establish conditions for filter invertibility. Filter invertibility is an essential property for establishing the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$. It requires that the effect of the filter initialization by an arbitrary nonrandom $\widehat{\boldsymbol{\gamma}}_0$ vanishes sufficiently fast. To formulate the result, we express the SRE in (2.9) as $\widehat{\boldsymbol{\gamma}}_{i+1}(\boldsymbol{\theta}) = \boldsymbol{\psi}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\big)$, where

$$\boldsymbol{\psi}_i\big(\boldsymbol{\gamma},\boldsymbol{\theta}\big) = \boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{A}\,\boldsymbol{S}\big(\boldsymbol{\gamma},\boldsymbol{\nu}\big)\nabla\big(\boldsymbol{Y}_i,\boldsymbol{\gamma},\boldsymbol{\nu}\big). \tag{3.4}$$

Furthermore, we define the $r$-fold convolution $\boldsymbol{\psi}_i^{(r)}(\,\cdot\,,\boldsymbol{\theta})$ of the function $\boldsymbol{\psi}_i(\,\cdot\,,\boldsymbol{\theta})$ as $\boldsymbol{\psi}_i^{(r)}(\,\cdot\,,\boldsymbol{\theta}) = \boldsymbol{\psi}_i(\,\cdot\,,\boldsymbol{\theta}) \circ \boldsymbol{\psi}_{i-1}(\,\cdot\,,\boldsymbol{\theta}) \circ \ldots \circ \boldsymbol{\psi}_{i-r+1}(\,\cdot\,,\boldsymbol{\theta})$ for $r \in \mathbb{Z}^+$. Let $\log^+(x) = \max\{\log(x), 0\}$ for $x > 0$. We now have the following result.

**Proposition 3** (Invertibility). *Suppose Assumptions A1 - A2 hold. Let $\{\mathbf{Y}_i,\, i \in \mathbb{Z}\}$ be an SE sequence and $\boldsymbol{\Theta}$ the compact parameter space. In addition, assume*

*IV1* $\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\boldsymbol{S}(\widehat{\gamma}_0,\boldsymbol{\nu})\right\| < \infty$ *for some* $\widehat{\gamma}_0 \in \boldsymbol{\mathcal{G}}$*;*

*IV2* $\mathbb{E}\Bigg(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\boldsymbol{B} + \boldsymbol{A}\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\Big(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\nabla\big(\mathbf{Y}_i,\boldsymbol{\gamma},\boldsymbol{\nu}\big)\Big)\right\|\Bigg) < \infty$*;*

*IV3* $\mathbb{E}\Bigg(\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{\psi}_i^{(r)}\big(\boldsymbol{\gamma},\boldsymbol{\theta}\big)\right\|\Bigg) < 0$ *for some integer* $r \geq 1$*.*

*Then, for a fixed $N \in \mathbb{Z}^+$, the sequence $\big\{\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\, i \in \mathbb{Z}^+\big\}$ initialized at any starting value $\widehat{\boldsymbol{\gamma}}_0$ converges exponentially fast almost surely (e.a.s.) to a unique SE solution $\big\{\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\, i \in \mathbb{Z}\big\}$ of $\widehat{\boldsymbol{\gamma}}_{i+1}(\boldsymbol{\theta}) = \boldsymbol{\psi}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\big)$ as $i \to \infty$, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$; i.e., there exists some $\rho > 1$ such that $\rho^i \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\right\| \xrightarrow{a.s.} 0$ as $i \to \infty$.*

The assumptions in Proposition 3 are similar to those in Proposition 2 in terms of the existence of a log moment and the contraction of the SRE. The main difference lies in the fact that the SRE for stationarity and ergodicity (in terms of $\boldsymbol{\varepsilon}_i$) is generally substantially different from the SRE for invertibility (in terms of $\mathbf{Y}_i$). Note that the invertibility of the filter immediately implies the invertibility of the process $\big\{\widehat{\boldsymbol{f}}_i(\cdot),\, i \in \mathbb{Z}^+\big\}$, where $\widehat{\boldsymbol{f}}_i(\cdot) = \widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\theta})\boldsymbol{\phi}_K(\cdot)$, with $\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) = \mathrm{vec}\Big(\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\theta})^\top\Big)$, provided that the basis functions are uniformly bounded over $\boldsymbol{\mathcal{T}}$.

With the existence of a DGP and filter invertibility being established, we can now prove the strong consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$. We first make the following additional assumptions.

**Assumptions:** *SC1* $\forall \boldsymbol{f} \in \mathbb{R}^{n_f}$*, there exists some constants* $C_\sigma > 0$ *and* $\eta \geq 0$ *such that*

$$\big|\log g_\sigma(\boldsymbol{f})\big| \leq C_\sigma \|\boldsymbol{f}\|^\eta, \qquad \big|\log g_\sigma(\boldsymbol{f}^0) - \log g_\sigma(\boldsymbol{f})\big| \leq C_\sigma \big\|\boldsymbol{f}^0 - \boldsymbol{f}\big\|^\eta. \tag{3.5}$$

*If* (3.5) *holds only for some* $\eta > 0$, *and not for* $\eta = 0$, *we additionally require: (1)* $\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| < \infty$; *(2)* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{B}\| < 1$; *(3)* $\sum_{j=1}^{N} \left\|\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)\right\|^{\eta} < \infty$ *for* $N \in \mathbb{Z}^+$. *Similarly,* $\forall \boldsymbol{f} \in \mathbb{R}^{n_f}$, *there exists some constants* $C_\mu > 0$ *and* $\zeta \geq 0$ *such that*

$$\left|g_\mu(\boldsymbol{f}^0) - g_\mu(\boldsymbol{f})\right| \leq C_\mu \|\boldsymbol{f}^0 - \boldsymbol{f}\|^{\zeta}. \tag{3.6}$$

*If* (3.6) *only holds for some* $\zeta > 0$, *and not for* $\zeta = 0$, *then we also require* $\sum_{j=1}^{N} \left\|\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)\right\|^{\zeta} < \infty$ *for* $N \in \mathbb{Z}^+$.

*SC2* $p\big(\boldsymbol{y}\,\big|\,\boldsymbol{\gamma},\boldsymbol{\nu}\big) = p\big(\boldsymbol{y}\,\big|\,\tilde{\boldsymbol{\gamma}},\tilde{\boldsymbol{\nu}}\big)$ *for almost every* $\boldsymbol{y} \in \mathbb{R}^N$ *with respect to the Lebesgue measure if and only if* $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$ *and* $\boldsymbol{\nu} = \tilde{\boldsymbol{\nu}}$.

*SC3* $\|\boldsymbol{B}_0\| < 1$ *and* $\det(\boldsymbol{A}_0) \neq 0$.

Assumption SC1 imposes that the link functions for the location and scale parameters are sufficiently regular in $\boldsymbol{f}_i(\cdot)$. Assumptions SC2 and SC3 ensure identification by assuming the density $p$ of $\boldsymbol{Y}_i$ changes noticeably with changing values of $\boldsymbol{\gamma}_i$ and/or $\nu_1$, and by requiring the data to have a non-negligible influence on $\boldsymbol{\gamma}_i$. We now have the following strong consistency result.

**Theorem 1** (Strong consistency)**.** *Suppose the discretely observed data is a subset of the realized path of an SE stochastic process* $\{\boldsymbol{Y}_i,\, i \in \mathbb{Z}\}$ *generated by* (2.9) *and* (2.10) *with* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. *Let the assumptions of Proposition 3 and Assumptions SC1 - SC3 hold. Then, for any filter initialization* $\widehat{\boldsymbol{\gamma}}_0 \in \boldsymbol{\mathcal{G}}$, *we have* $\widehat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ *as* $T \to \infty$.

The following assumptions are required to analyze the asymptotic properties of derivative processes, which are subsequently used to derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_T$.

**Assumptions:** *AN1 Let* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{B}\| < 1$. *Assume* $\boldsymbol{S}(\cdot,\cdot) \in \mathcal{C}^3\big(\boldsymbol{\mathcal{G}} \times \mathbb{R}^{n_\nu}, \mathbb{R}^{n_\gamma \times n_\gamma}\big)$ *with derivatives up to order 3 uniformly bounded on* $\boldsymbol{\mathcal{G}} \times \boldsymbol{\Theta}_\nu$.[2] *Similarly, assume* $\boldsymbol{\Lambda}(\cdot)^{-1} \in \mathcal{C}^2\big(\mathbb{R}^{n_\nu - 1}, \mathbb{R}^{N\times N}\big)$ *with derivatives up to order 2 uniformly bounded on* $\boldsymbol{\Theta}_\nu$.

[2]For instance, $\sup_{(\boldsymbol{\gamma},\boldsymbol{\nu})\in\boldsymbol{\mathcal{G}}\times\boldsymbol{\Theta}_\nu} \left|\partial^k S_{i,j}(\boldsymbol{\gamma},\boldsymbol{\nu})/(\partial\gamma_l^{k_l}\partial\nu_m^{k_m}\partial\nu_n^{k_n})\right| < \infty$ for $k_l + k_m + k_n \leq 3$, where $S_{i,j}(\cdot,\cdot)$ is the $(i,j)_{\text{th}}$ element of $\boldsymbol{S}(\cdot,\cdot)$, $\gamma_l$ is the $l_{\text{th}}$ element of $\boldsymbol{\gamma}$, while $\nu_m$ and $\nu_n$ are the $m_{\text{th}}$ and $n_{\text{th}}$ elements of $\boldsymbol{\nu}$, respectively.

*AN2 Suppose the link functions* $g_\mu \in \mathcal{C}^4(\mathbb{F},\mathbb{R})$ *and* $g_\sigma \in \mathcal{C}^4(\mathbb{F},\mathbb{R}^+)$. *We assume that each component of* $\dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma})^\top \boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}$ *and* $\dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma})^\top \boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}$ *has derivatives with respect to* $\boldsymbol{\gamma}$ *up to order 3, which are uniformly bounded in* $\boldsymbol{\gamma} \in \boldsymbol{\mathcal{G}}$.

*AN3 For* $i \in \mathbb{Z}$, *there exists a constant* $\kappa > 0$ *and a positive real sequence* $\{\varrho_\kappa(j),\, j \geq 1\}$ *such that*

$$\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{\psi}_i^{(j)}(\boldsymbol{\gamma},\boldsymbol{\theta})\right\|^\kappa\right) \leq \varrho_\kappa(j), \qquad j \geq 1, \tag{3.7}$$

*with* $\sum_{j=1}^\infty \left[\varrho_\kappa(j)\right]^{1/\kappa} < \infty$, *where* $\boldsymbol{\psi}_i^{(j)}(\,\cdot\,,\boldsymbol{\theta})$ *is the* $j$*-fold convolution as defined below* (3.4), *and* $j \in \mathbb{Z}^+$.

*AN4 The compact parameter space* $\boldsymbol{\Theta}$ *coincides with the closure of its (open) interior.*

We see that Assumption AN3 is stronger than Assumptions IV2 and IV3 by Jensen's inequality. It not only ensures the stationarity and ergodicity of the derivative processes of $\{\boldsymbol{\gamma}_i(\cdot)\}$, but Assumption AN3 also guarantees the existence of their moments. These are needed to apply Theorem 2.10 of Straumann and Mikosch (2006) to perturbed processes. Note that Assumption AN1 is also stronger than IV1. Therefore, under the assumptions of Proposition 4 below, Proposition 3 continues to hold. Finally, Assumption AN4 excludes compact parameter spaces that have an empty interior or isolated points in the standard topology of $\mathbb{R}^{n_\theta}$, as these would complicate the existence of derivatives.

Before proceeding, we define the following derivative processes for the time-varying parameter $\boldsymbol{\gamma}_i(\cdot)$. For $k \in \mathbb{Z}^+$, let $\boldsymbol{\gamma}_i^{(k)}(\boldsymbol{\theta}) = \operatorname{vec}\left(\partial\boldsymbol{\gamma}_i^{(k-1)}(\boldsymbol{\theta})\big/\partial\boldsymbol{\theta}^\top\right)$, where $\boldsymbol{\gamma}_i^{(0)}(\boldsymbol{\theta}) = \boldsymbol{\gamma}_i(\boldsymbol{\theta})$. Similarly, $\widehat{\boldsymbol{\gamma}}_i^{(k)}(\boldsymbol{\theta}) = \operatorname{vec}\left(\partial\widehat{\boldsymbol{\gamma}}_i^{(k-1)}(\boldsymbol{\theta})\big/\partial\boldsymbol{\theta}^\top\right)$, with $\widehat{\boldsymbol{\gamma}}_i^{(0)}(\boldsymbol{\theta}) = \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})$. For $k = 1, 2$, the explicit expressions of $\boldsymbol{\gamma}_i^{(k)}(\boldsymbol{\theta})$ are provided in Lemma F.7.

**Proposition 4** (Derivative processes)**.** *Let Assumptions A1 - A2 and AN1 - AN4 hold. Let* $\{\boldsymbol{Y}_i,\, i \in \mathbb{Z}\}$ *be an SE sequence. For* $k = 1, 2$, *there exist unique SE sequences* $\{\boldsymbol{\gamma}_i^{(k)}(\boldsymbol{\theta}),\, i \in \mathbb{Z}\}$

*such that* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\gamma}_i^{(k)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(k)}(\boldsymbol{\theta}) \right\| \xrightarrow{e.a.s.} 0$*, as* $i \to \infty$.

Proposition 4 establishes that the first- and second-order derivative processes enjoy the same invertibility property as the filter itself. This is crucial for deriving an asymptotic approximation of $\widehat{\boldsymbol{\theta}}_T$, which depends on the uniform convergence of the Hessian matrix of the objective function (2.16).

**Remark 1.** It is well established in the literature on M-estimators that a local dominance condition on the Hessian matrix $\partial^2 \mathcal{L}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ enables the application of a uniform law of large numbers and is sufficient for establishing asymptotic normality; see, for instance, Newey and McFadden (1994, Section 3). This, in turn, allows us to impose a weaker condition to guarantee the existence of certain finite moments for the first- and second-order derivative processes only within a local compact neighborhood $\mathcal{V}(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$, rather than across $\boldsymbol{\Theta}$. Specifically, we can replace (3.7) with $\mathbb{E}\big(\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta}_0)} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}_0} \left\|\partial\boldsymbol{\psi}_i^{(j)}(\boldsymbol{\gamma},\boldsymbol{\theta})/\partial\boldsymbol{\gamma}^\top\right\|^\kappa\big) \leq \varrho_\kappa(j)$, $j \geq 1$, where $\left\{\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\, i \in \mathbb{Z},\, \boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}_0)\right\} \subset \boldsymbol{\mathcal{G}}_0 \subset \boldsymbol{\mathcal{G}}$. However, it remains unclear whether this relaxation provides enough empirical benefits to off-set the additional notational complexity. We therefore adhere to the stronger condition in AN3.

**Theorem 2** (Asymptotic normality)**.** *Assume the discretely observed data is a subset of the realized path of an SE stochastic process* $\left\{\boldsymbol{Y}_i,\, i \in \mathbb{Z}\right\}$ *generated by* (2.9) *and* (2.10)*, where* $\boldsymbol{\theta}_0$ *lies in the interior of* $\boldsymbol{\Theta}$*. Moreover, let Assumptions A1 - A2, SC1 - SC3, and AN1 - AN4 hold, with Assumption AN3 satisfied for* $\kappa \geq 3$*. If* $\boldsymbol{\mathcal{I}}_0 = \mathbb{E}\Big(\partial^2 \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\Big)$ *is invertible, then* $\sqrt{T}\big(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\big) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\mathcal{I}}_0^{-1}\big)$ *as* $T \to \infty$.

Theorem 2 derives an asymptotic approximation of the static parameter estimator. To establish the theorem, we rely on the condition that the total number of basis functions, $\sum_{j=1}^d K_j$, is finite. Allowing this quantity to diverge to infinity presents significant challenges, as also discussed in Cerovecki et al. (2019), making it nontrivial to extend our asymptotic framework to accommodate an infinite number of basis functions. In our context, it would require (i) establishing the invertibility property of an infinite-dimensional time-varying parameter filter process along with its derivative processes, and (ii) proving weak convergence results in infinite-dimensional spaces to obtain a limiting approximation of the static parameter estimator. Both challenges remain open.

# 4 Simulations

In this section, we evaluate the performance of the proposed method in a controlled setting. Section 4.1 examines an example of time-varying volatility curves ($d = 1$). Section 4.2 explores an example of time-varying surfaces ($d = 2$), mimicking real-life scenarios such as air pollution surfaces. In both examples, we focus on cases of model misspecification, i.e., settings where the true dynamics of the DGP do not correspond to the observation-driven dynamics of the new method. The results show that despite such possible misspecification, the new method performs well in recovering the true underlying location and scale dynamics.

## 4.1 Time-varying volatility curves

Let the process $\{Y_i(t),\, i \in [\![ T ]\!],\, t \in [0,1]\}$ be generated by $Y_i(t) = \sigma_i(t)\varepsilon_i(t)$, where $\varepsilon_i(t) \sim t_{\nu_1}$ is independent across $i \in \mathbb{Z}^+$, with the Ornstein-Uhlenbeck covariance kernel $\text{cov}(\varepsilon_i(t), \varepsilon_i(s)) = \exp\big(-|s-t|/\nu_2\big)$, and where

$$\sigma_i^2(t) = 4 + 4\left(2t - 1 - \sin\left(\frac{4\pi\, i}{2000} - u_1\right)\right)^2 + 2\sin\left(\frac{2\pi\, i}{2000} - u_2\right), \tag{4.1}$$

for $u_1, u_2$ that are uniform on $[0, 2\pi]$ and drawn independently for each Monte Carlo replication. The DGP in (4.1) produces quadratic shapes in $t$ for each $i$, where the height and the location of the minimum of this quadratic shape slowly over $i$. The shapes are also randomized in terms of phase shifts across simulations via the random variables $u_1$ and $u_2$, which do not depend on $i$ or $t$. An example of the complex volatility pattern for $u_1 = u_2 = 0$ is presented in Figure 1. Note that the fGAS model is misspecified for the current DGP. However, it is still able to recover the paths of $\sigma_i(t)$ across both $i$ and $t$. We assume the process $Y_i(t)$ is observed on an equally-spaced grid $\mathbb{T} = \{t_j = j/N, \, j \in [\![ N ]\!]\}$.

We adopt a functional version of the E-GARCH model of Nelson (1990) or the E-GAS model of Harvey (2013) by setting $g_\mu(\cdot) \equiv 0$ and $g_\sigma(f) = \exp(f/2)$ in (2.1), i.e., we model the log variances or standard deviations. The updating equation for $\boldsymbol{\gamma}_i$ is given by (3.3) from Section 3. To improve computational efficiency in the simulations, we avoid brute-force inversions of large-dimensional matrices like $\boldsymbol{\Lambda}(\nu_2)$. Instead, we use the modified Cholesky decomposition from Pourahmadi (1999), which provides a straightforward expression for the inverse of $\boldsymbol{\Lambda}(\nu_2)$, as discussed in Beutner et al. (2023). We obtain $\boldsymbol{\Lambda}(\nu_2)^{-1} = \boldsymbol{F}(\nu_2)^\top \boldsymbol{D}(\nu_2) \boldsymbol{F}(\nu_2)$, where $\boldsymbol{F}(\nu_2)$ is a lower triangular matrix with 1 on the main diagonal and $-\rho_{\nu_2} = -\exp(-1/[(N-1)\nu_2])$ on the first subdiagonal, and zero elsewhere, and $\boldsymbol{D}(\nu_2) = (1-\rho_{\nu_2}^2)^{-1} \cdot \operatorname{diag}\left(1-\rho_{\nu_2}^2, 1, \ldots, 1\right)$. It directly follows that $\log \left|\boldsymbol{\Lambda}(\nu_2)\right| = (N-1)\log\left(1-\rho_{\nu_2}^2\right)$. Since we observe in preliminary experiments that the estimates of the degrees of freedom parameter $\nu_1$ may be downward biased in small samples, particularly when the true value of $\nu_1$ is large, we consider two estimators for $\boldsymbol{\theta}$: one with $\nu_1$ is estimated, denoted as fGAS$(\hat{\nu}_1)$, and another where $\nu_1$ is (infeasibly) fixed at its true value, denoted as fGAS$(\nu_1)$.

As our benchmark we use the functional GARCH model (denoted as fGARCH) of Cerovecki et al. (2019); see also Section 2.2. Their objective function differs from ours as presented in Section 2.4 as we adhere to the original construction outlined in Cerovecki et al. (2019, Eq. (3.5)). We also follow the standard convention of approximating Riemann integrals using a Riemann sum (see, e.g., Rice et al., 2020), for instance, $\int \boldsymbol{\phi}_K(s) Y_i^2(s)\,\mathrm{d}s \approx N^{-1}\sum_{j=1}^N \boldsymbol{\phi}_K(s_j) Y_i^2(s_j)$, where $\boldsymbol{\phi}_K(\cdot) = \big(\phi_1(\cdot), \ldots, \phi_K(\cdot)\big)^\top$ is defined earlier in Section 2.2.

All methods use $K = 7$ B-spline basis functions of order 4, resulting in a piecewise polynomial, positive function of degree 3, with three equidistant interior control knots at $t = 0.25, 0.5, 0.75$ (Ramsay et al., 2009, Chapter 3.3.4). To further reduce the computational cost of the simulation experiment, we let $\boldsymbol{A}$ and $\boldsymbol{B}$ be diagonal.

We evaluate the out-of-sample (rather than the in-sample) simulated mean absolute errors (MAEs) to avoid differences in performance due to overfitting. For this, we generate $T + T_{\text{out}}$ discrete paths of $\{Y_i(\cdot)\}$ for some $T_{\text{out}} \in \mathbb{Z}^+$. The first $T$ of these, $\{Y_1(\cdot), \ldots, Y_T(\cdot)\}$, are used to estimate the static parameters. The remaining $T_{\text{out}}$ paths, $\{Y_{T+1}(\cdot), \ldots, Y_{T+T_{\text{out}}}(\cdot)\}$, serve as the out-of-sample set for computing the fitted curves $\left\{\widehat{\sigma}_{T+1}(\cdot), \ldots, \widehat{\sigma}_{T+T_{\text{out}}}(\cdot)\right\}$ obtained via either our proposed method (fGAS) or the benchmark approach (fGARCH). For fGAS, we define $\widehat{\sigma}_i = g_\sigma(\hat{f}_i)$, where $\hat{f}_i(\cdot) = \boldsymbol{\phi}_K(\cdot)^\top \widehat{\boldsymbol{\gamma}}_i(\widehat{\boldsymbol{\theta}}_T)$, with $\widehat{\boldsymbol{\theta}}_T$ obtained from (2.16) based on the first $T$ observed paths. For fGARCH, which requires $\mathbb{E}[\varepsilon_i(t)] = 1$ for $t \in [0, 1]$, we adjust the estimated curves as $\widehat{\sigma}_i = (1 - 2\nu_1^{-1})^{1/2}\widetilde{\sigma}_i$, where $\widetilde{\sigma}_i^2(\cdot) = \boldsymbol{\phi}_K(\cdot)^\top \widetilde{\boldsymbol{\gamma}}_i(\widetilde{\boldsymbol{\theta}}_T)$ and $\widetilde{\boldsymbol{\gamma}}_i$ follows the recursion in (2.5), and where $\widetilde{\boldsymbol{\theta}}_T$ is obtained from Cerovecki et al. (2019, Eq. (3.5)). This ensures that the estimated paths of the dynamic scales for both the fGAS and the fGARCH are comparable. For each Monte Carlo replication, we compute the full MAE as $(NT_{\text{out}})^{-1} \sum_{j=1}^{N} \sum_{i=1}^{T_{\text{out}}} \left|\widehat{\sigma}_{T+i}(t_j) - \sigma_{T+i}(t_j)\right|$. We also compute the MAE separately across dimensions $i$ and $t$ as $N^{-1} \sum_{j=1}^{N} \left|\widehat{\sigma}_{T+i}(t_j) - \sigma_{T+i}(t_j)\right|$, for $i \in [\![\, T_{\text{out}} \,]\!]$, and $T_{\text{out}}^{-1} \sum_{i=1}^{T_{\text{out}}} \left|\widehat{\sigma}_{T+i}(t_j) - \sigma_{T+i}(t_j)\right|$, for $j \in [\![\, N \,]\!]$. The MAEs are then averaged over the Monte Carlo simulations.

We consider parameters that cover both sparse and relatively dense grids $\mathbb{T}$ on which $Y_i(t)$ is observed by using $N \in \{25, 100\}$. We also consider different sample sizes $T \in \{500, 1500\}$, fat-tailed $\nu_1 = 3$ and relatively light-tailed $\nu_1 = 10$ distributions, and weak $\nu_2 = 0.01$ and strong $\nu_2 = 0.1$ dependence. We set the number of out-of-sample observations to $T_{\text{out}} = 1000$. Table 1 presents the results.

Table 1 clearly shows that the fGAS model outperforms the fGARCH benchmark. This holds particularly strongly if the DGP is fat-tailed ($\nu_1 = 3$). Also for the thinner-tailed DGP ($\nu_1 = 10$), the fGAS performs better than the fGARCH for smaller sample sizes ($T = 500$). Even when the sample size is large ($T = 1500$) and $\nu_1 = 10$, the fGAS still performs better if there is a somewhat stronger correlation structure ($\nu_2 = 0.1$). Only when the DGP is thinner-tailed ($\nu_1 = 10$), the sample size is large ($T = 1500$), and there is hardly any cross-sectional correlation ($\nu_2 = 0.01$),

**Table 1:** Simulated out-of-sample MAE $(NT_{\text{out}})^{-1}\sum_{j=1}^{N}\sum_{i=1}^{T_{\text{out}}}\left|\widehat{\sigma}_{T+i}(t_j)-\sigma_{T+i}(t_j)\right|$ with $T_{\text{out}}=1000$ for time-varying volatility curves

| | | | $N=25$ | | | $N=100$ | | |
|---|---|---|---|---|---|---|---|---|
| $T$ | $\nu_1$ | $\nu_2$ | fGAS($\hat{\nu}_1$) | fGAS($\nu_1$) | fGARCH | fGAS($\hat{\nu}_1$) | fGAS($\nu_1$) | fGARCH |
| 500 | 3 | 0.1 | 0.525 | 0.519 | 1.040 | 0.542 | 0.521 | 0.997 |
| | 3 | 0.01 | 0.528 | 0.522 | 1.035 | 0.571 | 0.559 | 1.062 |
| | 10 | 0.1 | 0.384 | 0.381 | 0.698 | 0.336 | 0.316 | 0.695 |
| | 10 | 0.01 | 0.370 | 0.364 | 0.622 | 0.332 | 0.307 | 0.518 |
| 1500 | 3 | 0.1 | 0.237 | 0.235 | 0.534 | 0.310 | 0.298 | 0.545 |
| | 3 | 0.01 | 0.241 | 0.239 | 0.488 | 0.315 | 0.305 | 0.493 |
| | 10 | 0.1 | 0.162 | 0.159 | 0.207 | 0.167 | 0.162 | 0.207 |
| | 10 | 0.01 | 0.157 | 0.158 | 0.177 | 0.168 | 0.163 | 0.160 |

then fGARCH behaves at par with the new fGAS model for a sparse grid size ($N=25$), or slightly better for a dense grid ($N=100$). For all other cases, the fGAS does better in terms of MAE, and often by a factor close to 2.

We also see that the fGAS as proposed in Section 2.4 recovers the underlying functional volatility dynamics better when $T$ increases, even when the dynamics are misspecified: the MAEs decrease consistently in $T$. As $N$ increases, we sometimes see a slight improvement in MAE, but overall the MAEs remain quite similar for sparse versus dense grids. It is also interesting to see that the estimation of $\nu_1$ has little impact on the consistency of the paths: the MAEs are not affected substantially by whether $\nu_1$ is estimated (fGAS($\hat{\nu}_1$) column) or fixed at its true value (fGAS($\nu_1$) column).

Figure 1 confirms these findings. The left panels show the true volatility pattern for $u_1=u_2=0$ and its fGAS($\hat{\nu}_1$) fit for one specific situation. We see that the fitted volatility pattern reflects the true dynamics very well. The middle panels show the average MAE for a fat-tailed DGP with stronger correlation structure as a function of the day $i$ (top-middle) or time $t$ within the day (bottom-middle). The curves clearly show that the fGAS($\hat{\nu}_1$) consistently outperforms the fGARCH benchmark across either $i$ or $t$. If the DPG is thinner tailed ($\nu_1=10$) with relatively weak dependence ($\nu_2=0.01$), the right-hand panels in the figure shows that the performance ranking of the two models continues, but that the differences are much less pronounced.

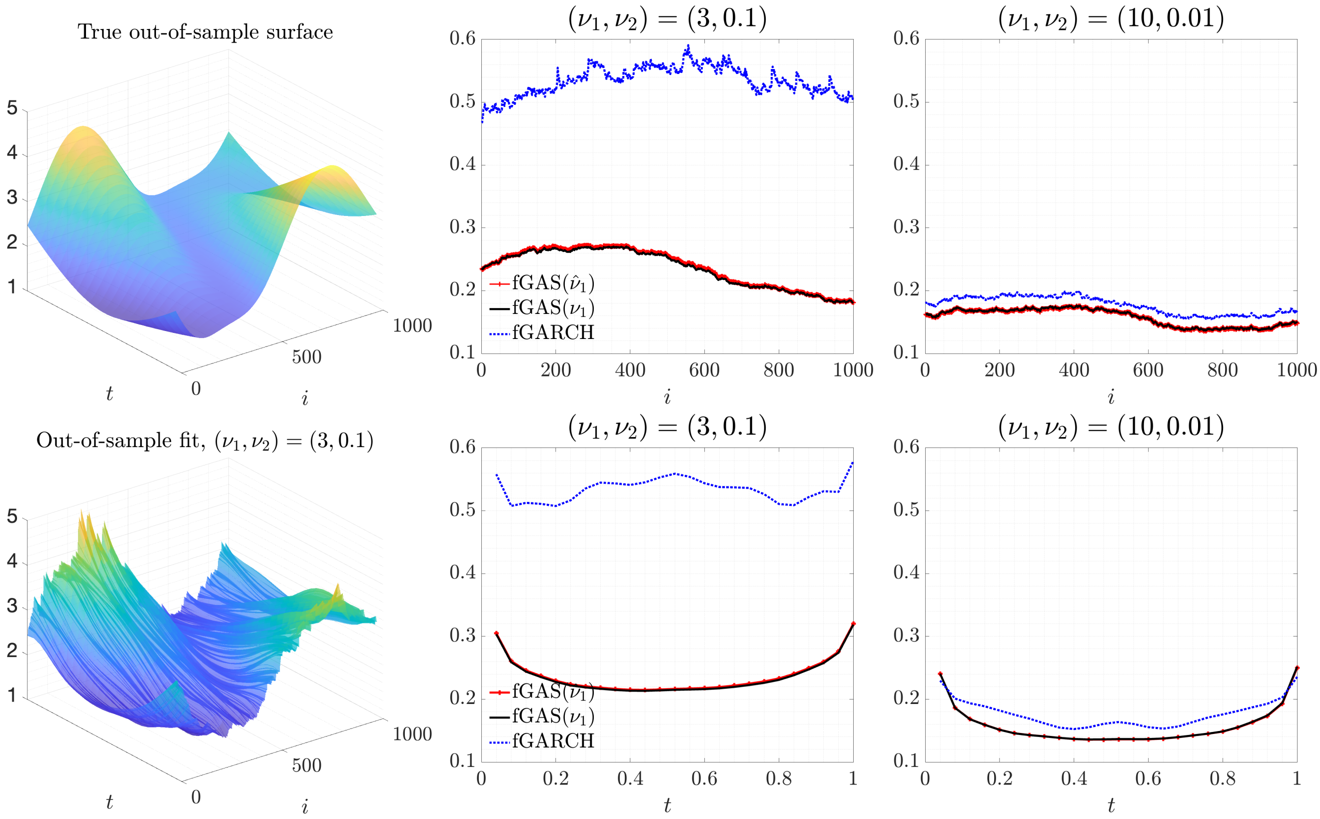


**Figure 1:** The top-left figure illustrates the shape of $\{\sigma_i(t), t \in [0, 1]\}$ for $i = T+1, \ldots, T+T_{\text{out}}$, where its square is defined in Eq. (4.1) with $u_1 = u_2 = 0$, for the example of time-varying volatility curves in Section 4.1. The bottom-left figure presents an example of an out-of-sample prediction using fGAS($\hat{\nu}_1$) given $(N, T, T_{\text{out}},, \nu_1, \nu_2) = (25, 1500, 1000, 3, 0.1)$, where $N$ is the grid size, $T$ the in-sample period, and $T_{\text{out}}$ the out-of-sample period. The middle and right panels plot the out-of-sample MAE across separate dimensions $i$ (top panel) and $t$ (bottom panel), computed as $N^{-1} \sum_{j=1}^{N} \left|\widehat{\sigma}_{T+i}(t_j) - \sigma_{T+i}(t_j)\right|$ with $i \in [\![ T_{\text{out}} ]\!]$ and $T_{\text{out}}^{-1} \sum_{i=1}^{T_{\text{out}}} \left|\widehat{\sigma}_{T+i}(t_j) - \sigma_{T+i}(t_j)\right|$ with $j \in [\![ N ]\!]$, for $(N, T, T_{\text{out}}) = (25, 1500, 1000)$.

Concluding, we see that the new fGAS approach works well in capturing the true volatility curves for the $d = 1$ dimensional case, even if the model is misspecified. This holds in particular for fat-tailed processes, where the robustness property of the fGAS plays out most. The estimated paths also become more accurate if the number of observed curves ($T$) increases, whereas the size of the observation grid ($N$) for each curve has less of an effect and the accuracy of the estimated paths.

## 4.2 Time-varying two-dimensional levels and irregular spacing

In this section, we examine the performance of the new fGAS model for $d = 2$, where we also add the complication of irregularly spaced observations, similar to one of our empirical illustrations in Section 5. We let $\boldsymbol{t} = (t_1, t_2)^\top \in [0, 1]^2$ and consider a time-varying location model

$Y_i(\boldsymbol{t}) = \mu_i(\boldsymbol{t}) + \sigma\varepsilon_i(\boldsymbol{t})$. We assume that the $\varepsilon_i(\boldsymbol{t}) \sim t_{\nu_1}$ are independent across $i \in \mathbb{Z}^+$, with the covariance kernel $\text{cov}(\varepsilon_i(\boldsymbol{t}), \varepsilon_i(\boldsymbol{s})) = \exp\big(-\|\boldsymbol{s}-\boldsymbol{t}\|/\nu_2\big)$. We set $\sigma = 1$ and

$$\mu_i(\boldsymbol{t}) = 2\left[\cos\left(\frac{i}{200} - u_1\right) + \sin\left(\frac{i \cdot t_1}{200} - u_2\right) + \cos(t_2 - u_3) + \sin(4t_1 - u_4)\cos(4t_2 - u_5)\right], \tag{4.2}$$

where the $u_j$ for $j = 1, \ldots, 5$ are independent uniform on $[0, 2\pi]$ and fixed in each Monte Carlo replication, inducing random phase shifts in the surface in the $i$ and $\boldsymbol{t}$ directions across simulations. The fGAS model is again clearly misspecified, but it turns out that it is still able to recover the dynamic surfaces $\mu_i(\boldsymbol{t})$ both across $i$ and $\boldsymbol{t}$. In each Monte Carlo iteration, the process $Y_i(\cdot)$ is observed at randomly generated locations $\boldsymbol{t}_j \sim \text{U}[0,1]^2$, for $j \in [\![ N ]\!]$. The observations are thus irregularly spaced, introducing a further complication for the model.

We set $g_\mu(f) = f$ and $g_\sigma(\cdot) \equiv \sigma$ and use $\boldsymbol{S}(\,\cdot\,,\,\cdot\,) \equiv \boldsymbol{I}_{n_\gamma}$. This yields the updating equation

$$\begin{aligned}\boldsymbol{\gamma}_{i+1} = \boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{\gamma}_i + \boldsymbol{A}\Bigg\{&\frac{\nu_1 + N}{\nu_1}\sigma^{-2}\Big(1 + \nu_1^{-1}\sigma^{-2}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\gamma}_i)\big)^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma}_i)\big)\Big)^{-1} \\ &\times \Big(\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_1), \ldots, \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_N)\Big)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma}_i)\big)\Bigg\},\end{aligned} \tag{4.3}$$

where $\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}) = \boldsymbol{\phi}_K(t_1) \otimes \boldsymbol{\phi}_K(t_2)$ contains all the products and cross-products of the basis functions for $K = 7$ as in Section 4.1. To limit the number of parameters in the simulation, we assume $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal. We again investigate the empirical MAE of the new fGAS($\hat{\nu}_1$) model. To the best of our knowledge, there is currently no directly comparable method available in the literature. We can nevertheless compare the results to the infeasible benchmark model fGAS($\nu_1$), where $\nu_1$ is fixed at the true value.

All simulation settings remain the same as in Section 4.1, except that we slightly increase $N$ using $N \in \{100, 200\}$. Table 2 presents the results. We again see that the estimation of $\nu_1$ has little effect on the performance of the method compared to the infeasible benchmark where $\nu_1$ is fixed at its true value. As the sample size $T$ increases, the MSEs go down, illustrating

**Table 2:** Full empirical out-of-sample MAE, computed as $(NT_{\text{out}})^{-1} \sum_{j=1}^{N} \sum_{i=1}^{T_{\text{out}}} \left|\widehat{\mu}_{T+i}(\boldsymbol{t}_j) - \mu_{T+i}(\boldsymbol{t}_j)\right|$ with $T_{\text{out}} = 1000$, for the example of time-varying surfaces (4.2) of Section 4.2.

| | | | $N = 100$ | | $N = 200$ | |
|---|---|---|---|---|---|---|
| $T$ | $\nu_1$ | $\nu_2$ | fGAS($\hat{\nu}_1$) | fGAS($\nu_1$) | fGAS($\hat{\nu}_1$) | fGAS($\nu_1$) |
| 500 | 3 | 0.1 | 0.605 | 0.582 | 0.612 | 0.598 |
| | 3 | 0.01 | 0.383 | 0.380 | 0.377 | 0.326 |
| | 10 | 0.1 | 0.621 | 0.539 | 0.637 | 0.590 |
| | 10 | 0.01 | 0.419 | 0.348 | 0.371 | 0.304 |
| 1500 | 3 | 0.1 | 0.364 | 0.352 | 0.405 | 0.378 |
| | 3 | 0.01 | 0.380 | 0.380 | 0.370 | 0.362 |
| | 10 | 0.1 | 0.432 | 0.374 | 0.472 | 0.353 |
| | 10 | 0.01 | 0.381 | 0.359 | 0.369 | 0.324 |

that the time-varying surfaces are estimated with increasing accuracy, even though the model is misspecified. Increasing the number $N$ of irregularly placed grid points again has a much smaller effect than increasing the number of curves $T$.

To further visualize the out-of-sample fit, we present the surface $\left\{\mu_i(\boldsymbol{t}),\, \boldsymbol{t} \in [0,1]^2\right\}$ and its estimated counterpart for $i = T, T+300, T+600, T+900$ in three-dimensional space, with the DGP parameters $(T, N, \nu_1, \nu_2) = (1500, 200, 3, 0.1)$; see Figure 2. The true surfaces provide a visualization of the DGP in Eq. (4.2).

To construct the estimated surfaces, we first obtain $\widehat{\boldsymbol{\theta}}_T$ using the first $T = 1500$ observed data points of $Y_i(\cdot)$ at $N = 200$ irregularly spaced locations. We subsequently obtain the estimated path $\left\{\widehat{\boldsymbol{\gamma}}_i(\widehat{\boldsymbol{\theta}}_T)\right\}$ using $\widehat{\boldsymbol{\theta}}_T$. Next, we randomly sample $N_{\text{plot}} = 12000$ discrete points (so many more than the $N = 200$ observed for estimation) from $\boldsymbol{s}_j \sim \text{U}[0,1]^2$, $j \in [\![ N_{\text{plot}} ]\!]$, and construct the estimated surface as $\widehat{\mu}_i(\boldsymbol{s}_j) = \boldsymbol{\phi}_{\boldsymbol{K}}^{\top}(\boldsymbol{s}_j)\widehat{\boldsymbol{\gamma}}_i(\widehat{\boldsymbol{\theta}}_T)$, following Eq. (2.9). The estimated surfaces, shown as blue dots in Figure 2, exhibit strong out-of-sample predictive performance in both level and shape dynamics. Even near the boundaries, the fit appears quite good. This is encouraging, as a good fit near the edges is typically challenging if data are sparse and is a well-documented issue in the nonparametric literature.

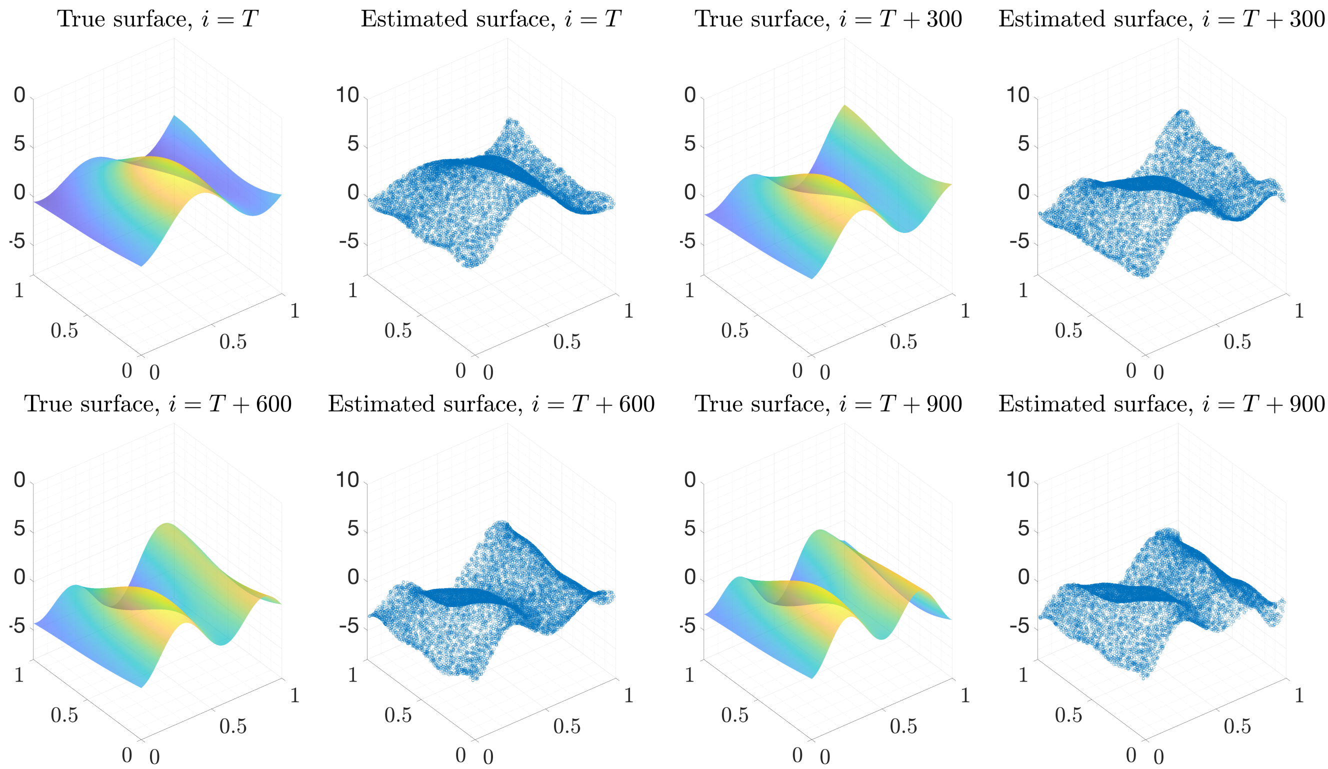


**Figure 2:** An example of (an out-of-sample) prediction using fGAS($\hat{\nu}_1$) for the surface $\{\mu_i(\boldsymbol{t}),\ \boldsymbol{t} \in [0,1]^2\}$, as defined in (4.2), in three-dimensional space, where $i = T, T+300, T+600, T+900$, with parameters $(T, N, \nu_1, \nu_2) = (1500, 200, 3, 0.1)$.

# 5 Empirical applications

We also illustrate the new fGAS method in two real-world examples. The first example examines stock volatility paths ($d = 1$), while the second explores air quality across the European continent ($d = 2$), measured by $PM_{2.5}$.

## 5.1 Volatility curves of stock returns

We collect minute-by-minute data from the TAQ database for a liquid stock, Pfizer (PFE). The dataset spans the period from 2 January 2015 to 29 December 2023, covering the volatile COVID-19 period, which is particularly relevant for a pharmaceutical company like Pfizer given its involvement in the development of one of the COVID-19 vaccines. We use the trading hours from 10 AM to 4 PM, resulting in 360 observations per day. We exclude days with more than 10 missing values, resulting in $T = 2247$. We impute the remaining missing values by averaging the two prices closest in time on both sides of the missing data point.

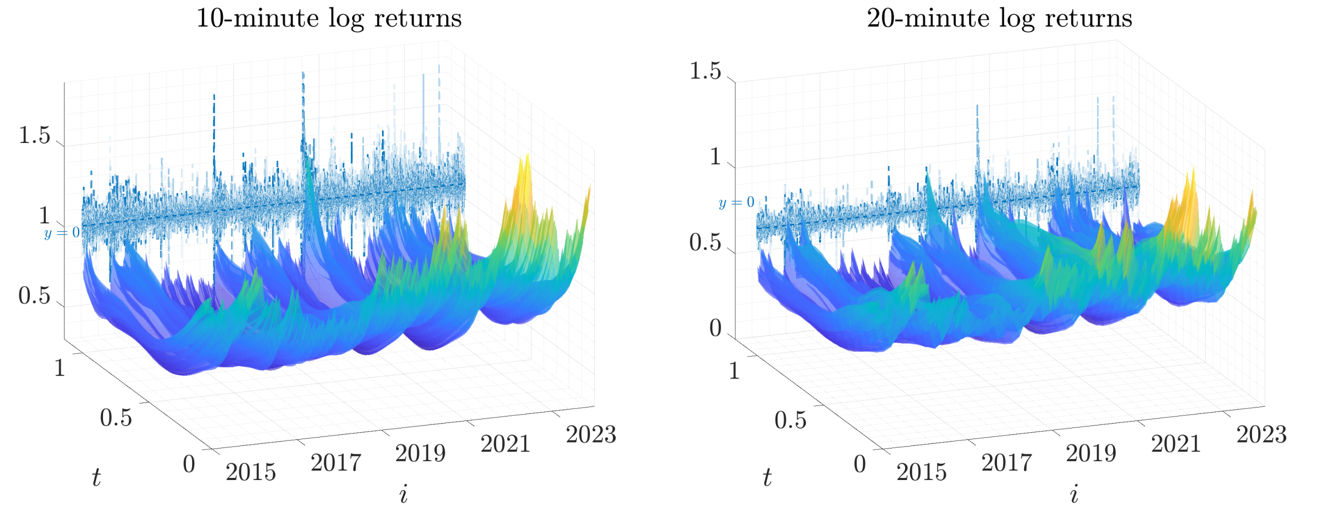


**Figure 3:** Fitted intraday volatility curves by fGAS($\hat{\nu}_1$) from 2 January 2015 up to 29 December 2023: The fitted volatility levels are annualized using $6 \cdot 6 \cdot \sqrt{252}$ for 10-minute intraday log returns and $6 \cdot 3 \cdot \sqrt{252}$ for 20-minute intraday log returns, ensuring they are displayed on a comparable scale. The left figure shows the fitted curves for 10-minute log returns, while the right figure represents the curves for 20-minute log returns. The blue dashed lines in the background represent the respective intraday log returns, scaled by a factor of 36 for the 10-minute returns and 18 for the 20-minute returns to convert them into daily returns for visualization. Each curve corresponds to the intraday log returns for a specific trading time.

Let $P_i(t)$ denote the price at time $t$ on day $i$. We define the $\Delta$-minute resolution intraday log returns as $Y_i(t) = \log P_i(t) - \log P_i(t - \Delta)$. Due to the presence of microstructure noise, it is common practice to set $\Delta \geq 10$ (see, e.g., Cerovecki et al., 2019). For illustration, we consider $\Delta \in \{10, 20\}$ minutes, resulting in $N \in \{35, 17\}$, respectively. Note that our method also performs well for higher resolutions such as $\Delta = 5$ minutes. As in Section 4.1, we again use unit scaling ($\boldsymbol{S}(\,\cdot\,,\,\cdot\,) \equiv \boldsymbol{I}_{n_\gamma}$), leading to the update equation in (3.3).

The fitted curves $\widehat{\sigma}_i(\cdot)$ for $i \in [\![\, T \,]\!]$ obtained using the new fGAS($\hat{\nu}_1$) model are displayed in Figure 3. We normalized time such that 10 AM corresponds to $t = 0$ and 4 PM to $t = 1$. The estimated volatilities exhibit a smile-like pattern, being higher at the beginning and end of the trading days compared to the middle. There is also clear evidence of time variation: both the levels and shapes of $\widehat{\sigma}_i(\cdot)$ change across $i$. Volatilities are generally highest during the outbreak of the COVID-19 pandemic. Finally, we observe mildly heavy-tailed behavior, with $\hat{\nu}_1 \approx 6.460$ for 10-minute log returns and $\hat{\nu}_1 \approx 6.905$ for 20-minute log returns.

Since the data exhibit moderately heavy tails, the existing fGARCH approach by Cerovecki et al.

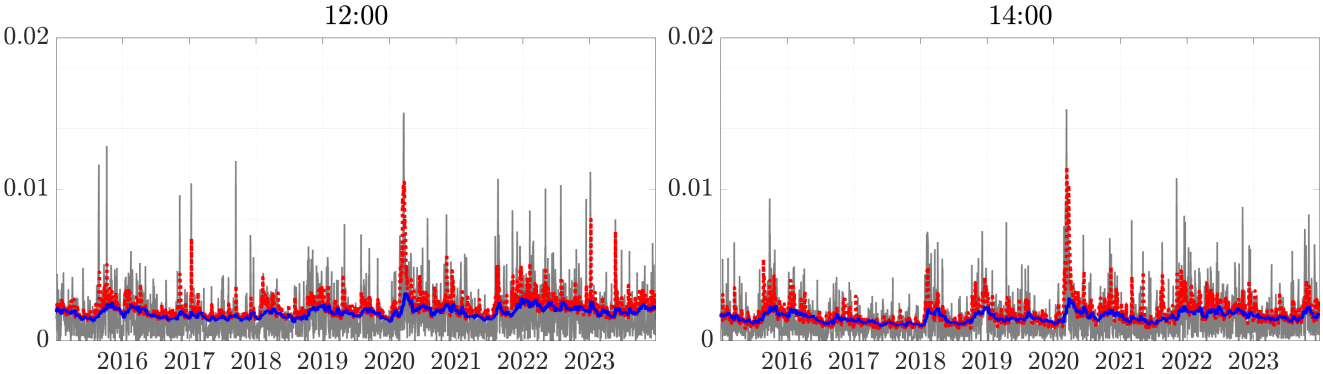


**Figure 4:** Fitted (nonannualized) volatility paths at 12 noon and 14:00 using fGAS (blue solid curves) and fGARCH (red dotted curves) from January 2, 2015 to December 29, 2023, based on 20-minute intraday log returns. Gray solid lines represent the absolute values of 20-minute intraday log returns.

(2019) can be substantially influenced by large shocks, making it less stable both numerically and theoretically. Figure 4 plots the filtered volatilities $\widehat{\sigma}_i(t)$ for both the fGAS($\hat{\nu}_1$) and the fGARCH models across days $i$ for two fixed moments of the day, namely 12 noon and 14:00. The figure is based on 20-minute intraday log returns. Both models produce similar secular patterns for the volatility dynamics, with particularly higher volatilities during the initial phase of the COVID-19 period in 2020 with its lockdowns and race for vaccine development. However, we also clearly see that the fGARCH estimated volatilities exhibit many more incidental sharp increases followed by rapid gradual declines than its fGAS counterpart, not only around the COVID-19 lockdowns of March 2020, but also in 2017, 2023, and elsewhere. By contrast, the filtered volatility pattern of our fGAS approach behaves much more gradually over time. Online Appendix G presents the fGARCH volatilities across all times of the day and further illustrates that the fGARCH volatilities are much more erratic and unstable, both at 10-minute and 20-minute frequencies, quite unlike the pattern in Figure 3.

## 5.2 $PM_{2.5}$ concentration

Fine particulate matter with an aerodynamic diameter of 2.5 $\mu$m or less, i.e., $PM_{2.5}$, has been widely associated with increased premature mortality (Zhang et al., 2017) and is therefore used by governments as a key air quality indicator. This example showcases how our method can be applied to estimate emission levels. We collect daily average $PM_{2.5}$ data from $N = 109$ sensors

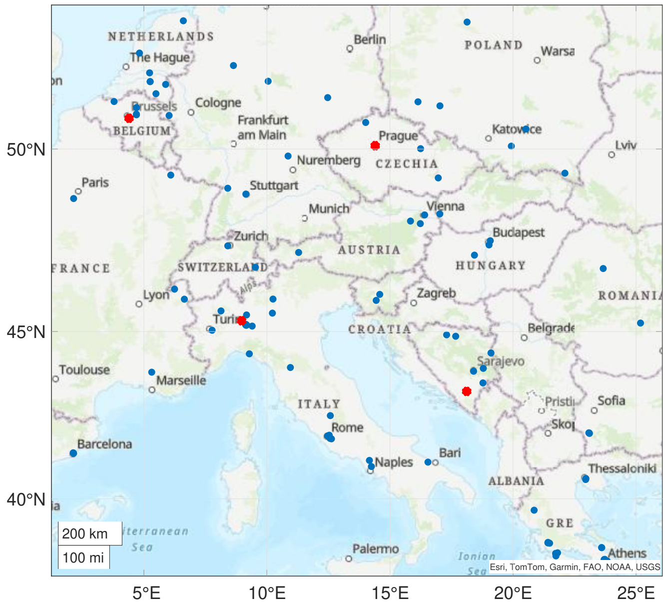


**Figure 5:** The locations of 109 sensors across Europe: The blue dots represent the specific sensor locations, while the red dots, located at different longitudes and latitudes, correspond to Brussels (Belgium), Prague (Czech Republic), Milan (Italy), Sarajevo (Bosnia and Herzegovina).

distributed across Europe, spanning from western Europe (France) to eastern Europe (Romania), over a period of $T = 700$ days, from 29 March 2023 to 25 February 2025. The data is obtained from PurpleAir (`https://www2.purpleair.com/`), a widely used citizen science network for real-time air quality monitoring that has been adopted in many recent atmospheric studies; see Bonas and Castruccio (2025) and references therein. The dots in Figure 5 represent the locations of these sensors, with their specific indices provided in Table G.1 in the Online Appendix. For missing data during the selected period, we again use imputation by averaging the two closest observations in the time dimension on both sides of the missing data point. For more advanced imputation methods, one may refer to, for instance, Cahan et al. (2023).

The sensors are sometimes subject to significant measurement errors, depending on their locations (e.g., indoors or outdoors), as observed in the time series plot in Figure G3 in the Online Appendix. It is therefore important to employ an outlier-robust method for analyzing this dataset. We apply our proposed method (fGAS($\hat{\nu}_1$)) using the updating scheme outlined in Section 4.2, which automatically downweights incidental large observations. For simplicity, we set $\boldsymbol{\Lambda}(\cdot) \equiv \boldsymbol{I}_N$, though one can of course further extend the model with an appropriate spatial correlation structure.

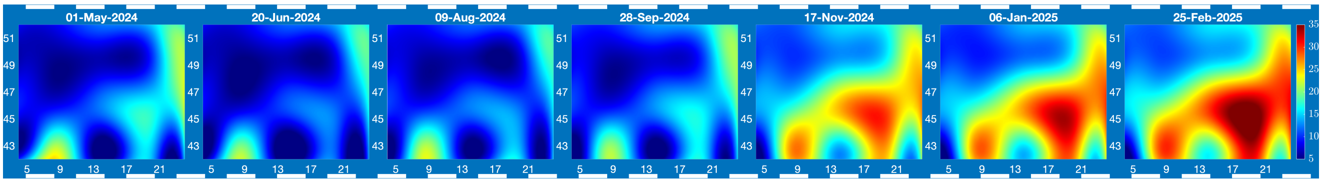


**Figure 6:** A filmstrip plot of the predicted $PM_{2.5}$ concentration levels, ranging from (4°E, 42°N) in the bottom-left to (24°E, 52°N) in the top-right, over the most recent 350 days, specifically at times $i = T - 300, T - 250, \ldots, T - 50, T$, where $T$ corresponds to Feb 25, 2025.

The fitted concentration levels are shown in Figure 6, where we use a time series of heatmaps (referred to as a filmstrip plot) for easy visualization.[3] Since nonparametric methods are typically subject to approximation bias near the boundaries of the map, as mentioned earlier in Section 4.2, we consider the fitted levels only within the region from (4°E, 42°N) to (24°E, 52°N); see Figure 5.

Figure 6 reveals that there is clear variation across different days of the year and different regions. $PM_{2.5}$ emissions are higher in central and eastern Europe compared to northwestern Europe, particularly during the winter months.

Next, we zoom in and focus on four specific sensors over time. These sensors represent locations with different longitudes and latitudes, as indicated by the red dots in Figure 5, and are located near Brussels (Belgium), Prague (Czech Republic), Milan (Italy), and Sarajevo (Bosnia and Herzegovina), respectively. Figure 7 displays the results. The first thing to notice is that the data are very noisy, such that it is non-trivial to filter out the mean surface dynamics across continental Europe. Second, the problem is even more complicated due to the spatial dimension: the fGAS need not only fit the time series dynamics per sensor, but also the pollution surface across space. It may therefore be that the model needs to sacrifice some of the temporal fit accuracy to have a better spatial fit given the chosen basis structure. Still, we observe that the predicted values $\widehat{Y}_{\text{fGAS}}$ capture the overall trend movements well. The model slightly underfits the spikes around January 2024 in Milan and Sarajevo. This is understandable given the second challenge above: the fGAS does not only need to fit the time-series dynamics, but it also needs to fit the surrounding sensors,

[3]The filmstrip plot is generated in a similar manner to Figure 2. Specifically, after estimating the static parameters, we randomly sample a dense set of locations ($N_{\text{plot}} = 15,000$, so much larger than the number of $N = 109$ sensors used for estimation), resulting in a predicted surface $\widehat{\mu}_i(\cdot)$ for $i \in [\![T]\!]$ at these sampled points. To construct the final filmstrip, we use linear interpolation to approximate the surface at unsampled locations.

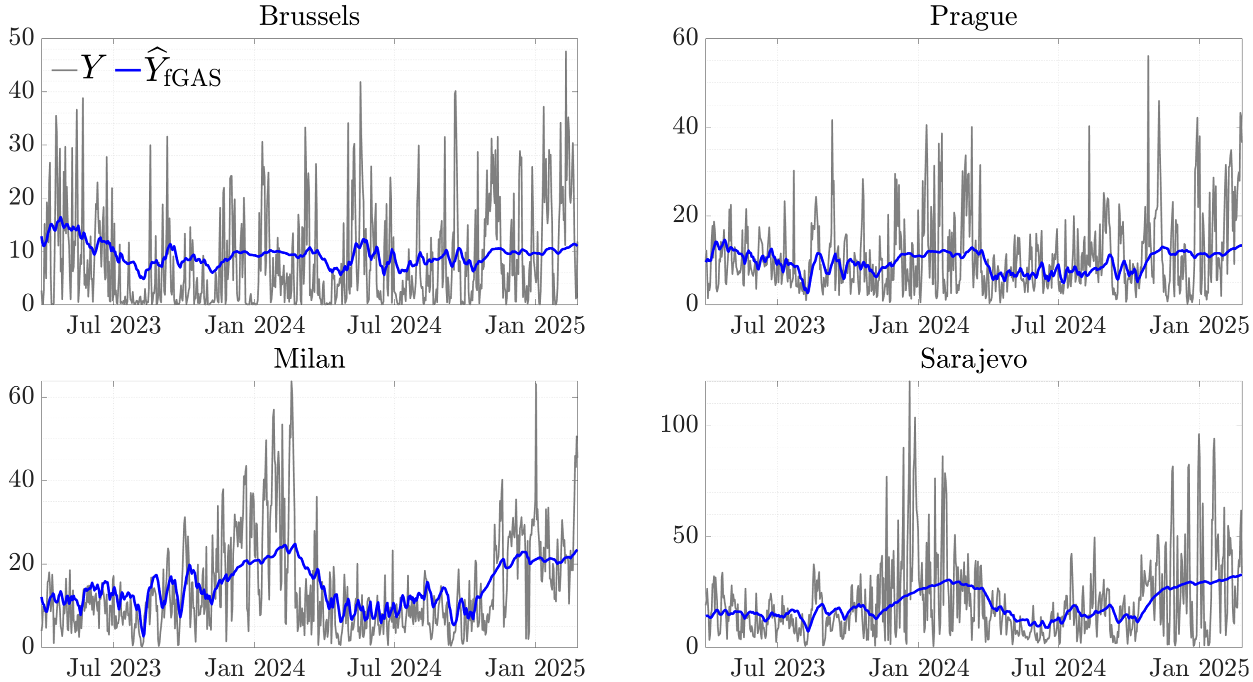


**Figure 7:** Predicted $PM_{2.5}$ concentration levels for Athens (Greece), Milan (Italy), Prague (Czech Republic), and Utrecht (the Netherlands), using the proposed functional approach (denoted as $\widehat{Y}_{\text{fGAS}}$, shown as blue solid curves). The real data is denoted as $Y$ and shown in gray.

making it possibly less reactive to fluctuations in a few nearby sensors.

# 6 Conclusion

In this paper we proposed a new class of location-scale models for sparsely observed functional data where the functions are defined on spaces of fixed, arbitrary dimension, and where we allow the location and scale functional parameters to exhibit latent time-varying dynamics. We showed that this new model performs well in both simulations and empirical applications, even in challenging situations where the functional observations are only observed on a limited set of grid points with considerable measurement noise.

The key of the new method was to project the functional location and scale onto a finite set of arbitrary basis functions and to let the projection coefficients vary over time using score-driven dynamics. The combination of an objective function based on fat-tailed Student's $t$ processes and the functional score-driven updates yielded less impact of incidental outliers on both parameter estimation and the functional dynamics themselves: outlying functional observations automatically received less weight in the updating mechanism in a data-driven way.

We also derived the asymptotic properties of the resulting estimator, establishing strong consistency as well as asymptotic normality of the static parameters governing the functional dynamics, thereby showing that the new approach allows us to recover the latent parameter variation from the observed data. The current method therefore provides a substantial step forward in modeling location-scale dynamics for functional objects $Y_i(\boldsymbol{t})$, with $\boldsymbol{t} \in \mathbb{R}^d$ for general $d \in \mathbb{Z}^+$, and can serve as a new benchmark model for further developments.

## Acknowledgment

We thank Daan Zunnenberg and Mingxuan Song for their excellent research assistance.

## References

Aue, A., L. Horváth, and D. F. Pellatt (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis 38*(1), 3–21.

Aue, A., D. D. Norinho, and S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association 110*(509), 378–392.

Bardsley, P., L. Horváth, P. Kokoszka, and G. Young (2017). Change point tests in functional factor models with application to yield curves. *Econometrics Journal 20*(1), 86–117.

Bastian, P. and H. Dette (2025). Gradual changes in functional time series. *Journal of Time Series Analysis*. In press.

Berild, M. O. and G.-A. Fuglstad (2023). Spatially varying anisotropy for Gaussian random fields in three-dimensional space. *Spatial Statistics 55*, 100750.

Beutner, E., Y. Lin, and S. Smeekes (2023). GLS estimation and confidence sets for the date of a single break in models with trends. *Econometric Reviews 42*(2), 195–219.

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

Bonas, M. and S. Castruccio (2025). High-resolution urban air quality monitoring from citizen science data with echo-state transformer networks. *Journal of the Royal Statistical Society Series C: Applied Statistics*, qlaf007.

Boniece, B. C., L. Horváth, and L. Trapani (2023). On changepoint detection in functional data using empirical energy distance. *arXiv preprint arXiv:2310.04853*.

Cahan, E., J. Bai, and S. Ng (2023). Factor-based imputation of missing values and covariances in panel data of large dimensions. *Journal of Econometrics 233*(1), 113–131.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics 136*(1), 163–188.

Cerovecki, C., C. Francq, S. Hörmann, and J.-M. Zakoïan (2019). Functional GARCH models: The quasi-likelihood approach and its applications. *Journal of Econometrics 209*(2), 353–375.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics 8*(2), 93–115.

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.

Cressie, N. (1993). *Statistics for Spatial Data* (Rev. ed.). New York: Wiley.

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

D'Innocenzo, E., A. Luati, and M. Mazzocchi (2023). A robust score-driven filter for multivariate time series. *Econometric Reviews 42*(5), 441–470.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods* (2nd ed.). Oxford University Press (Oxford).

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.

Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory 2*(3), 155–239.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

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

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

Hays, S., H. Shen, and J. Z. Huang (2012). Functional dynamic factor models with application to yield curve forecasting. *Annals of Applied Statistics 6*(3), 870–894.

Heyde, C. C. and N. N. Leonenko (2005). Student processes. *Advances in Applied Probability 37*(2), 342–365.

Hörmann, S., L. Horváth, and R. Reeder (2013). A functional version of the ARCH model. *Econometric Theory 29*(2), 267–288.

Hörmann, S., Ł. Kidziński, and M. Hallin (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society Series B: Statistical Methodology 77*(2), 319–348.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *Annals of Statistics 38*(3), 1845–1884.

Horváth, L., P. Kokoszka, and G. Rice (2014). Testing stationarity of functional time series. *Journal of Econometrics 179*(1), 66–82.

Hyndman, R. J. and H. Booth (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting 24*(3), 323–342.

Hyndman, R. J. and M. Shahid Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis 51*(10), 4942–4956.

Li, D., R. Li, and H. L. Shang (2024). Detection and estimation of structural breaks in high-dimensional functional time series. *Annals of Statistics 52*(4), 1716–1740.

Li, X., L. Wang, and H. J. Wang (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association 116*(536), 1994–2008.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics 45*(1-2), 7–38.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Paparoditis, E. and H. L. Shang (2023). Bootstrap prediction bands for functional time series. *Journal of the American Statistical Association 118*(542), 972–986.

Paul, D. and J. Peng (2009). Consistency of restricted maximum likelihood estimators of principal components. *Annals of Statistics 37*(3), 1229 – 1271.

Peng, J. and D. Paul (2009). A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics 18*(4), 995–1015.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika 86*, 677–690.

Ramsay, J. O., G. Hooker, and S. Graves (2009). *Functional Data Analysis with R and MATLAB*. New York: Springer-Verlag.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. New York: Springer-Verlag.

Rasmussen, C. E. and C. K. Williams (2006). *Gaussian Processes for Machine Learning*. MA: MIT press.

Rice, G., T. Wirjanto, and Y. Zhao (2020). Tests for conditional heteroscedasticity of functional data. *Journal of Time Series Analysis 41*(6), 733–758.

Shah, A., A. Wilson, and Z. Ghahramani (2014). Student-$t$ processes as alternatives to Gaussian processes. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, Volume 33 of *Proceedings of Machine Learning Research*, pp. 877–885.

Shang, H. L. and R. J. Hyndman (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics 26*(2), 330–343.

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), 2449–2495.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and Its Application 3*(1), 257–295.

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

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

Zhang, H. and Y. Li (2022). Unified principal component analysis for sparse and dense functional data under spatial dependency. *Journal of Business & Economic Statistics 40*(4), 1523–1537.

Zhang, Q., X. Jiang, D. Tong, S. J. Davis, H. Zhao, G. Geng, T. Feng, B. Zheng, Z. Lu, D. G. Streets, et al. (2017). Transboundary health impacts of transported global air pollution and international trade. *Nature 543*(7647), 705–709.

Zhang, X. and J.-L. Wang (2016). From sparse to dense functional data and beyond. *Annals of Statistics 44*(5), 2281–2321.

Zhu, C. and J.-L. Wang (2023). Testing homogeneity: the trouble with sparse functional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(3), 705–731.

Online Appendix to:

# Functional Location-Scale Models with Robust Observation-Driven Dynamics

*Yicong Lin*[1] *and Andre Lucas*[1]

[1] *Vrije Universiteit Amsterdam* and *Tinbergen Institute*

## Contents

# A Proof of Proposition 1 (score-driven dynamics)

Recall that $\boldsymbol{\gamma}_i = \text{vec}(\boldsymbol{\Gamma}_i^\top)$. Note that

$$\begin{aligned}
\frac{\partial \boldsymbol{\Gamma}_i \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})}{\partial \boldsymbol{\gamma}_i^\top} &= \boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top, \\
\frac{\partial \mu_i(\boldsymbol{t})}{\partial \boldsymbol{\gamma}_i} = \frac{\partial g_\mu\big(\boldsymbol{\Gamma}_i \, \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)}{\partial \boldsymbol{\gamma}_i} &= \left(\frac{\partial \boldsymbol{\Gamma}_i \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})}{\partial \boldsymbol{\gamma}_i^\top}\right)^\top \dot{\boldsymbol{g}}_\mu\big(\boldsymbol{f}_i(\boldsymbol{t})\big) = \dot{\boldsymbol{g}}_\mu\big(\boldsymbol{f}_i(\boldsymbol{t})\big) \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}), \\
\frac{\partial \sigma_i(\boldsymbol{t})}{\partial \boldsymbol{\gamma}_i} = \frac{\partial g_\sigma\big(\boldsymbol{\Gamma}_i \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)}{\partial \boldsymbol{\gamma}_i} &= \left(\frac{\partial \boldsymbol{\Gamma}_i \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})}{\partial \boldsymbol{\gamma}_i^\top}\right)^\top \dot{\boldsymbol{g}}_\sigma\big(\boldsymbol{f}_i(\boldsymbol{t})\big) = \dot{\boldsymbol{g}}_\sigma\big(\boldsymbol{f}_i(\boldsymbol{t})\big) \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}).
\end{aligned}$$

We then have

$$\begin{aligned}
\frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}_i} &= \frac{\partial \boldsymbol{\mu}(\boldsymbol{\gamma}_i)^\top}{\partial \boldsymbol{\gamma}_i} \frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \boldsymbol{\mu}(\boldsymbol{\gamma}_i)} + \frac{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)^\top}{\partial \boldsymbol{\gamma}_i} \frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)} \\
&= \dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma}_i)^\top \frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \boldsymbol{\mu}(\boldsymbol{\gamma}_i)} + \dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma}_i)^\top \frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)},
\end{aligned}$$

where $\operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)$ is a column vector that stacks the diagonal elements of $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)$. Define $\boldsymbol{V}(\boldsymbol{\gamma}_i) = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)$. We then also have

$$\begin{aligned}
\frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \boldsymbol{\mu}(\boldsymbol{\gamma}_i)} &= w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \; \boldsymbol{V}(\boldsymbol{\gamma}_i)^{-1}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma}_i)\big) = w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \; \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i), \\
\frac{\partial \log p\big(\boldsymbol{Y}_i \mid \boldsymbol{\gamma}_i; \boldsymbol{\nu}\big)}{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)} &= -\frac{\partial \log \big|\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big|}{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)} - \frac{1}{2} w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \; \frac{\partial \boldsymbol{e}_i(\boldsymbol{\gamma}_i)^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2) \boldsymbol{e}_i(\boldsymbol{\gamma}_i)}{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)} \\
&= -\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\iota}_N - w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \; \frac{\partial \boldsymbol{e}_i(\boldsymbol{\gamma}_i)^\top}{\partial \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)\big)} \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i) \\
&= w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \;\; \operatorname{diag}\big(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i)\big) \, \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i) - \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \boldsymbol{\iota}_N \\
&= \boldsymbol{\Sigma}(\boldsymbol{\gamma}_i)^{-1} \left( w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \; \big( (\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}_i)) \odot \boldsymbol{e}_i(\boldsymbol{\gamma}_i) \big) - \boldsymbol{\iota}_N \right).
\end{aligned}$$

This completes the proof. □

# B Proofs of Proposition 2 and Proposition 3

For $\boldsymbol{f} = \boldsymbol{\Gamma}\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})$ and $\boldsymbol{\gamma} = \text{vec}(\boldsymbol{\Gamma}^\top)$, define

$$\boldsymbol{\Phi}^\mu(\boldsymbol{t};\boldsymbol{\gamma}) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\dot{\boldsymbol{g}}_\mu(\boldsymbol{f}) \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)\,\dot{\boldsymbol{g}}_\mu(\boldsymbol{f}),$$
$$\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma}) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}) \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)\,\dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}).$$

Following Bougerol (1993, Section 3), let $(E, d_E)$ denote a complete separable metric space and define the Lipschitz coefficient $\rho$ for a random map $\lambda : E \to E$ as follows:

$$\rho(\lambda) = \sup_{x,y\in E,\, x\neq y}\left\{\frac{d_E\big(\lambda(x),\lambda(y)\big)}{d_E(x,y)}\right\}. \tag{B.1}$$

Note that if $\lambda$ is measurable, then $\rho(\lambda)$ is measurable (Bougerol, 1993, p. 955). In the subsequent proofs, we will frequently refer to the definition above, specifying the relevant space in each context. Finally, we use $C$ throughout the proofs to denote a generic positive constant that may vary from line to line.

***Proof of Proposition 2.*** We apply Theorem 3.1 from Bougerol (1993), considering the sequence $\{\widehat{\gamma}_i^\varepsilon,\, i \in \mathbb{Z}\}$ as a random process taking values in the separable Banach space $(E, d_E) = \big(\boldsymbol{\mathcal{G}}, \|\cdot\|\big)$; see Eq. (B.1). The SRE (3.2) can be equivalently written as $\widehat{\gamma}_{i+1}^\varepsilon = \boldsymbol{h}_i^\varepsilon\big(\widehat{\gamma}_i^\varepsilon\big)$, where

$$\begin{aligned}
\boldsymbol{h}_i^\varepsilon(\boldsymbol{\gamma}) &= \boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{A}\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\big(\nabla_i^{\mu,\varepsilon}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \nabla_i^{\sigma,\varepsilon}(\boldsymbol{\gamma},\boldsymbol{\nu})\big), \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \\
\nabla_i^{\mu,\varepsilon}(\boldsymbol{\gamma},\boldsymbol{\nu}) &= w_i^\varepsilon(\boldsymbol{\nu})\big(\boldsymbol{\Phi}^\mu(\boldsymbol{t}_1;\boldsymbol{\gamma}),\ldots,\boldsymbol{\Phi}^\mu(\boldsymbol{t}_N;\boldsymbol{\gamma})\big)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i, \\
\nabla_i^{\sigma,\varepsilon}(\boldsymbol{\gamma},\boldsymbol{\nu}) &= w_i^\varepsilon(\boldsymbol{\nu})\big(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_1;\boldsymbol{\gamma}),\ldots,\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_N;\boldsymbol{\gamma})\big)\,\text{diag}(\boldsymbol{\varepsilon}_i)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i - \sum_{j=1}^N \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma}),
\end{aligned} \tag{B.2}$$

with $w_i^\varepsilon(\boldsymbol{\nu}) = \frac{1+\nu_1^{-1}N}{1+\nu_1^{-1}\boldsymbol{\varepsilon}_i^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i}$. Since $\{\boldsymbol{\varepsilon}_i,\, i \in \mathbb{Z}\}$ is i.i.d. and thus SE, $\{\boldsymbol{h}_i^\varepsilon,\, i \in \mathbb{Z}\}$ is measurable and SE for any given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, see Krengel (1985, Proposition 4.3) or White (2001, Theorem 3.35). Following an approach similar to the proof of Proposition TA.1 in Blasques et al. (2022), the following (high-level, abbreviated as HL) conditions suffice to invoke Bougerol (1993, Theorem 3.1): For $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

HL1 $\mathbb{E}\Big(\log^+\big\|\boldsymbol{h}_1^\varepsilon(\widehat{\gamma}_0^\varepsilon) - \widehat{\gamma}_0^\varepsilon\big\|\Big) < \infty$ for some $\widehat{\gamma}_0^\varepsilon \in \boldsymbol{\mathcal{G}}$;

HL2 $\mathbb{E}\log\rho\big(\boldsymbol{h}_i^\varepsilon\big) < 0$.

**Verification of Condition HL1:** Note that for any matrices $\boldsymbol{X}_i$, $i \in [\![ K ]\!]$, where $K \in \mathbb{Z}^+$, with compatible dimensions, we have

$$\log^+ \left\| \prod_{i=1}^{K} \boldsymbol{X}_i \right\| \leq \log^+ \left( \prod_{i=1}^{K} \|\boldsymbol{X}_i\| \right) \leq \sum_{i=1}^{K} \log^+ \|\boldsymbol{X}_i\|, \tag{B.3}$$

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

By applying the inequalities (B.3) - (B.4) repeatedly, we obtain: for $N \in \mathbb{Z}^+$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\mathbb{E}\Big( \log^+ \big\|\boldsymbol{h}_1^{\varepsilon}(\widehat{\gamma}_0^{\varepsilon}) - \widehat{\gamma}_0^{\varepsilon}\big\|\Big) \leq \log(6) + \log^+ \|\boldsymbol{\omega}\| + \log^+ \big\|(\boldsymbol{B} - \boldsymbol{I}_{n_\gamma})\widehat{\gamma}_0^{\varepsilon}\big\| + \log^+ \|\boldsymbol{A}\|$$
$$+ \log^+ \big\|\boldsymbol{S}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\| + \mathbb{E}\Big( \log^+ \big\|\nabla_i^{\mu,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\|\Big) + \mathbb{E}\Big( \log^+ \big\|\nabla_i^{\sigma,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\|\Big). \tag{B.5}$$

Then, by Assumption SE1, it suffices to prove that the last two moments in (B.5) exist for the verification of Condition HL1. Note that for any random matrix $\boldsymbol{Z}$, if there exists an $r > 0$ such that $\mathbb{E}\|\boldsymbol{Z}\|^r < \infty$, then $\mathbb{E}\big( \log^+ \|\boldsymbol{Z}\|\big) < \infty$. It thus suffices to show that (1) $\mathbb{E}\big\|\nabla_i^{\mu,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\| < \infty$ and (2) $\mathbb{E}\big\|\nabla_i^{\sigma,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\| < \infty$. Recall that $\boldsymbol{\varepsilon}_i \sim t_{\nu_1}\big(\boldsymbol{0}, \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)\big)$. For any $\boldsymbol{\nu} \in \boldsymbol{\Theta}_\nu$, $\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)$ can be decomposed as $\boldsymbol{\Lambda}(\boldsymbol{\nu}_2) = \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}$, where $\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}$ is positive definite and symmetric. Then, we obtain $\boldsymbol{u}_i := \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\varepsilon}_i \sim t_{\nu_1}(\boldsymbol{0}, \boldsymbol{I}_N)$. For the term $\nabla_i^{\mu,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})$, we thus have

$$\mathbb{E}\big\|\nabla_i^{\mu,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\| = \mathbb{E}\left\| w_i^{\varepsilon}(\boldsymbol{\nu}) \sum_{j=1}^{N} \boldsymbol{\Phi}^{\mu}(\boldsymbol{t}_j; \widehat{\gamma}_0^{\varepsilon}) \operatorname{row}_j \big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\big)\boldsymbol{u}_i \right\|$$
$$\leq \sum_{j=1}^{N} \Big\|\boldsymbol{\Phi}^{\mu}(\boldsymbol{t}_j; \widehat{\gamma}_0^{\varepsilon})\Big\| \, \mathbb{E}\Big| \operatorname{row}_j \big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\big) w_i^{\varepsilon}(\boldsymbol{\nu})\boldsymbol{u}_i \Big|. \tag{B.6}$$

Note that the term $w_i^{\varepsilon}(\boldsymbol{\nu})\boldsymbol{u}_i$ can be written as

$$\frac{\nu_1 + N}{\nu_1}\big(1 + \nu_1^{-1}\boldsymbol{\varepsilon}_i^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i\big)^{-1}\boldsymbol{u}_i = \frac{\nu_1 + N}{\sqrt{\nu_1}}\left(1 + \frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu_1}\right)^{-1} \frac{\boldsymbol{u}_i}{\sqrt{\nu_1}}. \tag{B.7}$$

Furthermore, $\mathbb{E}|Z| \leq \big[\mathbb{E}(Z^2)\big]^{1/2}$ for any random variable $Z$. For $N \in \mathbb{Z}^+$, we obtain

$$\mathbb{E}\Big| \operatorname{row}_j \big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\big) w_i^{\varepsilon}(\boldsymbol{\nu})\boldsymbol{u}_i \Big| \leq \frac{\nu_1 + N}{\sqrt{(\nu_1 + N)(\nu_1 + N + 2)}} \big\| \operatorname{row}_j \big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\big)\big\|, \tag{B.8}$$

using (F.2) in Appendix F. Applying Lemma F.5 and using Eqs. (B.6) and (B.8), we obtain $\mathbb{E}\big\|\nabla_i^{\mu,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})\big\| < \infty$.

Now, we consider the term $\nabla_i^{\sigma,\varepsilon}(\widehat{\gamma}_0^{\varepsilon}, \boldsymbol{\nu})$. Note that the $j_{\text{th}}$ element of $\boldsymbol{\varepsilon}_i$ can be expressed as $\operatorname{row}_j \big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\big)\boldsymbol{u}_i$. This leads to $\operatorname{diag}(\boldsymbol{\varepsilon}_i)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\varepsilon}_i = \operatorname{diag}(\boldsymbol{\varepsilon}_i)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{u}_i = (e_1, \ldots, e_N)^\top$, where

$e_j = \left[(\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right)\boldsymbol{u}_i\right]\left[\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)\boldsymbol{u}_i\right]$ for $j \in [\![ N ]\!]$. As a result, one can write:

$$\nabla_i^{\sigma,\varepsilon}(\widehat{\gamma}_0^\varepsilon, \boldsymbol{\nu}) = \sum_{j=1}^{N} \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j; \widehat{\gamma}_0^\varepsilon)\Big(\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right) w_i^\varepsilon(\boldsymbol{\nu})\boldsymbol{u}_i\boldsymbol{u}_i^\top \mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)^\top - 1\Big).$$

This leads to

$$\mathbb{E}\big\|\nabla_i^{\sigma,\varepsilon}(\widehat{\gamma}_0^\varepsilon, \boldsymbol{\nu})\big\| \leq \sum_{j=1}^{N} \Big\|\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j; \widehat{\gamma}_0^\varepsilon)\Big\|\Big\{\mathbb{E}\Big|\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right) w_i^\varepsilon(\boldsymbol{\nu})\boldsymbol{u}_i\boldsymbol{u}_i^\top \left[\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)\right]^\top\Big| + 1\Big\}. \tag{B.9}$$

Note that $w_i^\varepsilon(\boldsymbol{\nu})\boldsymbol{u}_i\boldsymbol{u}_i^\top = (\nu_1 + N)\left(1 + \boldsymbol{u}_i^\top\boldsymbol{u}_i/\nu_1\right)^{-1}\boldsymbol{u}_i\boldsymbol{u}_i^\top/\nu_1$. By (F.3) and Lemma F.5, we have

$$\mathbb{E}\Big|\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right) w_i^\varepsilon(\boldsymbol{\nu})\boldsymbol{u}_i\boldsymbol{u}_i^\top \mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)^\top\Big| \leq \|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\|_{2,\infty}^{1/2}\,\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\|_{2,\infty}^{1/2}. \tag{B.10}$$

Combining (B.9) - (B.10), and by applying Assumptions A1, A2, and Lemma F.5, we arrive at

$$\mathbb{E}\big\|\nabla_i^{\sigma,\varepsilon}(\widehat{\gamma}_0^\varepsilon, \boldsymbol{\nu})\big\| \leq \left(\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\|_{2,\infty}^{1/2}\,\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\|_{2,\infty}^{1/2} + 1\right)\left(\sum_{j=1}^{N}\Big\|\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j; \widehat{\gamma}_0^\varepsilon)\Big\|\right) < \infty. \tag{B.11}$$

Now we can conclude that for a fixed $N \in \mathbb{Z}^+$ and any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, Condition HL1 holds.

**Verification of Condition HL2:** By applying a version of the mean value theorem for vector-valued functions (see, e.g., Rudin, 1976, Theorem 9.19), we have

$$\big\|\boldsymbol{h}_i^\varepsilon(\gamma_1) - \boldsymbol{h}_i^\varepsilon(\gamma_2)\big\| \leq \sup_{\gamma\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{h}_i^\varepsilon(\gamma)\right\|\,\|\gamma_1 - \gamma_2\|. \tag{B.12}$$

By the definition of $\rho(\cdot)$ in (B.1), we have $\rho\big(\boldsymbol{h}_i^\varepsilon\big) \leq \sup_{\gamma\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{h}_i^\varepsilon(\gamma)\right\|$, where $\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\boldsymbol{h}_i^\varepsilon(\gamma) = \boldsymbol{B} + \boldsymbol{A}\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\big[\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\big(\nabla_i^{\mu,\varepsilon}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \nabla_i^{\sigma,\varepsilon}(\boldsymbol{\gamma},\boldsymbol{\nu})\big)\big]$. By Assumption SE2, Condition HL2 is thus fulfilled.

The existence of a unique SE solution to (3.2) now follows from Bougerol (1993, Theorem 3.1). □

Further notation is required. As in Straumann and Mikosch (2006, Section 2.3), let $\mathcal{C}^0\big(\boldsymbol{\Theta}, \boldsymbol{\mathcal{G}}\big)$ be the space of continuous $\boldsymbol{\mathcal{G}}$-valued functions equipped with the supremum norm $\|\cdot\|_{\boldsymbol{\Theta}}$ defined as $\|\gamma\|_{\boldsymbol{\Theta}} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\gamma(\boldsymbol{\theta})\|$ for $\gamma \in \mathcal{C}^0\big(\boldsymbol{\Theta}, \boldsymbol{\mathcal{G}}\big)$. Note that $\big(\mathcal{C}^0\big(\boldsymbol{\Theta}, \boldsymbol{\mathcal{G}}\big), \|\cdot\|_{\boldsymbol{\Theta}}\big)$ is a separable Banach (or Polish) space provided that $\boldsymbol{\Theta}$ is compact.

***Proof of Proposition 3.*** Recall $\rho$ from Eq. (B.1), along with the explicit expressions for $\nabla_i^{\mu}(\boldsymbol{\gamma},\boldsymbol{\nu})$ and $\nabla_i^{\sigma}(\boldsymbol{\gamma},\boldsymbol{\nu})$ given in (F.24) and (F.25), respectively. Unlike Proposition 2, this proposition requires the convergence to be uniform in $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. To employ Bougerol (1993, Theorem 3.1), we view $\{\widehat{\gamma}_i(\cdot),\, i\in\mathbb{Z}^+\}$, initialized at some $\widehat{\gamma}_0\in\boldsymbol{\mathcal{G}}$, as a sequence of random functions residing in the separable Banach space $(E,d_E)=\big(\mathcal{C}^0(\boldsymbol{\Theta},\boldsymbol{\mathcal{G}}),\|\cdot\|_{\boldsymbol{\Theta}}\big)$. Note that $\widehat{\gamma}_{i+1}$ can be expressed as $\widehat{\gamma}_{i+1}=\boldsymbol{\lambda}_i(\widehat{\gamma}_i)$, where the random maps $\boldsymbol{\lambda}_i:\mathcal{C}^0(\boldsymbol{\Theta},\boldsymbol{\mathcal{G}})\to\mathcal{C}^0(\boldsymbol{\Theta},\boldsymbol{\mathcal{G}})$ are given by $\big[\boldsymbol{\lambda}_i(\boldsymbol{\gamma})\big](\boldsymbol{\theta}):=\boldsymbol{\psi}_i\big(\boldsymbol{\gamma}(\boldsymbol{\theta}),\boldsymbol{\theta}\big)$ with $\boldsymbol{\psi}_i$ defined in Eq. (3.4). Since $\{\boldsymbol{Y}_i,\, i\in\mathbb{Z}\}$ is SE by assumption, it follows that $\{\boldsymbol{\lambda}_i,\, i\in\mathbb{Z}\}$ is also SE. Similar to the proof of Proposition TA.3 in Blasques et al. (2022), the following (high-level, abbreviated as $\widetilde{\text{HL}}$) conditions are sufficient to apply Bougerol (1993, Theorem 3.1) and Straumann and Mikosch (2006, Theorem 2.8):

$\widetilde{\text{HL}}1$ $\mathbb{E}\Big(\log^+\big\|\boldsymbol{\lambda}_1(\widehat{\gamma}_0)-\widehat{\gamma}_0\big\|_{\boldsymbol{\Theta}}\Big)<\infty$, where $\widehat{\gamma}_0(\boldsymbol{\theta})=\widehat{\gamma}_0\in\boldsymbol{\mathcal{G}}$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$;

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

$\widetilde{\text{HL}}3$ $\mathbb{E}\Big(\log\rho\big(\boldsymbol{\lambda}_i^{(r)}\big)\Big)<0$ for some integer $r\geq 1$, where $\boldsymbol{\lambda}_i^{(r)}=\boldsymbol{\lambda}_i\circ\boldsymbol{\lambda}_{i-1}\circ\cdots\circ\boldsymbol{\lambda}_{i-r+1}$ is referred to as the $r$-fold convolution of the function $\boldsymbol{\lambda}_i$.

**Verification of Condition $\widetilde{\text{HL}}1$:** Note that, by applying the inequalities (B.3) - (B.4), we have

$$\begin{aligned}\mathbb{E}\Big(\log^+\big\|\boldsymbol{\lambda}_1(\widehat{\gamma}_0)-\widehat{\gamma}_0\big\|_{\boldsymbol{\Theta}}\Big)&\leq\log(6)+\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\omega}\|+\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{B}-\boldsymbol{I}_{n_\gamma}\big\|\\&\quad+\log^+\big\|\widehat{\gamma}_0\big\|+\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{A}\|+\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{S}(\widehat{\gamma}_0,\boldsymbol{\nu})\big\|\\&\quad+\mathbb{E}\Big(\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\nabla_i^{\mu}(\widehat{\gamma}_0,\boldsymbol{\nu})\big\|\Big)+\mathbb{E}\Big(\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\nabla_i^{\sigma}(\widehat{\gamma}_0,\boldsymbol{\nu})\big\|\Big),\end{aligned}\tag{B.13}$$

where $\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\omega}\|+\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{B}-\boldsymbol{I}_{n_\gamma}\big\|+\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{A}\|<\infty$ as $\boldsymbol{\Theta}$ is compact. Moreover, $\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{S}(\widehat{\gamma}_0,\boldsymbol{\nu})\big\|$ is a direct result of Assumption IV1 in Proposition 3. It thus suffices to show that the final two quantities in (B.13) are finite.

Recall $\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})=\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\big(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\gamma})\big)$ from (F.7). For a fixed $N\in\mathbb{Z}^+$, we obtain

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\big\|\nabla_i^{\mu}(\boldsymbol{\gamma},\boldsymbol{\nu})\big\|\\&=\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\nu_1+N}{\sqrt{\nu_1}}\sum_{j=1}^{N}\boldsymbol{\Phi}^{\mu}(\boldsymbol{t}_j;\boldsymbol{\gamma})\,\mathrm{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)\frac{\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})/\sqrt{\nu_1}}{1+\nu_1^{-1}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})}\right\|\\&\leq C\left(\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu}\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\|_{2,\infty}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\sum_{j=1}^{N}\big\|\boldsymbol{\Phi}^{\mu}(\boldsymbol{t}_j;\boldsymbol{\gamma})\big\|\right)\leq C,\end{aligned}\tag{B.14}$$

following from Assumption A1 and Lemma F.5. Therefore, $\mathbb{E}\Big(\log^+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\nabla_i^{\mu}(\widehat{\gamma}_0,\boldsymbol{\nu})\big\|\Big)<\infty$.

Similarly, one obtains

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\nabla_i^{\sigma}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\Bigg\|(\nu_1+N)\sum_{j=1}^{N}\boldsymbol{\Phi}^{\sigma}(\boldsymbol{t}_j;\boldsymbol{\gamma})\Bigg(\operatorname{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right)$$
$$\times\frac{\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^{\top}/\nu_1}{1+\nu_1^{-1}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^{\top}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})}\left[\operatorname{row}_j\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)\right]^{\top}-1\Bigg)\Bigg\|$$
$$\leq C\Bigg(\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_{\nu}}\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\|_{2,\infty}\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_{\nu}}\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\|_{2,\infty}+1\Bigg)\Bigg(\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}N^{-1}\sum_{j=1}^{N}\left\|\boldsymbol{\Phi}^{\sigma}(\boldsymbol{t}_j;\boldsymbol{\gamma})\right\|\Bigg)\leq C, \quad \text{(B.15)}$$

using Assumption A1 and Lemma F.5. We thus conclude that $\mathbb{E}\Big(\log^{+}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\nabla_i^{\sigma}(\widehat{\boldsymbol{\gamma}}_0,\boldsymbol{\nu})\right\|\Big)<\infty$. This establishes Condition $\widetilde{\text{HL}}1$.

**Verification of Condition $\widetilde{\text{HL}}2$:** Note that

$$\begin{aligned}
\rho(\boldsymbol{\lambda}_1) &= \sup_{\|\gamma_1-\gamma_2\|_{\boldsymbol{\Theta}}>0}\frac{\|\boldsymbol{\lambda}_1(\gamma_1)-\boldsymbol{\lambda}_1(\gamma_2)\|_{\boldsymbol{\Theta}}}{\|\gamma_1-\gamma_2\|_{\boldsymbol{\Theta}}}\\
&= \sup_{\|\gamma_1-\gamma_2\|_{\boldsymbol{\Theta}}>0}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{\frac{\left\|\boldsymbol{\psi}_1(\gamma_1(\boldsymbol{\theta}),\boldsymbol{\theta})-\boldsymbol{\psi}_1(\gamma_2(\boldsymbol{\theta}),\boldsymbol{\theta})\right\|}{\|\gamma_1(\boldsymbol{\theta})-\gamma_2(\boldsymbol{\theta})\|}\frac{\|\gamma_1(\boldsymbol{\theta})-\gamma_2(\boldsymbol{\theta})\|}{\|\gamma_1-\gamma_2\|_{\boldsymbol{\Theta}}}\right\}\\
&\leq \sup_{\|\gamma_1-\gamma_2\|_{\boldsymbol{\Theta}}>0}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{\frac{\left\|\boldsymbol{\psi}_1(\gamma_1(\boldsymbol{\theta}),\boldsymbol{\theta})-\boldsymbol{\psi}_1(\gamma_2(\boldsymbol{\theta}),\boldsymbol{\theta})\right\|}{\|\gamma_1(\boldsymbol{\theta})-\gamma_2(\boldsymbol{\theta})\|}\right\}\\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\substack{\bar{\gamma}_1,\bar{\gamma}_2\in\boldsymbol{\mathcal{G}},\\ \|\bar{\gamma}_1-\bar{\gamma}_2\|>0}}\left\{\frac{\left\|\boldsymbol{\psi}_1(\bar{\gamma}_1,\boldsymbol{\theta})-\boldsymbol{\psi}_1(\bar{\gamma}_2,\boldsymbol{\theta})\right\|}{\|\bar{\gamma}_1-\bar{\gamma}_2\|}\right\}\\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\boldsymbol{B}+\boldsymbol{A}\frac{\partial}{\partial\boldsymbol{\gamma}^{\top}}\Big(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\nabla_1(\boldsymbol{\gamma},\boldsymbol{\nu})\Big)\right\|, \qquad \text{(B.16)}
\end{aligned}$$

where the final step follows from the mean value theorem given in Rudin (1976, Theorem 9.19) applied to the vector-valued function $\boldsymbol{\psi}_1(\,\cdot\,,\boldsymbol{\theta})$. By Assumption IV2, $\mathbb{E}\Big(\log^{+}\rho(\boldsymbol{\lambda}_1)\Big)<\infty$.

**Verification of Condition $\widetilde{\text{HL}}3$:** Analogous to (B.16), we can derive

$$\mathbb{E}\Big(\log\rho(\boldsymbol{\lambda}_i^{(r)})\Big)\leq\mathbb{E}\left(\log\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^{\top}}\boldsymbol{\psi}_i^{(r)}(\boldsymbol{\gamma},\boldsymbol{\theta})\right\|\right). \qquad \text{(B.17)}$$

By Assumption IV3, Condition $\widetilde{\text{HL}}3$ immediately follows.

Proposition 3 now follows by applying Bougerol (1993, Theorem 3.1) or Straumann and Mikosch (2006, Theorem 2.8). □

# C Proof of Theorem 1 (consistency)

The following lemma plays a key role in establishing the strong consistency of static parameter estimators.

**Lemma C.1.** *Recall* $\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = T^{-1}\sum_{i=1}^{T} \ell_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)$ *from* (2.16)*. Let* $\mathcal{L}_T(\boldsymbol{\theta}) = T^{-1}\sum_{i=1}^{T} \ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)$ *and* $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\big(\ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\big)$*. Under the assumptions of Proposition 3 and Assumption SC1, we have*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0, \qquad as \quad T \to \infty. \tag{C.1}$$

*Moreover,* $\mathcal{L}(\cdot)$ *is continuous on* $\boldsymbol{\Theta}$*.*

***Proof of Lemma C.1.*** Recall the notation $\nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \nabla_i^{\mu}(\boldsymbol{\gamma}, \boldsymbol{\nu}) + \nabla_i^{\sigma}(\boldsymbol{\gamma}, \boldsymbol{\nu})$, see also (F.24) and (F.25). Note that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big|$. We deal with each term separately.

**First term** $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$**:** By a mean value theorem of $\ell_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)$ around $\ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)$, we obtain

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| &\leq T^{-1}\sum_{i=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\ell_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big) - \ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\big| \\ &\leq T^{-1}\sum_{i=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial \ell_i\big(\boldsymbol{\gamma}, \boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}}\right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\|, \end{aligned} \tag{C.2}$$

where $\partial \ell_i\big(\boldsymbol{\gamma}, \boldsymbol{\nu}\big)/\partial\boldsymbol{\gamma} = N^{-1}\nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})$. By Assumptions A1 - A2, it follows from (B.14) and (B.15) that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial \ell_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\right\| \leq C$. Since $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\| \xrightarrow{e.a.s.} 0$ (Proposition 3), there exists $\rho > 1$ such that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\| \leq C\rho^{-i}$ a.s. for all $i \in \mathbb{Z}^+$. Therefore, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| \leq CT^{-1}\sum_{i=1}^{\infty} \rho^{-i} = C(\rho-1)^{-1}T^{-1} \to 0$ almost surely as $T \to \infty$.

**Second term** $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$**:** We can apply the uniform law of large numbers provided in White (1996, Theorem A.2.2) to the SE sequence $\big\{\ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big),\ i \in \mathbb{Z}\big\}$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ provided that

$$\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big|\ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\Big|\Big) < \infty \tag{C.3}$$

is met. Note that the limit criterion $\mathcal{L}(\boldsymbol{\theta})$ depends on the unique limiting SE process $\{\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\ i \in \mathbb{Z}\}$ instead of the initialized process $\big\{\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}, \widehat{\boldsymbol{\gamma}}_0),\ i \in \mathbb{Z}^+\big\}$, and is therefore independent of the initial value $\widehat{\gamma}_0$. Furthermore, recall $\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma}) = \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma})\big)$ from (F.7), as well as the representation $\boldsymbol{f}_i(\cdot) = \boldsymbol{\Gamma}_i\,\boldsymbol{\phi}_{\boldsymbol{K}}(\cdot) = \mathrm{vec}\left(\boldsymbol{\phi}_{\boldsymbol{K}}(\cdot)^\top\boldsymbol{\Gamma}_i^\top\right) = \left(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\cdot)^\top\right)\boldsymbol{\gamma}_i$. Given the compactness of $\boldsymbol{\Theta}$ and Assumption A2, for a fixed $N\in\mathbb{Z}^+$, it suffices for establishing (C.3) to show that the following moment conditions hold:

M1 $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big|N^{-1}\sum_{j=1}^N\log\Big(g_\sigma\big[\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)^\top\big)\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big]\Big)\Big|\Big)<\infty$;

M2 $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big|\log\Big(1+\nu_1^{-1}\boldsymbol{u}_i\big(\boldsymbol{\nu}_2,\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^\top\boldsymbol{u}_i\big(\boldsymbol{\nu}_2,\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)\Big)\Big|\Big)<\infty$.

First consider M1. If (3.5) holds for $\eta=0$, M1 trivially follows. Consider (3.5) holds for some $\eta>0$. Note that $\|\boldsymbol{A}\otimes\boldsymbol{B}\| = \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$ for any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. By applying Assumption SC1, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|N^{-1}\sum_{j=1}^N\log\Big(g_\sigma\big[\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)^\top\big)\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big]\Big)\right| \le C_\sigma\,N^{-1}\sum_{j=1}^N\|\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)\|^\eta\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\gamma}_i(\boldsymbol{\theta})\|^\eta\right). \tag{C.4}$$

By (B.14) - (B.15), and $\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu}\big\|\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\big\|<\infty$ (Assumption SC1), it follows that $\sup_{i\in\mathbb{Z}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})\big\|\le C$. From (2.9), we iterate backward $m\in\mathbb{Z}^+$ times to obtain

$$\begin{aligned}\sup_{i\in\mathbb{Z}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\gamma}_{i+1}(\boldsymbol{\theta})\big\| &= \sup_{i\in\mathbb{Z}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{B}^m\boldsymbol{\gamma}_{i-m+1}(\boldsymbol{\theta})+\sum_{k=0}^{m-1}\boldsymbol{B}^k\big(\boldsymbol{\omega}+\boldsymbol{A}\boldsymbol{s}_{i-k}\big(\boldsymbol{\gamma}_{i-k}(\boldsymbol{\theta}),\boldsymbol{\nu}\big)\big)\right\| \\ &\le \left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{B}\|^m\right)\sup_{i\in\mathbb{Z}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\gamma}_{i-m+1}(\boldsymbol{\theta})\big\| + C\sum_{k=0}^{m-1}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{B}\|^k\right).\end{aligned}$$

Then, $\sup_{i\in\mathbb{Z}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\gamma}_{i+1}(\boldsymbol{\theta})\big\|$ is bounded if $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{B}\|<1$, as required in Assumption SC1. Consequently, there exists a constant $C_\gamma\in\mathbb{R}^+$ such that

$$\sup_{i\in\mathbb{Z}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\|^\varrho \le C_\gamma, \qquad \forall\varrho\ge 0. \tag{C.5}$$

By (C.4), (C.5), and Assumption SC1, the condition M1 is satisfied for $\eta>0$ as well.

We move on to establish Condition M2. First, observe that, similar to (B.4), we have the following: $\Big|\log\Big(\sum_{i=1}^K X_i\Big)\Big|\le\log(K)+\sum_{i=1}^K\big|\log(X_i)\big|$, where $X_i>0$ for $i\in[\![\,K\,]\!]$, where $K\in\mathbb{Z}^+$. Second, for a symmetric matrix $\boldsymbol{A}$, the inequality $|\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}|\le\|\boldsymbol{A}\|\,\|\boldsymbol{x}\|^2$ holds for any compatible vector $\boldsymbol{x}$. Third, we have $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big|\log\big(2\nu_1^{-1}\big\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\big\|\big)\big|<\infty$ because $\nu_1$ is bounded below away from 0 (owing to the compactness of $\boldsymbol{\Theta}$), and given Assumption A2. By applying these results and the $c_r$-inequality, we obtain

$$\begin{aligned}
&\Big|\log\Big(1+\nu_1^{-1}\boldsymbol{u}_i\big(\boldsymbol{\nu}_2,\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^\top \boldsymbol{u}_i\big(\boldsymbol{\nu}_2,\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)\Big)\Big| \\
&\le \Bigg|\log\bigg[1+2\nu_1^{-1}\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\|\Big(\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\Big\|^2\Big\|\boldsymbol{\mu}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)-\boldsymbol{\mu}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)\Big\|^2 \\
&\qquad + \Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)\boldsymbol{\varepsilon}_i\Big\|^2\Big)\bigg]\Bigg| \\
&\le \Bigg|\log\bigg[1+2\nu_1^{-1}\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\|\Big(\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\Big\|^2\widetilde{C}_\mu^2+\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)\Big\|^2\Big(1+\|\boldsymbol{\varepsilon}_i\|^2\Big)\Big)\bigg]\Bigg| \\
&\le C+2\Big|\log\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\Big\|\Big|+2\Big|\log\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)\Big\|\Big|+\Big|\log\Big(1+\|\boldsymbol{\varepsilon}_i\|^2\Big)\Big|,
\end{aligned}$$

where $\widetilde{C}_\mu > 0$ in the second inequality is a constant independent of $\boldsymbol{\theta}$. This result follows from $\|\cdot\| \le \|\cdot\|_1$, Assumption SC1, and (C.5):

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|\boldsymbol{\mu}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)-\boldsymbol{\mu}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)\Big\| \\
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sum_{j=1}^{N}\Big|g_\mu\big[\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)^\top\big)\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big]-g_\mu\big[\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)^\top\big)\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big]\Big| \\
&\le C_\mu\sum_{j=1}^{N}\|\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t}_j)\|^\zeta\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)-\boldsymbol{\gamma}_i(\boldsymbol{\theta})\|^\zeta \le \widetilde{C}_\mu.
\end{aligned} \tag{C.6}$$

Given $N$ is fixed, observe that $\mathbb{E}\big|\log\big(1+\|\boldsymbol{\varepsilon}_i\|^2\big)\big| < \infty$ follows from the existence of a power moment $\mathbb{E}\|\boldsymbol{\varepsilon}_i\|^{2q} < \infty$ for some $q \in (0, \nu_{10}/2)$, where $\nu_{10}$ is the true degree of freedom of $\boldsymbol{\varepsilon}_i$. Thus, to verify condition M2, it remains to be shown that $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big|\log\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\Big\|\Big|\Big) < \infty$ and $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big|\log\Big\|\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})\big)^{-1}\boldsymbol{\Sigma}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)\Big\|\Big|\Big) < \infty$. However, under Assumption SC1, these results follow directly from the arguments used for M1 above. This completes the proof for (C.3).

Finally, note that the continuity of the limit criterion function $\mathcal{L}(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ also follows from Theorem A.2.2 of White (1996). □

***Proof of Theorem 1.*** We follow the standard consistency arguments outlined in Blasques et al. (2022, Theorem 4.6), which rely on results such as Theorem 3.4 of White (1996) or Theorem 3.3 of Gallant and White (1988). Recall $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\big(\ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\big)$ from Lemma C.1. First, Lemma C.1 establishes the uniform convergence of the empirical criterion function $\widehat{\mathcal{L}}_T(\boldsymbol{\theta})$ to the limiting criterion function $\mathcal{L}(\boldsymbol{\theta})$ over the compact parameter space $\boldsymbol{\Theta}$.

To establish the identifiable uniqueness of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, we need to check (see White, 1996, Definition 3.3): (i) $\mathcal{L}(\boldsymbol{\theta}_0) > \mathcal{L}(\boldsymbol{\theta})$ for every $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (as shown below); (ii) the continuity of

$\boldsymbol{\theta} \mapsto \mathcal{L}(\boldsymbol{\theta})$ (as established in Lemma C.1); and (iii) the compactness of $\boldsymbol{\Theta}$ (as directly assumed). To show Part (i), note that $\mathcal{L}(\boldsymbol{\theta})$ depends on the unique limiting SE process $\{\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\, i \in \mathbb{Z}\}$ instead of the initialized process $\{\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}, \widehat{\boldsymbol{\gamma}}_0),\, i \in \mathbb{Z}^+\}$, and is therefore independent of the initial value $\widehat{\boldsymbol{\gamma}}_0$. Let $\mathcal{F}_i = \sigma(\boldsymbol{Y}_s,\, s \leq i)$ be a natural filtration and $\mathbb{E}_i(\cdot) = \mathbb{E}(\,\cdot \mid \mathcal{F}_i)$. In a manner similar to Blasques et al. (2022, Lemma TA.7), we have the following:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_0) - \mathcal{L}(\boldsymbol{\theta}) &= N^{-1}\mathbb{E}\left( \log \frac{p\big(\boldsymbol{Y}_i \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)}{p\big(\boldsymbol{Y}_i \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)} \right) \\
&= N^{-1}\mathbb{E}\left[ \mathbb{E}_{i-1}\left( \log \frac{p\big(\boldsymbol{Y}_i \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)}{p\big(\boldsymbol{Y}_i \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)} \right) \right] \\
&= N^{-1}\mathbb{E}\left[ \int p\big(\boldsymbol{y} \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big) \log \frac{p\big(\boldsymbol{y} \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)}{p\big(\boldsymbol{y} \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)} \, \mathrm{d}\boldsymbol{y} \right] \geq 0,
\end{aligned}
$$

using Gibbs' inequality for the Kullback-Leibler divergence (White, 1996, Theorem 2.3). Equality holds if and only if $p\big(\boldsymbol{y} \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big) = p\big(\boldsymbol{y} \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)$ for almost every $\boldsymbol{y} \in \mathbb{R}^N$, with respect to the density $p\big(\,\cdot \,\big|\, \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)$ (whose support is $\mathbb{R}^N$). Therefore, by Assumption SC2, $\mathcal{L}(\boldsymbol{\theta}_0) = \mathcal{L}(\boldsymbol{\theta})$ implies that $\boldsymbol{\nu} = \boldsymbol{\nu}_0$ and $\boldsymbol{\gamma}_i(\boldsymbol{\theta}) \overset{a.s.}{=} \boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)$ for arbitrary $i$. Next, we argue that if these two conditions hold, then it follows that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (uniqueness). Substituting these conditions into the recurrence equations for both $\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)$ and $\boldsymbol{\gamma}_i(\boldsymbol{\theta})$ and subtracting one from the other, we obtain:

$$
\mathbf{0} \overset{a.s.}{=} \boldsymbol{\omega} - \boldsymbol{\omega}_0 + (\boldsymbol{B} - \boldsymbol{B}_0)\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0) + (\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{s}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big). \tag{C.7}
$$

Since $\boldsymbol{s}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)$ depends on the innovation $\boldsymbol{\varepsilon}_i$, it is stochastic conditional on $\mathcal{F}_{i-1}$. As a result, we must have $\boldsymbol{A} = \boldsymbol{A}_0$. Moreover, given $\|\boldsymbol{B}_0\| < 1$ (Assumption SC3), by iterating backward into the infinite past (see, e.g., Straumann and Mikosch, 2006, Eq. (2.5)), we may write

$$
\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0) = \sum_{k=1}^{\infty} \boldsymbol{B}_0^k \big(\boldsymbol{\omega}_0 + \boldsymbol{A}_0 \boldsymbol{s}_{i-k}\big(\boldsymbol{\gamma}_{i-k}(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)\big) = (\boldsymbol{I}_{n_\gamma} - \boldsymbol{B}_0)^{-1}\boldsymbol{\omega}_0 + \boldsymbol{A}_0 \sum_{k=1}^{\infty} \boldsymbol{B}_0^k \boldsymbol{s}_{i-k}\big(\boldsymbol{\gamma}_{i-k}(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big).
$$

Note that $\sum_{k=1}^{\infty} \boldsymbol{B}_0^k \boldsymbol{s}_{i-k}\big(\boldsymbol{\gamma}_{i-k}(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)$ is almost surely nonzero. Given $\boldsymbol{A} = \boldsymbol{A}_0$ and $\boldsymbol{\nu} = \boldsymbol{\nu}_0$, to ensure (C.7), one must have $(\boldsymbol{B} - \boldsymbol{B}_0)\boldsymbol{A}_0 = \mathbf{O}$. Since $\det(\boldsymbol{A}_0) \neq 0$ (Assumption SC3), it follows that $\boldsymbol{A}_0$ has a trivial null space, which implies $\boldsymbol{B} = \boldsymbol{B}_0$. Consequently, from (C.7), we also conclude that $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and thus $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. This completes the proof. □

# D Proof of Proposition 4 (derivative processes)

We now establish the convergence of the first- and second-order derivative processes, which forms the foundation for proving Lemma E.1 - Lemma E.3 and, ultimately, the asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$.

***Proof of Proposition 4.*** We derive the results for $k = 1$ and $k = 2$ separately. The proof follows a similar approach used in Proposition 3.4 of Blasques et al. (2022) and Proposition 6.1 of Straumann and Mikosch (2006). For each $k$, we begin by deriving the general expression of the SRE for the derivative process and then establish the existence of an SE approximation to the perturbed system by verifying the conditions outlined in Straumann and Mikosch (2006, Theorem 2.10), similar to those applied in Proposition 3. Specifically, suppose the perturbed SRE for the process $\{\widehat{\boldsymbol{\gamma}}_i^{(k)}(\cdot),\, i \in \mathbb{Z}^+\}$, initialized at some $(\widehat{\boldsymbol{\gamma}}_0, \widehat{\boldsymbol{\gamma}}_0^{(1)}, \ldots, \widehat{\boldsymbol{\gamma}}_0^{(k)})$, can be expressed as $\widehat{\boldsymbol{\gamma}}_{i+1}^{(k)} = \widehat{\boldsymbol{\lambda}}_{(k),i}(\widehat{\boldsymbol{\gamma}}_i^{(k)})$, where the random maps $\widehat{\boldsymbol{\lambda}}_{(k),i} : \mathcal{C}^0(\boldsymbol{\Theta}, \mathbb{R}^{n_\gamma n_\theta^k}) \to \mathcal{C}^0(\boldsymbol{\Theta}, \mathbb{R}^{n_\gamma n_\theta^k})$ are defined on the Polish space $(E, d_E) = \left(\mathcal{C}^0(\boldsymbol{\Theta}, \mathbb{R}^{n_\gamma n_\theta^k}), \|\cdot\|_{\boldsymbol{\Theta}}\right)$, and are given by:

$$
\begin{aligned}
\big[\widehat{\boldsymbol{\lambda}}_{(k),i}(\boldsymbol{\gamma}^{(k)})\big](\boldsymbol{\theta}) &= \widehat{\boldsymbol{\psi}}_{(k),i}(\boldsymbol{\gamma}^{(k)}(\boldsymbol{\theta}), \boldsymbol{\theta}) \\
&:= \boldsymbol{Q}_{(k),i}(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\, \boldsymbol{\gamma}^{(k)}(\boldsymbol{\theta}) + \boldsymbol{q}_{(k),i}(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}), , \ldots, \widehat{\boldsymbol{\gamma}}_i^{(k-1)}(\boldsymbol{\theta}), \boldsymbol{\theta}),
\end{aligned} \tag{D.1}
$$

where $\boldsymbol{Q}_{(k),i}(\cdot)$ is a matrix independent of the derivatives of $\widehat{\boldsymbol{\gamma}}_i(\cdot)$, and $\boldsymbol{q}_{(k),i}(\cdot)$ is a vector that depends only on derivatives up to order $\widehat{\boldsymbol{\gamma}}_i^{(k-1)}(\cdot)$. Note that the perturbed sequence $\{\widehat{\boldsymbol{\gamma}}_i^{(k)}(\cdot),\, i \in \mathbb{Z}^+\}$ depends on the nonstationary initialized sequence $\left\{\left(\widehat{\boldsymbol{\gamma}}_i(\cdot), \widehat{\boldsymbol{\gamma}}_i^{(1)}(\cdot), \ldots, \widehat{\boldsymbol{\gamma}}_i^{(k-1)}(\cdot)\right),\, i \in \mathbb{Z}^+\right\}$ and is therefore only stationary in the limit. Moreover, let $\{\hat{\boldsymbol{d}}_i^{(k)}(\cdot),\, i \in \mathbb{Z}^+\}$ be the unperturbed sequence, initialized at zero, and depending solely on the limit sequence $\{(\boldsymbol{\gamma}_i(\cdot), \boldsymbol{\gamma}_i^{(1)}(\cdot), \ldots, \boldsymbol{\gamma}_i^{(k-1)}(\cdot)),\, i \in \mathbb{Z}\}$, which is associated with the random maps $\boldsymbol{\lambda}_{(k),i}$ defined by

$$
\begin{aligned}
\big[\boldsymbol{\lambda}_{(k),i}(\boldsymbol{d}^{(k)})\big](\boldsymbol{\theta}) &= \boldsymbol{\psi}_{(k),i}(\boldsymbol{d}^{(k)}(\boldsymbol{\theta}), \boldsymbol{\theta}) \\
&:= \boldsymbol{Q}_{(k),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\, \boldsymbol{d}^{(k)}(\boldsymbol{\theta}) + \boldsymbol{q}_{(k),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}), \ldots, \boldsymbol{\gamma}_i^{(k-1)}(\boldsymbol{\theta}), \boldsymbol{\theta}).
\end{aligned} \tag{D.2}
$$

The following high-level conditions (abbreviated as $\widehat{\text{HL}}$) are sufficient to apply Straumann and Mikosch (2006, Theorem 2.10): For $k \in \mathbb{Z}^+$, $\{\boldsymbol{\lambda}_{(k),i},\, i \in \mathbb{Z}\}$ is SE. Moreover,

$\widehat{\text{HL}}1$ $\mathbb{E}\left(\log^+ \left\|\boldsymbol{\lambda}_{(k),1}(\mathbf{0})\right\|_{\boldsymbol{\Theta}}\right) < \infty$;

$\widehat{\text{HL}}2$ $\mathbb{E}\left(\log^+ \rho(\boldsymbol{\lambda}_{(k),1})\right) < \infty$;

$\widehat{\text{HL}}3$ $\mathbb{E}\left(\log \rho(\boldsymbol{\lambda}_{(k),i}^{(r)})\right) < 0$ for some integer $r \geq 1$, where $\boldsymbol{\lambda}_{(k),i}^{(r)} = \boldsymbol{\lambda}_{(k),i} \circ \boldsymbol{\lambda}_{(k),i-1} \circ \ldots \circ \boldsymbol{\lambda}_{(k),i-r+1}$ is the $r$-fold convolution of $\boldsymbol{\lambda}_{(k),i}$;

$\widehat{\text{HL}}4$ $\mathbb{E}\Big(\log^+ \|\boldsymbol{d}_0^{(k)}\|_{\boldsymbol{\Theta}}\Big) < \infty$, where $\big\{\boldsymbol{d}_i^{(k)}(\cdot),\, i \in \mathbb{Z}\big\}$ is the unique SE solution of the unperturbed system (D.2) (with the existence guaranteed by Conditions $\widehat{\text{HL}}1$ - $\widehat{\text{HL}}3$);

$\widehat{\text{HL}}5$ $\Big\|\widehat{\boldsymbol{\lambda}}_{(k),i}(\mathbf{0}) - \boldsymbol{\lambda}_{(k),i}(\mathbf{0})\Big\|_{\boldsymbol{\Theta}} \xrightarrow{e.a.s.} 0$ and $\rho\Big(\widehat{\boldsymbol{\lambda}}_{(k),i} - \boldsymbol{\lambda}_{(k),i}\Big) \xrightarrow{e.a.s.} 0$ as $i \to \infty$.

Under these conditions, Theorem 2.10 in Straumann and Mikosch (2006) implies that $\big\{\boldsymbol{d}_i^{(k)}(\cdot),\, i \in \mathbb{Z}^+\big\}$ is an SE approximation of $\big\{\widehat{\gamma}_i^{(k)}(\cdot),\, i \in \mathbb{Z}^+\big\}$. Then, following the reasoning for Part (3) on p. 2483 of Straumann and Mikosch (2006), it follows that the $k_{\text{th}}$-order derivative $\boldsymbol{\gamma}_i^{(k)}$ of $\boldsymbol{\gamma}_i$ on $\boldsymbol{\Theta}$ coincide with $\boldsymbol{d}_i^{(k)}$ almost surely for every $i$, and thus the proposition is established. In the subsequent discussion, we repeatedly use the property $\|\boldsymbol{A} \otimes \boldsymbol{B}\| = \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$ for any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ without explicitly mentioning it.

**I. Proof of $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\gamma}_i^{(1)}(\boldsymbol{\theta}) - \gamma_i^{(1)}(\boldsymbol{\theta})\big\| \xrightarrow{e.a.s.} 0$:** The general SRE for the first-order derivative process (i.e., $k = 1$) is given in (F.33) of Lemma F.7. By directly matching the construction in (D.2) with (F.33) and Proposition 3, we conclude that $\big\{\boldsymbol{\lambda}_{(1),i},\, i \in \mathbb{Z}\big\}$ is SE. This follows from Theorem 3.35 of White (2001) and the fact that the sequence $\big\{(\boldsymbol{Y}_i, \boldsymbol{\gamma}_i),\, i \in \mathbb{Z}\big\}$ is SE, as $\boldsymbol{\gamma}_i$ is $\mathcal{F}_{i-1}$-measurable.

**Verification of Condition $\widehat{\text{HL}}1$:** Recall the definition of $\boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})$ in (F.34). Note that

$$\mathbb{E}\Big(\log^+ \|\boldsymbol{\lambda}_{(1),1}(\mathbf{0})\|_{\boldsymbol{\Theta}}\Big) = \mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{q}_{(1),1}\big(\gamma_1(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\Big\|\Big)$$
$$\leq \mathbb{E}\Bigg\{\log^+ C\Bigg(1 + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\gamma_1(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{s}_1\big(\gamma_1(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\Big\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Bigg\|\frac{\partial \boldsymbol{s}_1\big(\gamma_1(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}^\top}\Bigg\|\Bigg)\Bigg\}.$$

Using (B.3), (B.4), (C.5), and $\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \big\|\boldsymbol{s}_1(\boldsymbol{\gamma}, \boldsymbol{\nu})\big\| \leq C$, we obtain $\widehat{\text{HL}}1$ provided that $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\frac{\partial \boldsymbol{s}_1(\gamma_1(\boldsymbol{\theta}), \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\Big\|\Big) < \infty$. Note that (C.5) is guaranteed by Assumption AN1. Under Assumptions A1, A2, and AN1, along with the compactness of $\boldsymbol{\Theta}$, by Lemma F.7, one has

$$\sup_{i\in\mathbb{Z}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\frac{\partial \nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right\| \leq C. \tag{D.3}$$

Utilizing this result, along with the upper bounds in (B.14) and (B.15), as well as Assumption AN1, we arrive at

$$\sup_{i\in\mathbb{Z}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right\| \leq \sup_{i\in\mathbb{Z}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \Bigg\{\big\|\nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})\big\| \left\|\frac{\partial\, \mathrm{vec}\,\big(\boldsymbol{S}(\boldsymbol{\gamma}, \boldsymbol{\nu})\big)}{\partial \boldsymbol{\theta}^\top}\right\|$$
$$+ \big\|\boldsymbol{S}(\boldsymbol{\gamma}, \boldsymbol{\nu})\big\| \left\|\frac{\partial \nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right\|\Bigg\} \leq C. \tag{D.4}$$

Condition $\widehat{\text{HL}}1$ follows immediately.

**Verification of Condition $\widehat{\text{HL}}$2:** By definition, we have

$$\mathbb{E}\Big(\log^+ \rho(\boldsymbol{\lambda}_{(1),1})\Big) \leq \mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{Q}_{(1),1}(\boldsymbol{\gamma}_1(\boldsymbol{\theta}),\boldsymbol{\theta})\|\Big) \leq \mathbb{E}\Bigg(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \bigg\|\boldsymbol{B}+\boldsymbol{A}\,\frac{\partial \boldsymbol{s}_1(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\bigg\|\Bigg),$$

which is finite as shown below (B.16).

**Verification of Condition $\widehat{\text{HL}}$3:** Similarly, we have

$$\begin{aligned}\mathbb{E}\Big(\log \rho(\boldsymbol{\lambda}^{(r)}_{(1),i})\Big) &\leq \mathbb{E}\Bigg(\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bigg\|\prod_{j=1}^{r} \boldsymbol{Q}_{(1),i-j+1}\big(\boldsymbol{\gamma}_{i-j+1}(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\bigg\|\Bigg)\\ &= \mathbb{E}\Bigg(\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bigg\|\prod_{j=1}^{r} \bigg(\boldsymbol{B}+\boldsymbol{A}\,\frac{\partial \boldsymbol{s}_{i-j+1}\big(\boldsymbol{\gamma}_{i-j+1}(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}^\top}\bigg)\bigg\|\Bigg)\\ &\leq \mathbb{E}\Bigg(\log \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \bigg\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\boldsymbol{\psi}^{(r)}_i(\boldsymbol{\gamma},\boldsymbol{\theta})\bigg\|\Bigg) < 0, \qquad \text{(D.5)}\end{aligned}$$

for some $r \geq 1$, where the last step follows from Assumption IV3 as in Proposition 3.

**Verification of Condition $\widehat{\text{HL}}$4:** Conditions $\widehat{\text{HL}}1$ - $\widehat{\text{HL}}3$ ensure that the SE solution $\{\boldsymbol{d}^{(1)}_i\}$ of the unperturbed system admits an almost sure representation Straumann and Mikosch (2006, Theorem 2.8, Eq. (2.5)):

$$\boldsymbol{d}^{(1)}_i(\boldsymbol{\theta}) = \sum_{j=0}^{\infty}\bigg(\prod_{\ell=1}^{j}\boldsymbol{Q}_{(1),i-\ell}\big(\boldsymbol{\gamma}_{i-\ell}(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\bigg)\boldsymbol{q}_{(1),i-j-1}\big(\boldsymbol{\gamma}_{i-j-1}(\boldsymbol{\theta}),\boldsymbol{\theta}\big), \qquad \text{(D.6)}$$

with $\prod_{\ell=1}^{0}\cdot \equiv 1$. From the proof of $\widehat{\text{HL}}1$, one has $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{q}_{(1),i-j-1}\big(\boldsymbol{\gamma}_{i-j-1}(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\big\| \leq C$. For some $\kappa > 0$, by an application of the Minkowski inequality, we obtain

$$\mathbb{E}\Big(\|\boldsymbol{d}^{(1)}_i\|^{\kappa}_{\boldsymbol{\Theta}}\Big) \leq C\Bigg(\sum_{j=0}^{\infty}\Bigg[\mathbb{E}\Bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\bigg\|\prod_{\ell=1}^{j}\boldsymbol{Q}_{(1),i-\ell}\big(\boldsymbol{\gamma}_{i-\ell}(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\bigg\|\Bigg)^{\kappa}\Bigg]^{1/\kappa}\Bigg)^{\kappa}. \qquad \text{(D.7)}$$

As in the proof of $\widehat{\text{HL}}3$, it follows that

$$\begin{aligned}\mathbb{E}\Bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\bigg\|\prod_{\ell=1}^{j}\boldsymbol{Q}_{(1),i-\ell}\big(\boldsymbol{\gamma}_{i-\ell}(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\bigg\|\Bigg)^{\kappa} &= \mathbb{E}\Bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\bigg\|\prod_{\ell=1}^{j}\bigg(\boldsymbol{B}+\boldsymbol{A}\,\frac{\partial \boldsymbol{s}_{i-\ell}\big(\boldsymbol{\gamma}_{i-\ell}(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}^\top}\bigg)\bigg\|\Bigg)^{\kappa}\\ &\leq \mathbb{E}\Bigg(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\bigg\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\boldsymbol{\psi}^{(j)}_{i-1}(\boldsymbol{\gamma},\boldsymbol{\theta})\bigg\|^{\kappa}\Bigg). \qquad \text{(D.8)}\end{aligned}$$

By Assumption AN3, we have

$$\mathbb{E}\Big(\big\|\boldsymbol{d}_i^{(1)}\big\|_{\boldsymbol{\Theta}}^{\kappa}\Big) < \infty. \tag{D.9}$$

$\widehat{\text{HL}}4$ then follows from (D.9).

**Verification of Condition $\widehat{\text{HL}}5$:** Note that $\big\|\widehat{\boldsymbol{\lambda}}_{(1),i}(\mathbf{0}) - \boldsymbol{\lambda}_{(1),i}(\mathbf{0})\big\|_{\boldsymbol{\Theta}} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{q}_{(1),i}\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big) - \boldsymbol{q}_{(1),i}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\big\| \leq C\sum_{j=1}^{3} R_{j,i}$, where $R_{1,i} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\|$,

$$R_{2,i} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big) - \boldsymbol{s}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\Big\|, \qquad R_{3,i} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\frac{\partial \boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}^\top} - \frac{\partial \boldsymbol{s}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}^\top}\right\|.$$

By Proposition 3, $R_{1,i} \xrightarrow{e.a.s.} 0$ as $i \to \infty$. As in (B.12), by applying a mean value theorem for vector-valued functions (Rudin, 1976, Theorem 9.19), we obtain

$$R_{2,i} \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\|. \tag{D.10}$$

Given (F.45) in Lemma F.7, applying Assumptions A1, A2, and AN2, leads to

$$\sup_{i\in\mathbb{Z}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\frac{\partial \nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right\| \leq C. \tag{D.11}$$

Given the construction of $\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}$ in (F.44), using the upper bounds in (B.14) and (B.15), along with Assumption AN1, we obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right\| \leq C$. Therefore, $R_{2,i} \xrightarrow{e.a.s.} 0$ as $i \to \infty$ as well. Finally, note that

$$R_{3,i} \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right)\right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\|. \tag{D.12}$$

By (F.48) - (F.49) in Lemma F.7 and under Assumptions A1, A2, and AN1, we can deduce that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right)\right\| \leq C. \tag{D.13}$$

Combining (D.13) with (D.3) and (D.11), we obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right)\right\| \leq C$, leading to $R_{3,i} \xrightarrow{e.a.s.} 0$ as $i \to \infty$ by Proposition 3. It is now evident that $\big\|\widehat{\boldsymbol{\lambda}}_{(1),i}(\mathbf{0}) - \boldsymbol{\lambda}_{(1),i}(\mathbf{0})\big\|_{\boldsymbol{\Theta}} \xrightarrow{e.a.s.} 0$ as $i \to \infty$.

Furthermore, based on the discussion above and the compactness of $\boldsymbol{\Theta}$, we have, as $i \to \infty$,

$$\rho\Big(\widehat{\boldsymbol{\lambda}}_{(1),i} - \boldsymbol{\lambda}_{(1),i}\Big) \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\| \boldsymbol{Q}_{(1),i}\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big) - \boldsymbol{Q}_{(1),i}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big) \Big\|$$
$$= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \boldsymbol{A}\left( \frac{\partial \boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}^\top} - \frac{\partial \boldsymbol{s}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}^\top} \right) \right\| \le \left( \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{A}\| \right) R_{3,i} \xrightarrow{e.a.s.} 0,$$

which justifies the second part of $\widehat{\text{HL}}5$.

Under $\widehat{\text{HL}}1$ - $\widehat{\text{HL}}5$, Theorem 2.10 of Straumann and Mikosch (2006) implies that $\big\|\widehat{\boldsymbol{\gamma}}_i^{(1)} - \boldsymbol{d}_i^{(1)}\big\|_{\boldsymbol{\Theta}} \xrightarrow{e.a.s.} 0$ as $i \to \infty$. Finally, as noted, following the reasoning in Part (3) on p. 2483 of Straumann and Mikosch (2006), it follows that $\boldsymbol{\gamma}_i^{(1)} \equiv \boldsymbol{d}_i^{(1)}$.

**II. Proof of $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i^{(2)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\big\| \xrightarrow{e.a.s.} 0$:** The general SRE for the second-order derivative process is presented in (F.38) of Lemma F.7. By directly matching the construction in (D.2) with (F.38), invoking Proposition 3, and noting that $\big\{\boldsymbol{\gamma}_i^{(1)}, i \in \mathbb{Z}\big\}$ is $\mathcal{F}_{i-1}$-measurable, it becomes clear that $\big\{\boldsymbol{\lambda}_{(2),i}, i \in \mathbb{Z}\big\}$ is SE.

**Verification of Condition $\widehat{\text{HL}}1$:** Recall the definition of $\boldsymbol{q}_{(2),i}$ in (F.39). Moreover, note that $\mathbb{E}\big(\log^+ \|\boldsymbol{\lambda}_{(2),1}(\boldsymbol{0})\|_{\boldsymbol{\Theta}}\big) = \mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{q}_{(2),1}\big(\boldsymbol{\gamma}_1^{(1)}(\boldsymbol{\theta}), \boldsymbol{\gamma}_1(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\big\|\Big)$. Since $\|\operatorname{vec}(\cdot)\| = \|\cdot\|_F \le \sqrt{\operatorname{rank}(\cdot)}\, \|\cdot\|$, for every $i \in \mathbb{Z}$, one has

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\| \boldsymbol{q}_{(2),i}\big(\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}), \boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big) \Big\| \le C \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\Big\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \operatorname{vec}\big(\boldsymbol{Q}_{(1),i}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\big)}{\partial \boldsymbol{\theta}^\top} \right\|$$
$$+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top} \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\Big\|. \quad \text{(D.14)}$$

We claim that:

(IIa) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \operatorname{vec}(\boldsymbol{Q}_{(1),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^\top} \right\| \le C + C \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\Big\|$;

(IIb) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| \le C$;

(IIc) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top} \right\| \le C$.

If (IIa) - (IIc) hold, we immediately obtain $\widehat{\text{HL}}1$ using the inequalities (B.3) - (B.4) and the moment bound (D.9). Next, we show the claims in (IIa) - (IIc).

Given the norm equivalence in finite dimensional spaces, establishing the claim in (IIa) reduces to considering one of its blocks: $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \boldsymbol{B} + \boldsymbol{A} \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\|$. By the compactness of $\boldsymbol{\Theta}$

and using the chain rule, we arrive at

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\boldsymbol{B}+\boldsymbol{A}\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\| \\
&\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{\left\|\frac{\partial\operatorname{vec}(\boldsymbol{B})}{\partial\boldsymbol{\theta}^\top}\right\|+\left\|\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right\|\left\|\frac{\partial\operatorname{vec}(\boldsymbol{A})}{\partial\boldsymbol{\theta}^\top}\right\|+\|\boldsymbol{A}\|\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|\right\} \\
&\le C+C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\right\|+C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|.
\end{aligned}$$

By (D.11), Assumptions A1, AN2, and the results in Parts (iii) and (viii) of Lemma F.7, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|\le C. \tag{D.15}$$

As a result, we have $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|\le C$. Moreover, it has been shown above that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|\le C$. Combining these results gives the claim in (IIa).

For the claim in (IIb), we utilize Assumptions A1 and AN1, along with Eq. (D.3), the results in (F.36), and Parts (iii-1) and (vii) of Lemma F.7, obtaining

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right\|\le C. \tag{D.16}$$

Consequently, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right\|\le C$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right\|\le C$ in conjunction with (D.4).

Finally, the claim in (IIc) immediately follows from $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right\|\le C$.

**Verification of Conditions $\widehat{\text{HL2}}$ - $\widehat{\text{HL4}}$:** The steps for verifying $\widehat{\text{HL2}}$ - $\widehat{\text{HL4}}$ are similar to those in the proof for the first-order derivative processes. We only provide some details for $\widehat{\text{HL4}}$. Recall that for $\kappa>0$ defined in Assumption AN3, we have $\mathbb{E}\left(\|\boldsymbol{d}_i^{(1)}\|_{\boldsymbol{\Theta}}^{\kappa}\right)<\infty$ as established in (D.9). Using (D.9), (D.14) implies that $\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{q}_{(2),i}\left(\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}),\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\right)\right\|\right)^{\kappa/2}\le C$. Similar to (D.7), for $\tilde{\kappa}=\kappa/3>0$, by employing the Minkowski inequality and subsequently Hölder's inequality,

$$\begin{aligned}
\left[\mathbb{E}\left(\|\boldsymbol{d}_i^{(2)}\|_{\boldsymbol{\Theta}}^{\tilde{\kappa}}\right)\right]^{1/\tilde{\kappa}} &\le \sum_{j=0}^{\infty}\left\{\left[\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\prod_{\ell=1}^{j}\boldsymbol{Q}_{(2),i-\ell}\left(\boldsymbol{\gamma}_{i-\ell}(\boldsymbol{\theta}),\boldsymbol{\theta}\right)\right\|\right)^{3\tilde{\kappa}}\right]^{1/(3\tilde{\kappa})}\right. \\
&\qquad\left.\times\left[\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{q}_{(2),i}\left(\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}),\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\right)\right\|\right)^{3\tilde{\kappa}/2}\right]^{2/(3\tilde{\kappa})}\right\} \\
&\le C\sum_{j=0}^{\infty}\left[\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\prod_{\ell=1}^{j}\boldsymbol{Q}_{(2),i-\ell}\left(\boldsymbol{\gamma}_{i-\ell}(\boldsymbol{\theta}),\boldsymbol{\theta}\right)\right\|\right)^{\kappa}\right]^{1/\kappa}<\infty,
\end{aligned} \tag{D.17}$$

where the final step follows from Assumption AN3 as in (D.8). Therefore, we obtain $\mathbb{E}\Big(\|\boldsymbol{d}_i^{(2)}\|_{\boldsymbol{\Theta}}^{\kappa/3}\Big) < \infty$ and thus $\widehat{\text{HL}}4$.

**Verification of Condition $\widehat{\text{HL}}$5:** Note that, by $\|\operatorname{vec}(\cdot)\| \leq \sqrt{\operatorname{rank}(\cdot)}\,\|\cdot\|$,

$$\big\|\widehat{\boldsymbol{\lambda}}_{(2),i}(\mathbf{0}) - \boldsymbol{\lambda}_{(2),i}(\mathbf{0})\big\|_{\boldsymbol{\Theta}} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\|\boldsymbol{q}_{(2),i}\big(\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}), \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big) - \boldsymbol{q}_{(2),i}\big(\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}), \boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta}\big)\Big\| \leq C \sum_{j=1}^{3} \tilde{R}_{j,i},$$

where

$$\begin{aligned}
\tilde{R}_{1,i} &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \big(\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta})^\top \otimes \boldsymbol{I}_{n_\gamma n_\theta}\big) \frac{\partial \operatorname{vec}\big(\boldsymbol{Q}_{(1),i}(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}^\top} - \big(\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})^\top \otimes \boldsymbol{I}_{n_\gamma n_\theta}\big) \frac{\partial \operatorname{vec}\big(\boldsymbol{Q}_{(1),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}^\top} \right\| \\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta})\big\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \operatorname{vec}\big(\boldsymbol{Q}_{(1),i}(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}^\top} - \frac{\partial \operatorname{vec}\big(\boldsymbol{Q}_{(1),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}^\top} \right\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\big\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \operatorname{vec}\big(\boldsymbol{Q}_{(1),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}), \boldsymbol{\theta})\big)}{\partial \boldsymbol{\theta}^\top} \right\|,
\end{aligned} \tag{D.18}$$

and

$$\begin{aligned}
\tilde{R}_{2,i} &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \operatorname{vec}\left( \left.\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} \right) - \operatorname{vec}\left( \left.\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \right) \right\| \\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\|,
\end{aligned} \tag{D.19}$$

$$\begin{aligned}
\tilde{R}_{3,i} &= \left\| \left( \boldsymbol{I}_{n_\theta} \otimes \left.\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} \right) \widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \left( \boldsymbol{I}_{n_\theta} \otimes \left.\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \right) \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}) \right\| \\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top} \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\big\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta})\big\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\big\|.
\end{aligned} \tag{D.20}$$

Note that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\|$ is measurable, owing to the continuity of $\boldsymbol{\gamma}_i^{(1)}(\cdot)$ on $\boldsymbol{\Theta}$ and the compactness of $\boldsymbol{\Theta}$. Moreover, $\Big\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\|,\ i \in \mathbb{Z}\Big\}$ is SE with $\mathbb{E}\Big(\log^+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\gamma}_0^{(1)}(\boldsymbol{\theta})\|\Big) < \infty$ as shown in Eq. (D.9). Since $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0$ as $i \to \infty$, it follows from Lemma 2.1 of Straumann and Mikosch (2006) that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0$ as $i \to \infty$. Note further that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\|$. With the results in Parts (IIa) and (IIc) on Appendix p. 16, to establish that $\|\widehat{\boldsymbol{\lambda}}_{(2),i}(\mathbf{0}) - \boldsymbol{\lambda}_{(2),i}(\mathbf{0})\|_{\boldsymbol{\Theta}} \xrightarrow{e.a.s.} 0$ as $i \to \infty$, it suffices to prove that

(IId) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial \operatorname{vec}(\boldsymbol{Q}_{(1),i}(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^\top} - \frac{\partial \operatorname{vec}(\boldsymbol{Q}_{(1),i}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^\top} \right\| \xrightarrow{e.a.s.} 0$ as $i \to \infty$;

(IIe) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right) \right\| \leq C$;

(IIf) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| \leq C$.

For (IId), the proof reduces to showing that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \boldsymbol{A}\left( \frac{\partial \boldsymbol{s}_i(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} - \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right) \right\| \xrightarrow{e.a.s.} 0$ as $i \to \infty$ because of the norm equivalence. By the chain rule (Lütkepohl, 2005, Proposition A.1), we have

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \boldsymbol{A}\left( \frac{\partial \boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}^\top} - \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right) \right\| \\
&\leq C \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta}) \right\| \\
&\qquad + C \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}^\top} - \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| \\
&\leq C \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta}) \right\| + C \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}^\top} - \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\|. \qquad \text{(D.21)}
\end{aligned}$$

Recall that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| \leq C$. It remains to show the second term in (D.21) coverages to zero e.a.s. Using the chain rule again, we obtain

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\gamma}^\top} - \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| \\
\leq &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \bigg|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} \frac{\partial \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \frac{\partial \boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\| \\
+ &\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \bigg|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} - \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \right\| =: \Delta^{\gamma\theta}_{\Upsilon,1} + \Delta^{\gamma\theta}_{\Upsilon,2}. \quad \text{(D.22)}
\end{aligned}$$

Note that $\|\boldsymbol{C}_1\boldsymbol{D}_1 - \boldsymbol{C}_2\boldsymbol{D}_2\| \leq \|\boldsymbol{C}_1 - \boldsymbol{C}_2\| \, \|\boldsymbol{D}_2\| + \|\boldsymbol{D}_1 - \boldsymbol{D}_2\| \, \|\boldsymbol{C}_2\| + \|\boldsymbol{C}_1 - \boldsymbol{C}_2\| \, \|\boldsymbol{D}_1 - \boldsymbol{D}_2\|$. Together with a mean value theorem for vector-valued functions (see, e.g., Rudin, 1976, Theorem 9.19), this implies

$$\begin{aligned}
\Delta^{\gamma\theta}_{\Upsilon,1} &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\boldsymbol{\gamma}}^{(1)}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}^{(1)}_i(\boldsymbol{\theta}) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta}) \right\| \\
&\quad \times \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right) \right\| \left( \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \boldsymbol{\gamma}^{(1)}_i(\boldsymbol{\theta}) \right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\boldsymbol{\gamma}}^{(1)}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}^{(1)}_i(\boldsymbol{\theta}) \right\| \right), \\
\Delta^{\gamma\theta}_{\Upsilon,2} &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right) \right\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i(\boldsymbol{\theta}) \right\|.
\end{aligned}$$

We see that (D.22) converges to zero e.a.s. if $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}} \left\| \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left( \frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \right) \right) \right\| \leq C$

and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right)\right\|\leq C$.

Although these two conditions hold, the proof involves third-order derivatives and is exceedingly tedious. As the steps do not provide additional insights, we omit the full details and instead outline the key steps below. Given (F.52) and Assumptions AN1, AN2, by applying (F.12) and (F.22) repeatedly, we have

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right)\right\| &\leq C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| + C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right\| \\ &+ C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\| + C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right)\right\|.\end{aligned}$$

Note that the first three terms are bounded as shown in (B.14), (B.15), (D.11), (D.15). By applying the results from Parts *(iii-5)*, *(v)*, and *(viii)* of Lemma F.7, we conclude that the final term above is bounded as well. Therefore, one has $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right)\right\|\leq C$. The proof of $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right)\right\|\leq C$ follows similar steps as above. Combining these results gives (IId).

For (IIe) - (IIf), from (F.36) and (F.37), we have

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right)\right\| &\leq C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right\| \\ &+ C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\theta}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right)\right\|\leq C,\end{aligned}$$

and

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}^\top}\right)\right\| &\leq C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\| \\ &+ C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right)\right\|\leq C.\end{aligned}$$

It follows that $\left\|\widehat{\boldsymbol{\lambda}}_{(2),i}(\mathbf{0})-\boldsymbol{\lambda}_{(2),i}(\mathbf{0})\right\|_{\boldsymbol{\Theta}}\xrightarrow{e.a.s.}0$ as $i\to\infty$.

Furthermore, as $i\to\infty$, $\rho\big(\widehat{\boldsymbol{\lambda}}_{(2),i}-\boldsymbol{\lambda}_{(2),i}\big)\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{Q}_{(2),i}\big(\widehat{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\big)-\boldsymbol{Q}_{(2),i}\big(\gamma_i(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\right\|\xrightarrow{e.a.s.}0$ follows directly from the same argument used for $\rho\big(\widehat{\boldsymbol{\lambda}}_{(1),i}-\boldsymbol{\lambda}_{(1),i}\big)$ earlier.

Finally, using the reasoning in Part (3) on p. 2483 of Straumann and Mikosch (2006) again, one has $\boldsymbol{\gamma}_i^{(2)}=\boldsymbol{d}_i^{(2)}$ on $\boldsymbol{\Theta}$ a.s. This completes the proof. □

# E Proof of Theorem 2 (asymptotic normality)

The following lemmas (Lemma E.1 - Lemma E.3) establish the asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$.

**Lemma E.1.** *Suppose $\{\mathbf{Y}_i,\, i \in \mathbb{Z}\}$ is generated by (2.9) and (2.10) with $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Under the assumptions of Proposition 4, if $\mathbb{E}\Big(\big\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\big\|^2\Big) < \infty$, then $\sqrt{T}\,\frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\mathcal{I}}_0\big)$ as $T \to \infty$, where $\boldsymbol{\mathcal{I}}_0 = \mathbb{E}\Big(\frac{\partial \ell_i(\gamma_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0)}{\partial \boldsymbol{\theta}} \frac{\partial \ell_i(\gamma_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0)}{\partial \boldsymbol{\theta}^\top}\Big)$ is defined in Theorem 2.*

***Proof of Lemma E.1.*** We use the Cramér-Wold device to establish the asymptotic distribution $\sqrt{T}\,\frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$. Let $\boldsymbol{a} \in \mathbb{R}^{n_\theta}$ be any unit vector. Then, we consider the limiting distribution of

$$\sqrt{T}\left(\boldsymbol{a}^\top \frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) = \sum_{i=1}^{T} X_{T,i}, \qquad X_{T,i} = \frac{1}{\sqrt{T}}\,\boldsymbol{a}^\top \frac{\partial \ell_i\big(\gamma_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \tag{E.1}$$

as $T \to \infty$. We shall apply the CLT for a martingale difference array, as provided in Theorem 24.3 of Davidson (1994). We verify the following conditions, enabling us to apply the theorem:

CLT1 $\{X_{T,i}, \mathcal{F}_{T,i}\}$ is a martingale difference array, where $\mathcal{F}_{T,i} = \mathcal{F}_i = \sigma\big(\mathbf{Y}_s,\, s \le i\big)$ for all $i \le T$;

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

CLT3 $\max_{1 \le i \le T} |X_{T,i}| \xrightarrow{p} 0$.

**Verification of Condition CLT1:** Without confusion, define $\mathbb{E}_i(\cdot) = \mathbb{E}(\,\cdot \mid \mathcal{F}_{T,i})$. Note that CLT1 follows immediately if it holds a.s. that

$$\begin{aligned}\mathbb{E}_{i-1}\left(\frac{\partial \ell_i\big(\gamma_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) &= N^{-1} \int \frac{\partial \log p\big(\boldsymbol{y} \,\big|\, \gamma_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} p\big(\boldsymbol{y} \,\big|\, \gamma_i(\boldsymbol{\theta}_0), \boldsymbol{\nu}_0\big)\, \mathrm{d}\boldsymbol{y} \\ &= N^{-1} \int \frac{\partial\, p\big(\boldsymbol{y} \,\big|\, \gamma_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \mathrm{d}\boldsymbol{y} = \mathbf{0}.\end{aligned} \tag{E.2}$$

Note that $\int p\big(\boldsymbol{y} \,\big|\, \gamma_i(\boldsymbol{\theta}), \boldsymbol{\nu}\big)\, \mathrm{d}\boldsymbol{y} = 1$. Then (E.2) holds trivially if differentiation with respect to $\boldsymbol{\theta}$ and integration can be interchanged; see, for example, the conditions in Schilling (2017, Theorem 12.5) or Klenke (2020, Theorem 6.28). Using results from Lemma F.7 to verify the interchangeability is possible but cumbersome, and thus omitted here. On the other hand, one can also directly compute (E.2). Recall $\boldsymbol{\theta} = \big(\nu_1, \boldsymbol{\nu}_2^\top, \boldsymbol{\omega}^\top, \mathrm{vec}(\boldsymbol{A})^\top, \mathrm{vec}(\boldsymbol{B})^\top\big)^\top$ with true value $\boldsymbol{\theta}_0 = \big(\nu_1, \boldsymbol{\nu}_{20}^\top, \boldsymbol{\omega}_0^\top, \mathrm{vec}(\boldsymbol{A}_0)^\top, \mathrm{vec}(\boldsymbol{B}_0)^\top\big)^\top$. Adopt the notation $\boldsymbol{u}_i(\boldsymbol{\nu}_2, \gamma) = \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2} \boldsymbol{\Sigma}(\gamma)^{-1}\big(\mathbf{Y}_i - \boldsymbol{\mu}(\gamma)\big)$ from (F.7). Note that $\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20}, \gamma_i(\boldsymbol{\theta}_0)\big) = \boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1/2} \boldsymbol{\varepsilon}_i \overset{i.i.d.}{\sim} t_{\nu_{10}}\big(\mathbf{0}, \boldsymbol{I}_N\big)$, where $\boldsymbol{\varepsilon}_i$ is given in Eq.

(2.10). By the chain rule, we have

$$\left.\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left.\frac{\partial \boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^\top. \tag{E.3}$$

Note that $\left.\frac{\partial \boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is $\mathcal{F}_{T,i-1}$-measurable. Therefore, to prove (E.2), it suffices to show that

$$\mathbb{E}_{i-1}\left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = \mathbf{0}, \qquad \mathbb{E}_{i-1}\left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = \mathbf{0}, \quad a.s. \tag{E.4}$$

**i.** For the first part in (E.4), note that $\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \nu_1}, \frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\nu}_2^\top}, \frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\omega}^\top}, \frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \operatorname{vec}(\boldsymbol{A})^\top}, \frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \operatorname{vec}(\boldsymbol{B})^\top}\right)^\top$, where the last three terms are $\mathbf{0}$. For the first term, we have

$$\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \nu_1}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{2N}\left[\psi\left(\frac{\nu_{10}+N}{2}\right) - \psi\left(\frac{\nu_{10}}{2}\right) + \kappa_{1,i}^0 + \log(\kappa_{2,i}^0) - \frac{N}{\nu_{10}}\kappa_{2,i}^0\right],$$

where $\psi(x) = \mathrm{d}\log\Gamma(x)/\mathrm{d}x$ is the digamma function, and

$$\kappa_{1,i}^0 = \frac{\nu_{10}^{-1}\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)^\top \boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)}{1+\nu_{10}^{-1}\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)^\top \boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)}, \qquad \kappa_{2,i}^0 = 1-\kappa_{1,i}^0. \tag{E.5}$$

By (F.1) in Lemma F.4 and the moment properties of beta distribution, we know $\mathbb{E}_{i-1}(\kappa_{1,i}^0) = \mathbb{E}(\kappa_{1,i}^0) = \frac{N}{\nu_{10}+N}$ and $\mathbb{E}_{i-1}\log(\kappa_{2,i}^0) = \mathbb{E}\log(1-\kappa_{1,i}^0) = \psi\left(\frac{\nu_{10}}{2}\right) - \psi\left(\frac{\nu_{10}+N}{2}\right)$. Therefore, we obtain

$$\mathbb{E}_{i-1}\left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \nu_1}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = 0. \tag{E.6}$$

We now address the second term in $\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}$. Note that, for a differentiable map $\boldsymbol{A}(\cdot):\mathbb{R}\to\mathbb{R}^{N\times N}$, one has the well-known Jacobi's formula:

$$\frac{\mathrm{d}}{\mathrm{d}x}\det\big(\boldsymbol{A}(x)\big) = \det\big(\boldsymbol{A}(x)\big)\operatorname{tr}\left(\boldsymbol{A}(x)^{-1}\frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{A}(x)\right), \tag{E.7}$$

provided $\boldsymbol{A}(x)$ is invertible (see, e.g., Horn and Johnson, 2012, Eq. (0.8.10.1)). Let $\nu_{2,j}$ be the $j_{\text{th}}$ element of $\boldsymbol{\nu}_2$. Using Eq. (E.7), we can obtain

$$\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \nu_{2,j}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{2N}\operatorname{tr}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})\frac{\partial}{\partial \nu_{2,j}}\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1}\big)\right) - \frac{1}{2N}\frac{\nu_{10}+N}{\nu_{10}}\kappa_{2,i}^0\,\boldsymbol{\varepsilon}_i^\top\frac{\partial}{\partial \nu_{2,j}}\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1}\big)\boldsymbol{\varepsilon}_i.$$

By the identity (F.6), we have

$$\begin{aligned}
&\mathbb{E}_{i-1}\left(\frac{\nu_{10}+N}{\nu_{10}}\,\kappa_{2,i}^{0}\,\boldsymbol{\varepsilon}_i^{\top}\frac{\partial}{\partial \nu_{2,j}}\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1}\big)\boldsymbol{\varepsilon}_i\right)\\
&=(\nu_{10}+N)\operatorname{tr}\left\{\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{1/2}\frac{\partial}{\partial \nu_{2,j}}\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1}\big)\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{1/2}\mathbb{E}\left(\frac{\nu_{10}^{-1}\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)^{\top}}{1+\nu_{10}^{-1}\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)^{\top}\boldsymbol{u}_i\big(\boldsymbol{\nu}_{20},\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0)\big)}\right)\right\}\\
&=(\nu_{10}+N)\operatorname{tr}\left\{\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{1/2}\frac{\partial}{\partial \nu_{2,j}}\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1}\big)\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{1/2}\,\mathbb{E}(\kappa_{1,i}^{0})\,\mathbb{E}\big(\tilde{\boldsymbol{\mathcal{U}}}_i\tilde{\boldsymbol{\mathcal{U}}}_i^{\top}\big)\right\}\\
&=\operatorname{tr}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})\,\frac{\partial}{\partial \nu_{2,j}}\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1}\big)\right),
\end{aligned}$$

where $\tilde{\boldsymbol{\mathcal{U}}}_i \in \mathbb{R}^N$ is uniformly distributed on the unit sphere in $\mathbb{R}^N$ with $\mathbb{E}\big(\tilde{\boldsymbol{\mathcal{U}}}_i\tilde{\boldsymbol{\mathcal{U}}}_i^{\top}\big) = N^{-1}\boldsymbol{I}_N$ (Fang et al., 2018, Eq. (2.17)). As a result, we have

$$\mathbb{E}\left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \nu_{2,j}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)=0, \qquad j\in [\![\, n_\nu - 1\,]\!]. \tag{E.8}$$

Combining these results gives the first part in (E.4).

**ii.** For the second part in (E.4), we have

$$\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \nabla_i^{\mu,\varepsilon}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\boldsymbol{\nu}_0)+\nabla_i^{\sigma,\varepsilon}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\boldsymbol{\nu}_0), \tag{E.9}$$

where $\nabla_i^{\mu,\varepsilon}(\cdot)$ and $\nabla_i^{\sigma,\varepsilon}(\cdot)$ are shown in (B.2). Note that $\mathbb{E}_{i-1}\Big(\nabla_i^{\mu,\varepsilon}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\boldsymbol{\nu}_0)\Big)=\boldsymbol{0}$ by (F.5), and $\mathbb{E}_{i-1}\Big(\nabla_i^{\sigma,\varepsilon}(\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\boldsymbol{\nu}_0)\Big)=\boldsymbol{0}$ is a result of the identity $\operatorname{row}_j\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{1/2}\big)\big[\operatorname{row}_j\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_{20})^{-1/2}\big)\big]^{\top}=1$ for $j\in [\![\, N\,]\!]$ and (F.6). The second part in (E.4) is thus obtained.

**Verification of Conditions CLT2 – CLT3:** If $\Big\{\boldsymbol{a}^{\top}\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},\ i\in\mathbb{Z}\Big\}$ is additionally SE with finite variance, then Condition CLT2 follows from Theorem 13.12, and Condition CLT3 follows from Theorem 23.16 of Davidson (1994), as it implies that the Lindeberg condition holds for $\{X_{T,i}\}$; also see, e.g., Theorem 18.3 of Billingsley (1999) and the discussion in Chapter 24.3 of Davidson (1994, p. 385). From (E.3), we see that $\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is a continuous function of $\big(\boldsymbol{\varepsilon}_i,\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\big)$. As the tuple $\big(\boldsymbol{\varepsilon}_i,\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\big)$ can be expressed in terms of $(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_{i-1},\ldots)$ under the assumptions of Proposition 4, it follows that it is (jointly) SE. Consequently, we conclude that the sequence $\Big\{\boldsymbol{a}^{\top}\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},\ i\in\mathbb{Z}\Big\}$ is also SE (Krengel, 1985, Proposition 4.3). Next, we show that the second-order moment of $\boldsymbol{a}^{\top}\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is finite. By the identity (E.3), we obtain

$$\mathbb{E}\left\|\boldsymbol{a}^\top \frac{\partial \ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|^2 \le C\Bigg\{\mathbb{E}\left\|\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|^2 + \mathbb{E}\left(\left\|\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|^2 \left\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\right\|^2\right)\Bigg\}. \tag{E.10}$$

Note that the first term inside the curly brackets above is bounded by applying Lemma F.4. Furthermore, from (E.9), $\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ depends only on the present innovation $\boldsymbol{\varepsilon}_i$, whereas $\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)$ depends on the past $(\boldsymbol{\varepsilon}_{i-1},\boldsymbol{\varepsilon}_{i-2},\ldots)$; thus, they are independent. Therefore, using Lemma F.4, the second term can be bounded by

$$\mathbb{E}\left(\left\|\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|^2\right)\mathbb{E}\left(\left\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\right\|^2\right) \le C\,\mathbb{E}\left(\left\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\right\|^2\right) < \infty, \tag{E.11}$$

where the final step is due to the assumption $\mathbb{E}\big(\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\|^2\big) < \infty$. This completes the proof. □

**Lemma E.2.** *Under the assumptions of Proposition 4, if* $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\big\|^2\Big) < \infty$ *and* $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\|\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\big\|\Big) < \infty$, *we have*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \boldsymbol{\mathcal{J}}(\boldsymbol{\theta})\right\| \xrightarrow{a.s.} 0, \qquad \text{as} \quad T\to\infty, \tag{E.12}$$

*where* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}) = \mathbb{E}\Big(\frac{\partial^2 \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big)$.

***Proof of Lemma E.2.*** Similar to the proof of Lemma C.1, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\| \le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \mathcal{L}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \mathcal{L}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \boldsymbol{\mathcal{J}}(\boldsymbol{\theta})\right\|.$$

Thus, it is enough to show that each term on the right-hand side converges to zero almost surely.

**First term** $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \mathcal{L}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\| \xrightarrow{a.s.} 0$**:** Note that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \mathcal{L}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\| \le T^{-1}\sum_{i=1}^{T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \ell_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\|. \tag{E.13}$$

Using the argument below (C.2), it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial^2 \ell_i\big(\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{\partial^2 \ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right\| \xrightarrow{e.a.s.} 0, \tag{E.14}$$

as $i \to \infty$. From (E.3), applying the chain rule and (F.18), we obtain

$$\begin{aligned}\frac{\partial^2 \ell_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} + \frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})^\top}{\partial\boldsymbol{\theta}}\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} + \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\\ &+ \frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})^\top}{\partial\boldsymbol{\theta}}\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} + \left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \otimes \boldsymbol{I}_{n_\theta}\right)\boldsymbol{\Gamma}_i(\boldsymbol{\theta}),\end{aligned} \tag{E.15}$$

with $\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{pmatrix} \frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\nu_1^2} & \frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\nu_1\partial\boldsymbol{\nu}_2^\top} & \\ \frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_2\partial\nu_1} & \frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_2\partial\boldsymbol{\nu}_2^\top} & \\ & & \mathbf{O}\end{pmatrix}$, $\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} = \left.\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top}{\partial\boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}$, $\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} = \left.\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} + \frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})^\top}{\partial\boldsymbol{\theta}}\left.\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top}{\partial\boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}$, $\boldsymbol{\Gamma}_i(\boldsymbol{\theta}) = \begin{pmatrix}\frac{\partial^2\gamma_{i,1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\\ \vdots \\ \frac{\partial^2\gamma_{i,n_\gamma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\end{pmatrix}$ and $\gamma_{i,j}(\cdot)$ is the $j_{\text{th}}$ element of $\boldsymbol{\gamma}_i(\cdot)$ for $j \in [\![ n_\gamma ]\!]$.

Therefore, to establish (E.14), we shall prove

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} - \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \right\| \xrightarrow{e.a.s.} 0, \tag{E.16}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})}\frac{\partial\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} \right\| \xrightarrow{e.a.s.} 0, \tag{E.17}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})^\top}{\partial\boldsymbol{\theta}}\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})}\frac{\partial\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})^\top}{\partial\boldsymbol{\theta}}\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} \right\| \xrightarrow{e.a.s.} 0, \tag{E.18}$$

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} \otimes \boldsymbol{I}_{n_\theta}\right)\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\theta}) - \left(\left.\frac{\partial \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})} \otimes \boldsymbol{I}_{n_\theta}\right)\boldsymbol{\Gamma}_i(\boldsymbol{\theta}) \right\| \xrightarrow{e.a.s.} 0, \tag{E.19}$$

where $\widehat{\boldsymbol{\Gamma}}_i(\boldsymbol{\theta})$ corresponds to $\boldsymbol{\Gamma}_i(\boldsymbol{\theta})$ with $\boldsymbol{\gamma}_i(\cdot)$ replaced by $\widehat{\boldsymbol{\gamma}}_i(\cdot)$. Consider (E.16) first. By examining the elements of the matrix, we obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} - \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\right\| \leq C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})-\boldsymbol{\gamma}_i(\boldsymbol{\theta})\right\| \xrightarrow{e.a.s.} 0$ as $i\to\infty$. Moreover, by applying the inequality $\|\boldsymbol{C}_1\boldsymbol{D}_1-\boldsymbol{C}_2\boldsymbol{D}_2\| \leq \|\boldsymbol{C}_1-\boldsymbol{C}_2\|\,\|\boldsymbol{D}_2\| + \|\boldsymbol{D}_1-\boldsymbol{D}_2\|\,\|\boldsymbol{C}_2\| + \|\boldsymbol{C}_1-\boldsymbol{C}_2\|\,\|\boldsymbol{D}_1-\boldsymbol{D}_2\|$ and the reasoning below (D.20), we can obtain (E.17), if (i) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta}) - \boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\right\| \xrightarrow{e.a.s.} 0$, (ii) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})} - \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\right\| \xrightarrow{e.a.s.} 0$, and (iii) $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\right\|^\kappa\Big) < \infty$ for some $\kappa > 0$. The first point is guaranteed by Proposition 4. For the second point, using a mean value theorem for vector-valued functions (see, e.g., Rudin, 1976, Theorem 9.19), we obtain

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\bigg|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})}-\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\right\|\\
&\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})-\boldsymbol{\gamma}_i(\boldsymbol{\theta})\right\|\\
&+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta})-\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\right\|\\
&+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i^{(1)}(\boldsymbol{\theta})-\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\right\|.
\end{aligned}
$$

By (D.11), (D.13), and (D.15), it is clear that the terms on the right-hand side above converge to zero e.a.s. Finally, $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\Big\|^\kappa\Big)<\infty$ follows from (D.3) (D.9), and (D.11). Therefore, (E.17) holds true. One can similarly obtain (E.18) using (D.9) and (D.15). Note that (E.19) is bounded by

$$
\begin{aligned}
&C\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i^{(2)}(\boldsymbol{\theta})-\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\right\|+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\left\|\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})-\boldsymbol{\gamma}_i(\boldsymbol{\theta})\right\|\right.\\
&\left.\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\right\|\right\}\leq C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i^{(2)}(\boldsymbol{\theta})-\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\right\|+C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})-\boldsymbol{\gamma}_i(\boldsymbol{\theta})\right\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\right\|.
\end{aligned}
$$

Since $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\widehat{\boldsymbol{\gamma}}_i^{(2)}(\boldsymbol{\theta})-\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\|\xrightarrow{e.a.s.}0$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\theta})-\boldsymbol{\gamma}_i(\boldsymbol{\theta})\|\xrightarrow{e.a.s.}0$, and the existence of a finite moment for $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\|$ (i.e., Eq. (D.17)), one has Eq. (E.19).

**Second term $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2\mathcal{L}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}-\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})\right\|\xrightarrow{a.s.}0$:** We use the same approach as in Lemma C.1. Specifically, we apply the uniform law of large numbers provided in White (1996, Theorem A.2.2) to the SE sequence $\left\{\frac{\partial^2\ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},\ i\in\mathbb{Z}\right\}$ for $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. This requires $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|\frac{\partial^2\ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big\|\Big)<\infty$. Since $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\Big\|\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big\|\Big)<\infty$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\Big\|\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}^\top}\Big\|\leq C+C\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\|$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\mathcal{G}}}\Big\|\frac{\partial^2\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\Big\|\leq C$, and

$$
\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\left(\frac{\partial\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\bigg|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\otimes\boldsymbol{I}_{n_\theta}\right)\boldsymbol{\Gamma}_i(\boldsymbol{\theta})\right\|\right)\leq C\,\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\right\|\right).
$$

From (E.15), we obtain $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|\frac{\partial^2\ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big\|\Big)<\infty$ given $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})\|^2\Big)<\infty$ and $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{\gamma}_i^{(2)}(\boldsymbol{\theta})\|\Big)<\infty$ hold. □

**Lemma E.3.** *Under the assumptions of Proposition 4, we have*

$$\sqrt{T}\left\|\frac{\partial\widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}-\frac{\partial\mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\|\xrightarrow{a.s.}0,\qquad as\quad T\to\infty. \tag{E.20}$$

***Proof of Lemma E.3.*** Similar to (E.13), by applying the argument below (C.2), it is sufficient to show that $\left\|\frac{\partial\ell_i(\widehat{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}-\frac{\partial\ell_i(\gamma_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|\xrightarrow{e.a.s.}0$, as $i\to\infty$. By (E.3), we have $\left\|\frac{\partial\ell_i(\widehat{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}-\frac{\partial\ell_i(\gamma_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|\leq\Delta^{\theta}_{\ell,1}+\Delta^{\theta}_{\ell,2}$, where

$$\Delta^{\theta}_{\ell,1}=\left\|\frac{\partial\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\bigg|_{\gamma=\widehat{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}-\frac{\partial\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\bigg|_{\gamma=\gamma_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|,$$
$$\Delta^{\theta}_{\ell,2}=\left\|\frac{\partial\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\bigg|_{\gamma=\widehat{\gamma}_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\frac{\partial\widehat{\gamma}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}-\frac{\partial\ell_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\bigg|_{\gamma=\gamma_i(\boldsymbol{\theta}_0),\,\boldsymbol{\theta}=\boldsymbol{\theta}_0}\frac{\partial\gamma_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|.$$

Using similar arguments as in the proof of Lemma E.2, along with Proposition 3 (which holds under the assumptions of Proposition 4), Proposition 4, and the fact that $\mathbb{E}\left(\left\|\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta}_0)\right\|^{\kappa}\right)<\infty$ for some $\kappa>0$ (see Eq. (D.9)), it follows immediately that $\Delta^{\theta}_{\ell,1}\xrightarrow{e.a.s.}0$ and $\Delta^{\theta}_{\ell,2}\xrightarrow{e.a.s.}0$, as $i\to\infty$. □

***Proof of Theorem 2.*** As in Blasques et al. (2022, Theorem 4.15), our proof builds on standard arguments for establishing the asymptotic normality of M-estimators; see, e.g., White (1996, Theorem 6.2) and Hayashi (2000, Proposition 7.8). Specifically, the asymptotic normality result in Theorem 2 holds if the following conditions are met:

AN1 the strong consistency of $\widehat{\boldsymbol{\theta}}_T\xrightarrow{a.s.}\boldsymbol{\theta}_0\in\text{int}(\boldsymbol{\Theta})\neq\emptyset$, as $T\to\infty$, where $\text{int}(\boldsymbol{\Theta})$ denotes the interior of $\boldsymbol{\Theta}$;

AN2 the a.s. twice continuous differentiability of $\widehat{\mathcal{L}}_T$ on $\boldsymbol{\Theta}$;

AN3 the asymptotic normality of the score, i.e., $\sqrt{T}\,\frac{\partial\widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\xrightarrow{d}\mathcal{N}\big(\mathbf{0},\boldsymbol{\mathcal{I}}_0\big)$, as $T\to\infty$;

AN4 the uniform convergence of the Hessian matrix $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2\widehat{\mathcal{L}}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}-\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})\right\|\xrightarrow{e.a.s.}0$, as $T\to\infty$, where $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})=\mathbb{E}\left(\frac{\partial^2\ell_i(\gamma_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right)$ is defined in Lemma E.2;

AN5 the information matrix equality $\boldsymbol{\mathcal{I}}_0=-\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_0)$ and the nonsingularity of $\boldsymbol{\mathcal{I}}_0$.

Condition AN1 follows from Theorem 1 and the assumption that $\boldsymbol{\theta}_0\in\text{int}(\boldsymbol{\Theta})$, where $\text{int}(\boldsymbol{\Theta})\neq\emptyset$ by Assumption AN4. AN2 holds under Assumptions AN1 and AN2, along with the expression in (E.15). Note that $\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\gamma_i^{(1)}(\boldsymbol{\theta})\right\|^2\right)<\infty$ and $\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\gamma_i^{(2)}(\boldsymbol{\theta})\right\|\right)<\infty$ follow from (D.9) and (D.17), respectively, provided that Assumption AN3 holds for $\kappa\geq 3$. Then, AN3 follows directly from Lemma E.1 and Lemma E.3, and AN4 is ensured by Lemma E.2. Furthermore, the information matrix equality in AN5 follows from standard textbook arguments. Specifically, one first establishes that the order of differentiation and integration for $p(\,\cdot\,|\,\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})$ and $\partial p(\,\cdot\,|\,\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})/\partial\boldsymbol{\theta}$ can be interchanged, and then shows that

$$\int p\big(\boldsymbol{y}\,\big|\,\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\frac{\partial \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\,\mathrm{d}\boldsymbol{y} = -\int p\big(\boldsymbol{y}\,\big|\,\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)\frac{\partial^2 \ell_i(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\,\mathrm{d}\boldsymbol{y}.$$

Evaluating the expression above at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ and applying the law of iterated expectations yields the information matrix equality. Finally, the nonsingularity of $\boldsymbol{\mathcal{I}}_0$ follows by assumption, completing the proof. □

# F Auxiliary lemmas

We establish some lemmas that are useful for the main proofs.

**Lemma F.4.** *Suppose $\boldsymbol{u}_i \sim t_\nu(\boldsymbol{0}, \boldsymbol{I}_N)$, where $N \in \mathbb{Z}^+$ and $\nu > 0$. We have*

$$\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu} \sim Beta\left(\frac{N}{2}, \frac{\nu}{2}\right), \qquad \frac{1}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu} \sim Beta\left(\frac{\nu}{2}, \frac{N}{2}\right). \tag{F.1}$$

*Moreover, for any non-random vectors $\boldsymbol{a}, \boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^N$, we have the following moment inequalities:*

$$\mathbb{E}\left[\boldsymbol{a}^\top \left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1} \frac{\boldsymbol{u}_i}{\sqrt{\nu}}\right]^2 = \frac{\nu}{(\nu+N)(\nu+N+2)} \|\boldsymbol{a}\|^2, \tag{F.2}$$

$$\mathbb{E}\left|\boldsymbol{a}_1^\top \left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1} \frac{\boldsymbol{u}_i \boldsymbol{u}_i^\top}{\nu} \boldsymbol{a}_2\right| \le \frac{1}{\nu+N} \|\boldsymbol{a}_1\|^{1/2} \|\boldsymbol{a}_2\|^{1/2}, \tag{F.3}$$

$$\mathbb{E}\left[\boldsymbol{a}_1^\top \left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1} \frac{\boldsymbol{u}_i \boldsymbol{u}_i^\top}{\nu} \boldsymbol{a}_2\right]^2 = \frac{1}{(\nu+N)(\nu+N+2)} \left[2\left(\boldsymbol{a}_1^\top \boldsymbol{a}_2\right)^2 + \|\boldsymbol{a}_1\|^2 \|\boldsymbol{a}_2\|^2\right]. \tag{F.4}$$

***Proof of Lemma F.4.*** For (F.1), see Harvey (2013, Proposition 39, p. 211).

Consider (F.2). Given that $\boldsymbol{u}_i/\sqrt{\nu}$ has a spherical distribution, we know $\left(1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu\right)^{-1} \boldsymbol{u}_i/\sqrt{\nu}$ also follows a spherical distribution. Theorem 2.3 of Fang et al. (2018) implies that

$$\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1} \frac{\boldsymbol{u}_i}{\sqrt{\nu}} \overset{d}{=} \left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu} \frac{1}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right)^{1/2} \boldsymbol{\mathcal{U}}_i, \tag{F.5}$$

where the notation "$\overset{d}{=}$" denotes equality in distribution, and $\boldsymbol{\mathcal{U}}_i$ is uniformly distributed on the unit sphere in $\mathbb{R}^N$, i.e., $\|\boldsymbol{\mathcal{U}}_i\| = 1$. Furthermore, $\boldsymbol{\mathcal{U}}_i$ is independent of the term inside curly brackets in (F.5). Since $\left(1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu\right)^{-1} \boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu \sim \text{Beta}(N/2, \nu/2)$ by (F.1), we obtain

$$\mathbb{E}\left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu} \frac{1}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right) = \mathbb{E}\left[\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu} \left(1-\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right)\right] = \frac{N\nu}{(\nu+N)(\nu+N+2)}.$$

Therefore, by Eq. (2.17) of Fang et al. (2018), we further obtain

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{a}^\top \left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1} \frac{\boldsymbol{u}_i}{\sqrt{\nu}}\right]^2 &= \mathbb{E}\left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu} \frac{1}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right) \mathbb{E}\left(\boldsymbol{a}^\top \boldsymbol{\mathcal{U}}_i \boldsymbol{\mathcal{U}}_i^\top \boldsymbol{a}\right) \\ &= \frac{\nu}{(\nu+N)(\nu+N+2)} \|\boldsymbol{a}\|^2.\end{aligned}$$

Next, consider (F.3). Similar to (F.5), the spherical property of $\left(1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu\right)^{-1/2} \boldsymbol{u}_i/\sqrt{\nu}$

implies that

$$\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1}\frac{\boldsymbol{u}_i\boldsymbol{u}_i^\top}{\nu} = \left[\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1/2}\frac{\boldsymbol{u}_i}{\sqrt{\nu}}\right]\left[\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1/2}\frac{\boldsymbol{u}_i}{\sqrt{\nu}}\right]^\top \overset{d}{=} \left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right)\tilde{\boldsymbol{\mathcal{U}}}_i\tilde{\boldsymbol{\mathcal{U}}}_i^\top, \tag{F.6}$$

where $\tilde{\boldsymbol{\mathcal{U}}}_i \in \mathbb{R}^N$ is uniformly distributed with $\|\tilde{\boldsymbol{\mathcal{U}}}_i\| = 1$, and is independent of $\left(1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu\right)^{-1}\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu$. Using (F.6) and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\mathbb{E}\left|\boldsymbol{a}_1^\top\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1}\frac{\boldsymbol{u}_i\boldsymbol{u}_i^\top}{\nu}\boldsymbol{a}_2\right| &= \mathbb{E}\left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right)\mathbb{E}\left|\boldsymbol{a}_1^\top\tilde{\boldsymbol{\mathcal{U}}}_i\tilde{\boldsymbol{\mathcal{U}}}_i^\top\boldsymbol{a}_2\right| \\
&\le \mathbb{E}\left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right)\left[\mathbb{E}\left(\boldsymbol{a}_1^\top\tilde{\boldsymbol{\mathcal{U}}}_i\right)^2\right]^{1/2}\left[\mathbb{E}\left(\boldsymbol{a}_2^\top\tilde{\boldsymbol{\mathcal{U}}}_i\right)^2\right]^{1/2} \\
&\le \frac{N}{\nu+N}\left(N^{-1}\|\boldsymbol{a}_1\|\right)^{1/2}\left(N^{-1}\|\boldsymbol{a}_2\|\right)^{1/2} = \frac{1}{\nu+N}\|\boldsymbol{a}_1\|^{1/2}\|\boldsymbol{a}_2\|^{1/2},
\end{aligned}$$

where the second-to-last step is due to $\mathbb{E}\left(\tilde{\boldsymbol{\mathcal{U}}}_i\tilde{\boldsymbol{\mathcal{U}}}_i^\top\right) = N^{-1}\boldsymbol{I}_N$ (see Fang et al., 2018, Eq. (2.17)).

Finally, consider (F.4). Let $a_{1,k}, a_{2,k}$, and $\tilde{\mathcal{U}}_{i,k}$ represent the $k_{\text{th}}$ elements of $\boldsymbol{a}_1, \boldsymbol{a}_2$, and $\tilde{\boldsymbol{\mathcal{U}}}_i$, respectively, for $k \in [\![ N ]\!]$. By (F.6), we then have

$$\begin{aligned}
\mathbb{E}\left[\left(\boldsymbol{a}_1^\top\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1}\frac{\boldsymbol{u}_i\boldsymbol{u}_i^\top}{\nu}\boldsymbol{a}_2\right)^2\right] &= \mathbb{E}\left(\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}{1+\boldsymbol{u}_i^\top \boldsymbol{u}_i/\nu}\right)^2\mathbb{E}\left(\boldsymbol{a}_1^\top\tilde{\boldsymbol{\mathcal{U}}}_i\tilde{\boldsymbol{\mathcal{U}}}_i^\top\boldsymbol{a}_2\right)^2 \\
&= \frac{N(N+2)}{(\nu+N)(\nu+N+2)}\sum_{j,k,\ell,m=1}^N a_{1,j}a_{1,\ell}a_{2,k}a_{2,m}\mathbb{E}\left(\tilde{\mathcal{U}}_{i,j}\tilde{\mathcal{U}}_{i,k}\tilde{\mathcal{U}}_{i,\ell}\tilde{\mathcal{U}}_{i,m}\right).
\end{aligned} \tag{F.7}$$

By Song and Gupta (1997, Theorem 2.1), is it not hard to drive: $\mathbb{E}\left(\tilde{\mathcal{U}}_{i,j}^4\right) = 3\left[N(N+2)\right]^{-1}$, $\mathbb{E}\left(\tilde{\mathcal{U}}_{i,j}^2\tilde{\mathcal{U}}_{i,k}^2\right) = \left[N(N+2)\right]^{-1}$, $\mathbb{E}\left(\tilde{\mathcal{U}}_{i,j}\tilde{\mathcal{U}}_{i,k}^3\right) = \mathbb{E}\left(\tilde{\mathcal{U}}_{i,j}\tilde{\mathcal{U}}_{i,k}\tilde{\mathcal{U}}_{i,\ell}\tilde{\mathcal{U}}_{i,m}\right) = 0$, for $j \ne k \ne \ell \ne m$. That is,

$$\mathbb{E}\left(\tilde{\mathcal{U}}_{i,j}\tilde{\mathcal{U}}_{i,k}\tilde{\mathcal{U}}_{i,\ell}\tilde{\mathcal{U}}_{i,m}\right) = \frac{1}{N(N+2)}\left(\delta_{jk}\delta_{\ell m}+\delta_{j\ell}\delta_{km}+\delta_{jm}\delta_{k\ell}\right), \qquad j,k,\ell,m \in [\![ N ]\!], \tag{F.8}$$

where $\delta_{pq}$ is the Kronecker delta, i.e., for any $p, q \in \mathbb{Z}^+$, $\delta_{pq} = 1$ if $p = q$ and 0 otherwise. Therefore,

$$\begin{aligned}
&E\left[\boldsymbol{a}_1^\top\left(1+\frac{\boldsymbol{u}_i^\top \boldsymbol{u}_i}{\nu}\right)^{-1}\frac{\boldsymbol{u}_i\boldsymbol{u}_i^\top}{\nu}\boldsymbol{a}_2\right]^2 \\
&= \frac{N(N+2)}{(\nu+N)(\nu+N+2)}\frac{1}{N(N+2)}\sum_{j,k,\ell,m=1}^N a_{1,j}a_{1,\ell}a_{2,k}a_{2,m}\left(\delta_{jk}\delta_{\ell m}+\delta_{j\ell}\delta_{km}+\delta_{jm}\delta_{k\ell}\right) \\
&= \frac{1}{(\nu+N)(\nu+N+2)}\left[2\left(\sum_{j=1}^N a_{1,j}a_{2,j}\right)^2+\sum_{j=1}^N a_{1,j}^2\sum_{j=1}^N a_{2,j}^2\right].
\end{aligned} \tag{F.9}$$

Observe that the expression within the square brackets simplifies to $2\big(\boldsymbol{a}_1^\top \boldsymbol{a}_2\big)^2 + \|\boldsymbol{a}_1\|^2\|\boldsymbol{a}_2\|^2$. This completes the proof. □

**Lemma F.5.** *For any matrix $\boldsymbol{A}$, let $\mathrm{row}_j(\boldsymbol{A})$ denote its $j_{th}$ row. Under Assumption A2, for $N \in \mathbb{Z}^+$ and $j \in [\![ N ]\!]$, we have*

$$\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\| \mathrm{row}_j \left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right)\right\| \le \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right\|_{2,\infty} < \infty, \tag{F.10}$$

$$\sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\| \mathrm{row}_j \left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right)\right\| \le \sup_{\boldsymbol{\nu}\in\boldsymbol{\Theta}_\nu} \left\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\right\|_{2,\infty} < \infty. \tag{F.11}$$

***Proof of Lemma F.5.*** Note that $\|\boldsymbol{A}\|_{2,\infty} = \max_{1\le j\le N} \|\mathrm{row}_j(\boldsymbol{A})\|$ for any matrix $\boldsymbol{A}$ with $N$ rows. Moreover, $\|\boldsymbol{B}\|_{2,\infty} \le \|\boldsymbol{B}\|$ for any matrix $\boldsymbol{B}$. Utilizing the identity $\left\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{1/2}\right\| = \left\|\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)\right\|^{1/2}$ for any $\boldsymbol{\nu}_2$ and Assumption A2, one can immediately obtain Lemma F.5. □

Next, we summarize crucial rules of the vectorization operator and matrix derivatives, which will be repeatedly used in the subsequent proofs.

**Lemma F.6.** *Let $k, l, m, n, p \in \mathbb{Z}^+$. Moreover, $\|\cdot\|_F$ denotes the Frobenius norm.*
***(i)*** *For any $\boldsymbol{A} \in \mathbb{R}^{l\times m}$, $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, and $\boldsymbol{b} \in \mathbb{R}^{n\times 1}$,*

$$\mathrm{vec}(\boldsymbol{A}\boldsymbol{B}) = (\boldsymbol{I}_n \otimes \boldsymbol{A})\,\mathrm{vec}(\boldsymbol{B}) = (\boldsymbol{B}^\top \otimes \boldsymbol{I}_l)\,\mathrm{vec}(\boldsymbol{A}), \tag{F.12}$$

$$\mathrm{vec}(\boldsymbol{b}^\top \otimes \boldsymbol{A}) = \boldsymbol{b} \otimes \mathrm{vec}(\boldsymbol{A}). \tag{F.13}$$

*Moreover, if $\boldsymbol{C} = \big(\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_m\big)$, where $\boldsymbol{C}_j \in \mathbb{R}^{l\times k_j}$ with $k_j \in \mathbb{Z}^+$ for any $j \in [\![ m ]\!]$, then*

$$\mathrm{vec}(\boldsymbol{C}) = \left(\mathrm{vec}(\boldsymbol{C}_1)^\top, \mathrm{vec}(\boldsymbol{C}_2)^\top, \ldots, \mathrm{vec}(\boldsymbol{C}_m)^\top\right)^\top. \tag{F.14}$$

***(ii-1)*** *Consider $\boldsymbol{A}(\boldsymbol{x}) = \big(\boldsymbol{a}_1(\boldsymbol{x}), \ldots, \boldsymbol{a}_l(\boldsymbol{x})\big)^\top \in \mathbb{R}^{l\times m}$, $\boldsymbol{b}(\boldsymbol{x}) \in \mathbb{R}^{m\times 1}$, $\boldsymbol{a}(\boldsymbol{x}) \in \mathbb{R}^{l\times 1}$, where $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$. Let $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times m}$ be symmetric. Then,*

$$\frac{\partial\big(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{b}(\boldsymbol{x})\big)}{\partial \boldsymbol{x}^\top} = \big(\boldsymbol{b}(\boldsymbol{x})^\top \otimes \boldsymbol{I}_l\big)\frac{\partial\,\mathrm{vec}\left(\boldsymbol{A}(\boldsymbol{x})\right)}{\partial \boldsymbol{x}^\top} + \boldsymbol{A}(\boldsymbol{x})\frac{\partial \boldsymbol{b}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top}, \tag{F.15}$$

$$\frac{\partial}{\partial \boldsymbol{x}^\top}\big(\boldsymbol{b}(\boldsymbol{x}) \otimes \boldsymbol{a}(\boldsymbol{x})\big) = \big(\boldsymbol{I}_m \otimes \boldsymbol{a}(\boldsymbol{x})\big)\frac{\partial \boldsymbol{b}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top} + \big(\boldsymbol{b}(\boldsymbol{x}) \otimes \boldsymbol{I}_l\big)\frac{\partial \boldsymbol{a}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top}, \tag{F.16}$$

$$\frac{\partial\big(\boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\Lambda}\boldsymbol{b}(\boldsymbol{x})\big)}{\partial \boldsymbol{x}^\top} = 2\boldsymbol{b}(\boldsymbol{x})^\top \boldsymbol{\Lambda}\frac{\partial \boldsymbol{b}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top}. \tag{F.17}$$

$$\frac{\partial \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{A}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left(\frac{\partial \boldsymbol{a}(\boldsymbol{x})^\top}{\partial \boldsymbol{x}}\right)\boldsymbol{A}(\boldsymbol{x}) + \big(\boldsymbol{a}(\boldsymbol{x})^\top \otimes \boldsymbol{I}_n\big)\boldsymbol{\mathcal{A}}(\boldsymbol{x}), \tag{F.18}$$

*where* $\boldsymbol{\mathcal{A}}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial \boldsymbol{a}_1(\boldsymbol{x})^\top}{\partial \boldsymbol{x}} \\ \vdots \\ \frac{\partial \boldsymbol{a}_l(\boldsymbol{x})^\top}{\partial \boldsymbol{x}} \end{pmatrix}$ *with* $\|\boldsymbol{\mathcal{A}}(\boldsymbol{x})\| \leq \left\| \frac{\partial \operatorname{vec}(\boldsymbol{A}(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \right\|_F$.

***(ii-2)*** *For* $\boldsymbol{C}(\boldsymbol{x}) = \big(\boldsymbol{C}_1(\boldsymbol{x})^\top, \ldots, \boldsymbol{C}_m(\boldsymbol{x})^\top\big)^\top$, *where* $\boldsymbol{C}_j(\boldsymbol{x}) \in \mathbb{R}^{l \times k}$ *with* $j \in [\![ m ]\!]$, *we have*

$$\frac{\partial}{\partial \boldsymbol{x}^\top} \operatorname{vec}\Big(\big(\boldsymbol{b}(\boldsymbol{x})^\top \otimes \boldsymbol{I}_l\big)\boldsymbol{C}(\boldsymbol{x})\Big) = \big(\boldsymbol{b}(\boldsymbol{x})^\top \otimes \boldsymbol{I}_{lk}\big)\boldsymbol{\mathcal{C}}_1(\boldsymbol{x}) + \boldsymbol{\mathcal{C}}_2(\boldsymbol{x})\frac{\partial \boldsymbol{b}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top}, \tag{F.19}$$

*with* $\|\boldsymbol{\mathcal{C}}_1(\boldsymbol{x})\| \leq \left\| \frac{\partial \operatorname{vec}(\boldsymbol{C}(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \right\|_F$ *and* $\|\boldsymbol{\mathcal{C}}_2(\boldsymbol{x})\| \leq \|\boldsymbol{C}(\boldsymbol{x})\|_F$ , *where*

$$\boldsymbol{\mathcal{C}}_1(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial \operatorname{vec}(\boldsymbol{C}_1(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \\ \frac{\partial \operatorname{vec}(\boldsymbol{C}_2(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \\ \vdots \\ \frac{\partial \operatorname{vec}(\boldsymbol{C}_m(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \end{pmatrix}, \quad \boldsymbol{\mathcal{C}}_2(\boldsymbol{x}) = \Big(\operatorname{vec}\big(\boldsymbol{C}_1(\boldsymbol{x})\big), \operatorname{vec}\big(\boldsymbol{C}_2(\boldsymbol{x})\big), \ldots, \operatorname{vec}\big(\boldsymbol{C}_m(\boldsymbol{x})\big)\Big). \tag{F.20}$$

*We see that* $\boldsymbol{\mathcal{C}}_1(\boldsymbol{x}) \in \mathbb{R}^{lkm \times n}$ *and* $\boldsymbol{\mathcal{C}}_2(\boldsymbol{x}) \in \mathbb{R}^{lk \times m}$. *Due to the identity* $\|\boldsymbol{X}_1 \otimes \boldsymbol{X}_2\| = \|\boldsymbol{X}_1\|\,\|\boldsymbol{X}_2\|$ *for any matrices* $\boldsymbol{X}_1, \boldsymbol{X}_2$ *and the property* $\|\cdot\|_F \leq \sqrt{\operatorname{rank}(\cdot)}\,\|\cdot\|$, (F.19) *further implies*

$$\left\| \frac{\partial}{\partial \boldsymbol{x}^\top} \operatorname{vec}\Big(\big(\boldsymbol{b}(\boldsymbol{x})^\top \otimes \boldsymbol{I}_l\big)\boldsymbol{C}(\boldsymbol{x})\Big) \right\| \leq C\left( \|\boldsymbol{b}(\boldsymbol{x})\| \left\| \frac{\partial \operatorname{vec}(\boldsymbol{C}(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \right\| + \left\| \frac{\partial \boldsymbol{b}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top} \right\| \|\boldsymbol{C}(\boldsymbol{x})\| \right). \tag{F.21}$$

*Similarly, for* $\boldsymbol{B}(\boldsymbol{x}) \in \mathbb{R}^{p \times m}$, *one can obtain*

$$\begin{aligned} &\left\| \frac{\partial}{\partial \boldsymbol{x}^\top} \operatorname{vec}\Big(\big(\boldsymbol{B}(\boldsymbol{x}) \otimes \boldsymbol{I}_l\big)\boldsymbol{C}(\boldsymbol{x})\Big) \right\| \\ &\qquad \leq C\left( \|\boldsymbol{B}(\boldsymbol{x})\| \left\| \frac{\partial \operatorname{vec}(\boldsymbol{C}(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} \right\| + \left\| \frac{\partial \operatorname{vec}\big(\boldsymbol{B}(\boldsymbol{x})\big)}{\partial \boldsymbol{x}^\top} \right\| \|\boldsymbol{C}(\boldsymbol{x})\| \right). \end{aligned} \tag{F.22}$$

***Proof of Lemma F.6.*** Note that (F.12), (F.15), and (F.17) are provided in Lütkepohl (2005, on pp. 662, 666, and 667, respectively). Furthermore, (F.22) follows directly from the identity

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{x}^\top} \operatorname{vec}\Big(\big(\boldsymbol{B}(\boldsymbol{x}) \otimes \boldsymbol{I}_l\big)\boldsymbol{C}(\boldsymbol{x})\Big) \\ &\qquad = \big(\boldsymbol{I}_k \otimes \boldsymbol{B}(\boldsymbol{x}) \otimes \boldsymbol{I}_l\big)\frac{\partial \operatorname{vec}(\boldsymbol{C}(\boldsymbol{x}))}{\partial \boldsymbol{x}^\top} + \big(\boldsymbol{C}(\boldsymbol{x})^\top \otimes \boldsymbol{I}_{lp}\big)\big(\boldsymbol{I}_m \otimes \boldsymbol{G}\big)\frac{\partial \operatorname{vec}\big(\boldsymbol{B}(\boldsymbol{x})\big)}{\partial \boldsymbol{x}^\top}, \end{aligned} \tag{F.23}$$

where $\boldsymbol{G} = \big(\boldsymbol{K}_{l,p} \otimes \boldsymbol{I}_l\big)\big(\boldsymbol{I}_p \otimes \operatorname{vec}(\boldsymbol{I}_l)\big)$ and $\boldsymbol{K}_{l,p}$ is some $lp \times lp$ commutation matrix. The remaining steps of the proof rely on straightforward linear algebra and are therefore omitted. □

The following lemma compiles key derivative results that are critical for examining the asymptotic behavior of first- and second-order derivative processes in Appendix D.

**Lemma F.7.** *Recall* $\boldsymbol{e}_i(\boldsymbol{\gamma}) = \boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma})\big)$ *from* (2.14), *and further define* $\boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma}) = \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\big(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\gamma})\big)$. *Then we can write*

$$w_i(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \frac{1 + \nu_1^{-1} N}{1 + \nu_1^{-1}\, \boldsymbol{e}_i(\boldsymbol{\gamma})^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma})} = \frac{1 + \nu_1^{-1} N}{1 + \nu_1^{-1}\, \boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma})^\top \boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma})}.$$

*For convenience, we introduce the following notation for the subsequent proofs:*

*(a) For* $i \in \mathbb{Z}$, *let* $\boldsymbol{s}_i(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \boldsymbol{S}(\boldsymbol{\gamma}, \boldsymbol{\nu})\nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})$, *where* $\nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \nabla_i^\mu(\boldsymbol{\gamma}, \boldsymbol{\nu}) + \nabla_i^\sigma(\boldsymbol{\gamma}, \boldsymbol{\nu})$, *with* $\nabla_i^\mu(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \nabla^\mu(\boldsymbol{Y}_i, \boldsymbol{\gamma}, \boldsymbol{\nu})$ *and* $\nabla_i^\sigma(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \nabla^\sigma(\boldsymbol{Y}_i, \boldsymbol{\gamma}, \boldsymbol{\nu})$.

*(b) Moreover, let* $\kappa_{1,i}(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \dfrac{\nu_1^{-1}\boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma})^\top \boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma})}{1 + \nu_1^{-1}\boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma})^\top \boldsymbol{u}_i(\boldsymbol{\nu}_2, \boldsymbol{\gamma})}$ *and* $\kappa_{2,i}(\boldsymbol{\gamma}, \boldsymbol{\nu}) = 1 - \kappa_{1,i}(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \dfrac{\nu_1}{\nu_1 + N} w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})$.

*Before proceeding, we express* $\nabla_i^\mu(\boldsymbol{\gamma}, \boldsymbol{\nu})$ *and* $\nabla_i^\sigma(\boldsymbol{\gamma}, \boldsymbol{\nu})$ *from their compact forms in* (2.14) *by the following expressions in the proofs below:*

$$\nabla_i^\mu(\boldsymbol{\gamma}, \boldsymbol{\nu}) = w_i(\boldsymbol{\gamma}, \boldsymbol{\nu})\big(\boldsymbol{\Phi}^\mu(\boldsymbol{t}_1; \boldsymbol{\gamma}), \ldots, \boldsymbol{\Phi}^\mu(\boldsymbol{t}_N; \boldsymbol{\gamma})\big)\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma}), \tag{F.24}$$

$$\nabla_i^\sigma(\boldsymbol{\gamma}, \boldsymbol{\nu}) = w_i(\boldsymbol{\gamma}, \boldsymbol{\nu})\big(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_1; \boldsymbol{\gamma}), \ldots, \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_N; \boldsymbol{\gamma})\big)\,\mathrm{diag}(\boldsymbol{e}_i(\boldsymbol{\gamma}))\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma}) - \sum_{j=1}^{N}\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j; \boldsymbol{\gamma}), \tag{F.25}$$

*where*

$$\boldsymbol{\Phi}^\mu(\boldsymbol{t}; \boldsymbol{\gamma}) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\dot{\boldsymbol{g}}_\mu(\boldsymbol{f}) \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)\,\dot{\boldsymbol{g}}_\mu(\boldsymbol{f}), \tag{F.26}$$

$$\boldsymbol{\Phi}^\sigma(\boldsymbol{t}; \boldsymbol{\gamma}) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}) \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big) = \frac{1}{g_\sigma(\boldsymbol{f})}\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)\,\dot{\boldsymbol{g}}_\sigma(\boldsymbol{f}), \tag{F.27}$$

*as previously defined in Appendix B, with* $\boldsymbol{f}$ *denoting* $\boldsymbol{\Gamma}\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})$, *and* $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{\Gamma}^\top)$. *This implies*

$$\frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t}; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} = \Big[g_\sigma(\boldsymbol{f})^{-1}\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)\boldsymbol{\mathcal{M}}(\boldsymbol{f})\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top\big)\Big]_{\boldsymbol{f} = (\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top)\boldsymbol{\gamma}}, \tag{F.28}$$

$$\frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} = \Big[g_\sigma(\boldsymbol{f})^{-1}\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)\boldsymbol{\mathcal{S}}(\boldsymbol{f})\big(\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top\big)\Big]_{\boldsymbol{f} = (\boldsymbol{I}_{n_f} \otimes \boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top)\boldsymbol{\gamma}}, \tag{F.29}$$

*where* $\boldsymbol{\mathcal{M}}(\boldsymbol{f}) = \dfrac{\partial^2 g_\mu(\boldsymbol{f})}{\partial \boldsymbol{f}\partial \boldsymbol{f}^\top} - g_\sigma(\boldsymbol{f})^{-1}\dfrac{\partial g_\mu(\boldsymbol{f})}{\partial \boldsymbol{f}}\dfrac{\partial g_\sigma(\boldsymbol{f})}{\partial \boldsymbol{f}^\top}$ *and* $\boldsymbol{\mathcal{S}}(\boldsymbol{f}) = \dfrac{\partial^2 g_\sigma(\boldsymbol{f})}{\partial \boldsymbol{f}\partial \boldsymbol{f}^\top} - g_\sigma(\boldsymbol{f})^{-1}\dfrac{\partial g_\sigma(\boldsymbol{f})}{\partial \boldsymbol{f}}\dfrac{\partial g_\sigma(\boldsymbol{f})}{\partial \boldsymbol{f}^\top}$.

*Furthermore, let* $\left(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\right)^{\otimes 2}=\left(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\right)\otimes\left(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\right)$. *Then, we have*

$$\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\boldsymbol{\Phi}^\mu(\boldsymbol{t};\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^\top}\right)=\Bigg[g_\sigma(\boldsymbol{f})^{-1}\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\otimes d}(\boldsymbol{t})\big)^{\otimes 2}\Bigg(\frac{\partial\operatorname{vec}(\boldsymbol{\mathcal{M}}(\boldsymbol{f}))}{\partial\boldsymbol{f}^\top}\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\otimes d}(\boldsymbol{t})^\top\big)$$

$$-\operatorname{vec}\big(\boldsymbol{\mathcal{M}}(\boldsymbol{f})\big)\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})^\top\Bigg)\Bigg]_{\boldsymbol{f}=(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\otimes d}(\boldsymbol{t})^\top)\boldsymbol{\gamma}}, \tag{F.30}$$

*and*

$$\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^\top}\right)=\Bigg[g_\sigma(\boldsymbol{f})^{-1}\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})\big)^{\otimes 2}\Bigg(\frac{\partial\operatorname{vec}(\boldsymbol{\mathcal{S}}(\boldsymbol{f}))}{\partial\boldsymbol{f}^\top}\big(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top\big)$$

$$-\operatorname{vec}\big(\boldsymbol{\mathcal{S}}(\boldsymbol{f})\big)\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})^\top\Bigg)\Bigg]_{\boldsymbol{f}=(\boldsymbol{I}_{n_f}\otimes\boldsymbol{\phi}_{\boldsymbol{K}}(\boldsymbol{t})^\top)\boldsymbol{\gamma}}. \tag{F.31}$$

*These identities will be repeatedly used later.*

***(i)*** *Applying* (F.15) *twice and using the chain rule (see, e.g.,* Lütkepohl, 2005, *Proposition A.1) on* (2.9)*, we obtain*

$$\frac{\partial\boldsymbol{\gamma}_{i+1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}=\frac{\partial\boldsymbol{\omega}}{\partial\boldsymbol{\theta}^\top}+\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta})^\top\otimes\boldsymbol{I}_{n_\gamma}\big)\frac{\partial\operatorname{vec}(\boldsymbol{B})}{\partial\boldsymbol{\theta}^\top}+\Big(\boldsymbol{s}_i\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\nu}\big)^\top\otimes\boldsymbol{I}_{n_\gamma}\Big)\frac{\partial\operatorname{vec}(\boldsymbol{A})}{\partial\boldsymbol{\theta}^\top}$$

$$+\boldsymbol{A}\left.\frac{\partial\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}+\left(\boldsymbol{B}+\boldsymbol{A}\left.\frac{\partial\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_i(\boldsymbol{\theta})}\right)\frac{\partial\boldsymbol{\gamma}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}, \tag{F.32}$$

*Applying* $\operatorname{vec}(\cdot)$ *to both sides of Eq.* (F.32)*, and subsequently using* (F.12)*, we arrive at*

$$\boldsymbol{\gamma}_{i+1}^{(1)}(\boldsymbol{\theta})=\boldsymbol{Q}_{(1),i}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\big)\,\boldsymbol{\gamma}_i^{(1)}(\boldsymbol{\theta})+\boldsymbol{q}_{(1),i}\big(\boldsymbol{\gamma}_i(\boldsymbol{\theta}),\boldsymbol{\theta}\big), \tag{F.33}$$

*where* $\boldsymbol{Q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})=\boldsymbol{I}_{n_\theta}\otimes\left(\boldsymbol{B}+\boldsymbol{A}\,\frac{\partial\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top}\right)$ *and*

$$\boldsymbol{q}_{(1),i}(\boldsymbol{\gamma},\boldsymbol{\theta})=\operatorname{vec}\left(\frac{\partial\boldsymbol{\omega}}{\partial\boldsymbol{\theta}^\top}+\big(\boldsymbol{\gamma}^\top\otimes\boldsymbol{I}_{n_\gamma}\big)\frac{\partial\operatorname{vec}(\boldsymbol{B})}{\partial\boldsymbol{\theta}^\top}+\Big(\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top\otimes\boldsymbol{I}_{n_\gamma}\Big)\frac{\partial\operatorname{vec}(\boldsymbol{A})}{\partial\boldsymbol{\theta}^\top}+\boldsymbol{A}\,\frac{\partial\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right)$$

$$=\operatorname{vec}\left(\frac{\partial\boldsymbol{\omega}}{\partial\boldsymbol{\theta}^\top}\right)+\left(\frac{\partial\operatorname{vec}(\boldsymbol{B})^\top}{\partial\boldsymbol{\theta}}\otimes\boldsymbol{I}_{n_\gamma}\right)\big(\boldsymbol{\gamma}\otimes\operatorname{vec}(\boldsymbol{I}_{n_\gamma})\big)$$

$$+\left(\frac{\partial\operatorname{vec}(\boldsymbol{A})^\top}{\partial\boldsymbol{\theta}}\otimes\boldsymbol{I}_{n_\gamma}\right)\Big(\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})\otimes\operatorname{vec}(\boldsymbol{I}_{n_\gamma})\Big)+\operatorname{vec}\left(\boldsymbol{A}\,\frac{\partial\boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\right). \tag{F.34}$$

*The second equality above follows from the linearity of* $\operatorname{vec}(\cdot)$ *and the identity* (F.13).

***(ii)*** *Using Eqs.* (F.15) *and* (F.33)*, along with the chain rule, we obtain*

$$\frac{\partial \gamma_{i+1}^{(1)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \left(\gamma_i^{(1)}(\boldsymbol{\theta})^\top \otimes \boldsymbol{I}_{n_\gamma n_\theta}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{Q}_{(1),i}(\gamma_i(\boldsymbol{\theta}), \boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}^\top} + \boldsymbol{Q}_{(1),i}(\gamma_i(\boldsymbol{\theta}), \boldsymbol{\theta}) \frac{\partial \gamma_i^{(1)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \left.\frac{\partial \boldsymbol{q}_{(1),i}(\gamma, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\gamma=\gamma_i(\boldsymbol{\theta})} + \left.\frac{\partial \boldsymbol{q}_{(1),i}(\gamma, \boldsymbol{\theta})}{\partial \gamma^\top}\right|_{\gamma=\gamma_i(\boldsymbol{\theta})} \frac{\partial \gamma_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}, \tag{F.35}$$

*where, by the identities* (F.15), (F.16), *and* (F.34),

$$\frac{\partial \boldsymbol{q}_{(1),i}(\gamma, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \left(\frac{\partial \operatorname{vec}(\boldsymbol{A})^\top}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_{n_\gamma}\right)\left(\boldsymbol{I}_{n_\gamma} \otimes \operatorname{vec}(\boldsymbol{I}_{n_\gamma})\right) \frac{\partial \boldsymbol{s}_i(\gamma, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top} + \left(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{A}\right) \frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\gamma, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) + \left(\frac{\partial \boldsymbol{s}_i(\gamma, \boldsymbol{\nu})^\top}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_{n_\gamma}\right) \frac{\partial \operatorname{vec}(\boldsymbol{A})}{\partial \boldsymbol{\theta}^\top}, \tag{F.36}$$

$$\frac{\partial \boldsymbol{q}_{(1),i}(\gamma, \boldsymbol{\theta})}{\partial \gamma^\top} = \left(\frac{\partial \operatorname{vec}(\boldsymbol{B})^\top}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_{n_\gamma}\right)\left(\boldsymbol{I}_{n_\gamma} \otimes \operatorname{vec}(\boldsymbol{I}_{n_\gamma})\right) + \left(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{A}\right) \frac{\partial}{\partial \gamma^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\gamma, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) + \left(\frac{\partial \operatorname{vec}(\boldsymbol{A})^\top}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_{n_\gamma}\right)\left(\boldsymbol{I}_{n_\gamma} \otimes \operatorname{vec}(\boldsymbol{I}_{n_\gamma})\right) \frac{\partial \boldsymbol{s}_i(\gamma, \boldsymbol{\nu})}{\partial \gamma^\top}. \tag{F.37}$$

*Applying* vec(·) *to both sides of Eq.* (F.35)*, and subsequently using* (F.12)*, implies*

$$\gamma_{i+1}^{(2)}(\boldsymbol{\theta}) = \boldsymbol{Q}_{(2),i}(\gamma_i(\boldsymbol{\theta}), \boldsymbol{\theta})\, \gamma_i^{(2)}(\boldsymbol{\theta}) + \boldsymbol{q}_{(2),i}(\gamma_i^{(1)}(\boldsymbol{\theta}), \gamma_i(\boldsymbol{\theta}), \boldsymbol{\theta}), \tag{F.38}$$

*where* $\boldsymbol{Q}_{(2),i}(\gamma, \boldsymbol{\theta}) = \boldsymbol{I}_{n_\theta^2} \otimes \left(\boldsymbol{B} + \boldsymbol{A}\,\frac{\partial \boldsymbol{s}_i(\gamma, \boldsymbol{\nu})}{\partial \gamma^\top}\right)$ *and*

$$\boldsymbol{q}_{(2),i}(\gamma_i^{(1)}(\boldsymbol{\theta}), \gamma_i(\boldsymbol{\theta}), \boldsymbol{\theta}) = \operatorname{vec}\left(\left(\gamma_i^{(1)}(\boldsymbol{\theta})^\top \otimes \boldsymbol{I}_{n_\gamma n_\theta}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{Q}_{(1),i}(\gamma_i(\boldsymbol{\theta}), \boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}^\top}\right) + \operatorname{vec}\left(\left.\frac{\partial \boldsymbol{q}_{(1),i}(\gamma, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right|_{\gamma=\gamma_i(\boldsymbol{\theta})}\right) + \left(\boldsymbol{I}_{n_\theta} \otimes \left.\frac{\partial \boldsymbol{q}_{(1),i}(\gamma, \boldsymbol{\theta})}{\partial \gamma^\top}\right|_{\gamma=\gamma_i(\boldsymbol{\theta})}\right) \gamma_i^{(1)}(\boldsymbol{\theta}). \tag{F.39}$$

***(iii-1)*** $\frac{\partial w_i(\gamma_i, \boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top} = \left(\frac{\partial w_i(\gamma_i, \boldsymbol{\nu})}{\partial \nu_1}, \frac{\partial w_i(\gamma_i, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}_2^\top}, \mathbf{0}_{(2n_\gamma+1)n_\gamma}^\top\right)$*, where*

$$\frac{\partial w_i(\gamma_i, \boldsymbol{\nu})}{\partial \nu_1} = \nu_1^{-1} \kappa_{2,i}(\gamma, \boldsymbol{\nu})\left(\kappa_{1,i}(\gamma, \boldsymbol{\nu}) - \nu_1^{-1} N \kappa_{2,i}(\gamma, \boldsymbol{\nu})\right),$$

$$\frac{\partial w_i(\gamma_i, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}_2^\top} = -\kappa_{2,i}(\gamma, \boldsymbol{\nu}) w_i(\gamma_i, \boldsymbol{\nu})\left(\nu_1^{-1/2} \boldsymbol{e}_i(\gamma)^\top \otimes \nu_1^{-1/2} \boldsymbol{e}_i(\gamma)^\top\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial \boldsymbol{\nu}_2^\top}.$$

***(iii-2)*** *Furthermore,* $\frac{\partial^2 w_i(\gamma_i, \boldsymbol{\nu})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \operatorname{diag}\left(\boldsymbol{W}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{(2)}, \mathbf{O}\right)$*, where* $\boldsymbol{W}_{\boldsymbol{\nu}\boldsymbol{\nu}}^{(2)} = \begin{pmatrix} W_{\nu_1\nu_1}^{(2)} & W_{\nu_1\nu_2}^{(2)} \\ W_{\nu_2\nu_1}^{(2)} & W_{\nu_2\nu_2}^{(2)} \end{pmatrix}$*, with*

$$W_{\nu_1\nu_1}^{(2)} = \frac{\partial^2 w_i(\gamma_i, \boldsymbol{\nu})}{\partial \nu_1^2} = -2\nu_1^{-2} \kappa_{1,i}(\gamma, \boldsymbol{\nu}) \kappa_{2,i}(\gamma, \boldsymbol{\nu})^2 + 2\nu_1^{-3} N \kappa_{2,i}(\gamma, \boldsymbol{\nu})\left(1 - \kappa_{1,i}(\gamma, \boldsymbol{\nu}) \kappa_{2,i}(\gamma, \boldsymbol{\nu})\right),$$

$W^{(2)}_{\boldsymbol{\nu}_2\nu_1} = \left(W^{(2)}_{\nu_1\boldsymbol{\nu}_2}\right)^\top$, *and*

$$
\begin{aligned}
W^{(2)}_{\nu_1\boldsymbol{\nu}_2} = \frac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\nu_1\partial\boldsymbol{\nu}_2^\top} &= \nu_1^{-1}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})^2\Big(2\nu_1^{-1}N\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu}) + 2\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu}) - 1\Big) \\
&\quad \times \left(\nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})^\top\right)\frac{\partial\,\mathrm{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial\boldsymbol{\nu}_2^\top}, \\
W^{(2)}_{\boldsymbol{\nu}_2\boldsymbol{\nu}_2} = \frac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\nu}_2\partial\boldsymbol{\nu}_2^\top} &= 2\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\frac{\partial\,\mathrm{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)^\top}{\partial\boldsymbol{\nu}_2} \\
&\quad \times \left(\nu_1^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma})\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \nu_1^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma})\boldsymbol{e}_i(\boldsymbol{\gamma})^\top\right)\frac{\partial\,\mathrm{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial\boldsymbol{\nu}_2^\top} \\
&\quad - \kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\Big(\nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{I}_{n_\delta}\Big)\boldsymbol{\mathcal{X}}^{\theta\theta}(\boldsymbol{\nu}_2).
\end{aligned}
$$

*The expression of $W^{(2)}_{\boldsymbol{\nu}_2\boldsymbol{\nu}_2}$ is obtained using (F.18). Here, we omit the details of the construction of $\boldsymbol{\mathcal{X}}^{\theta\theta}(\boldsymbol{\nu}_2)$; however, with the norm equivalence in finite dimensional spaces and Assumption AN1, it follows that $\left|\boldsymbol{\mathcal{X}}^{\theta\theta}(\boldsymbol{\nu}_2)\right| \le C$.*

***(iii-3)*** *Next, by employing (F.17), we have*

$$
\begin{aligned}
\frac{\partial w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top} &= -2\nu_1^{-1}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\frac{\partial\boldsymbol{e}_i(\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^\top} \\
&= 2\nu_1^{-1}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top, \qquad \text{(F.40)}
\end{aligned}
$$

*where* $\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) = \dot{\boldsymbol{G}}_\mu(\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1} + \dot{\boldsymbol{G}}_\sigma(\boldsymbol{\gamma})^\top\boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}\,\mathrm{diag}\left(\boldsymbol{e}_i(\boldsymbol{\gamma})\right)$.

***(iii-4)*** *Furthermore,* $\dfrac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top\partial\boldsymbol{\theta}} = \begin{pmatrix} \dfrac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top\partial\nu_1} \\ \dfrac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top\partial\boldsymbol{\nu}_2} \\ \mathbf{O}_{(2n_\gamma+1)n_\gamma\times n_\gamma} \end{pmatrix}$, *with* $\dfrac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top\partial\nu_1} = 2\nu_1^{-2}\big(\kappa_{1,i}(\boldsymbol{\gamma},\boldsymbol{\nu}) - \kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu}) - 2\nu_1^{-1}N\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})\big)\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})^2\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top$,

$$
\begin{aligned}
\frac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}^\top\partial\boldsymbol{\nu}_2} &= -\frac{\partial\,\mathrm{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)^\top}{\partial\boldsymbol{\nu}_2}\Bigg\{\left(\nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})\otimes\nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})\right)\frac{\partial\big(\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\big)}{\partial\boldsymbol{\gamma}^\top} \\
&\quad + \kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\nu_1^{-1}\frac{\partial}{\partial\boldsymbol{\gamma}^\top}\Big(\boldsymbol{e}_i(\boldsymbol{\gamma})\otimes\boldsymbol{e}_i(\boldsymbol{\gamma})\Big)\Bigg\} \\
&= -\frac{\partial\,\mathrm{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)^\top}{\partial\boldsymbol{\nu}_2}\Bigg\{\left(\nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})\otimes\nu_1^{-1/2}\boldsymbol{e}_i(\boldsymbol{\gamma})\right) \\
&\quad \times\left(4\nu_1^{-1}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\right) \\
&\quad - \kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\nu_1^{-1}\Big(\big(\boldsymbol{I}_N\otimes\boldsymbol{e}_i(\boldsymbol{\gamma})\big) + \big(\boldsymbol{e}_i(\boldsymbol{\gamma})\otimes\boldsymbol{I}_N\big)\Big)\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\Bigg\},
\end{aligned}
$$

*following from* (F.16), (F.17), *and the identity* $\frac{\partial \boldsymbol{e}_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} = -\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top$.

***(iii-5)*** *Finally, by applying* (F.18) *to* (F.40), *we obtain*

$$\begin{aligned}\frac{\partial^2 w_i(\gamma_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}^\top} &= 8\nu_1^{-2}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})^2 w_i(\gamma_i,\boldsymbol{\nu})\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top \\ &\quad - 2\nu_1^{-1}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\gamma_i,\boldsymbol{\nu})\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top \\ &\quad + 2\nu_1^{-1}\kappa_{2,i}(\boldsymbol{\gamma},\boldsymbol{\nu})w_i(\gamma_i,\boldsymbol{\nu})\left(\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\otimes \boldsymbol{I}_{n_\gamma}\right)\boldsymbol{\mathcal{X}}^{\gamma\gamma}(\boldsymbol{\gamma}). \end{aligned} \tag{F.41}$$

*The construction of* $\boldsymbol{\mathcal{X}}^{\gamma\gamma}(\boldsymbol{\gamma})$ *is omitted. It nevertheless holds that* $\left\|\boldsymbol{\mathcal{X}}^{\gamma\gamma}(\boldsymbol{\gamma})\right\| \le \left\|\frac{\partial\, \mathrm{vec}(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top)}{\partial \boldsymbol{\gamma}^\top}\right\|$.

***(iv)*** *By employing* (F.15), *one has*

$$\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top} = \left(\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top \otimes \boldsymbol{I}_{n\gamma}\right)\frac{\partial\, \mathrm{vec}\left(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right)}{\partial \boldsymbol{\theta}^\top} + \boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}, \tag{F.42}$$

*where*

$$\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top} = w_i(\gamma_i,\boldsymbol{\nu})\left(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)\frac{\partial\, \mathrm{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial \boldsymbol{\theta}^\top} + \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})\frac{\partial w_i(\gamma_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}, \tag{F.43}$$

*with* $\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})$ *defined above in Part (iii-3).*

***(v)*** *Recall* $\frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}$ *and* $\frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}$ *from* (F.28) *and* (F.29), *respectively. Similar to* (F.42), *one has*

$$\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} = \left(\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top \otimes \boldsymbol{I}_{n\gamma}\right)\frac{\partial\, \mathrm{vec}\left(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right)}{\partial \boldsymbol{\gamma}^\top} + \boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}, \tag{F.44}$$

*where*

$$\begin{aligned}\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} &= w_i(\gamma_i,\boldsymbol{\nu})\left[\left(\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\otimes \boldsymbol{I}_{n_\gamma}\right)\frac{\partial\, \mathrm{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)}{\partial \boldsymbol{\gamma}^\top} - \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\right] \\ &\quad + \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})\frac{\partial w_i(\gamma_i,\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} - \sum_{j=1}^N \frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}, \end{aligned} \tag{F.45}$$

*with* $\frac{\partial\, \mathrm{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)}{\partial \boldsymbol{\gamma}^\top} = \dot{\boldsymbol{\Phi}}^\mu(\boldsymbol{\gamma}) + \left(\mathrm{diag}\left(\boldsymbol{e}_i(\boldsymbol{\gamma})\right)\otimes \boldsymbol{I}_{n_\gamma}\right)\dot{\boldsymbol{\Phi}}^\sigma(\boldsymbol{\gamma}) - \boldsymbol{\Phi}^\sigma_{\mathrm{diag}}(\boldsymbol{\gamma})\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top$, *and*

$$\dot{\boldsymbol{\Phi}}^\mu(\boldsymbol{\gamma}) = \begin{pmatrix} \frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t}_1;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} \\ \frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t}_2;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} \\ \vdots \\ \frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t}_N;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} \end{pmatrix}, \quad \dot{\boldsymbol{\Phi}}^\sigma(\boldsymbol{\gamma}) = \begin{pmatrix} \frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_1;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} \\ \frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_2;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} \\ \vdots \\ \frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_N;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} \end{pmatrix}, \quad \boldsymbol{\Phi}^\sigma_{\text{diag}}(\boldsymbol{\gamma}) = \text{diag}\Big(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_1;\boldsymbol{\gamma}), \dots, \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_N;\boldsymbol{\gamma})\Big).$$

***(vi)*** *By applying (F.19), we obtain the following identity from (F.42):*

$$\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\text{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) = \Big(\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top \otimes \boldsymbol{I}_{n_\gamma n_\theta}\Big)\boldsymbol{\mathcal{S}}_1^{\theta\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \boldsymbol{\mathcal{S}}_2^{\theta}(\boldsymbol{\gamma},\boldsymbol{\nu})\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}$$
$$+ \big(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\big)\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\text{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) + \left(\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right)^\top \otimes \boldsymbol{I}_{n_\gamma}\right)\frac{\partial\,\text{vec}\,\big(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\big)}{\partial \boldsymbol{\gamma}^\top}, \tag{F.46}$$

*where* $\left\|\boldsymbol{\mathcal{S}}_1^{\theta\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| \le C$ *and* $\left\|\boldsymbol{\mathcal{S}}_2^{\theta}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| \le C$ *provided that Assumption AN1 holds, using the equivalence of norms in finite dimensional spaces. The specific details of* $\boldsymbol{\mathcal{S}}_1^{\theta\gamma}(\cdot)$ *and* $\boldsymbol{\mathcal{S}}_2^{\theta}(\cdot)$ *are omitted for brevity. Note that, by (F.14), one has* $\text{vec}\,\big(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\big) = \big[\boldsymbol{I}_N \otimes \text{vec}\,\big(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\big)\big]\boldsymbol{e}_i(\boldsymbol{\gamma}) = \big(\boldsymbol{e}_i(\boldsymbol{\gamma}) \otimes \boldsymbol{I}_{n_\gamma N}\big)\,\text{vec}\,\big(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\big)$. *Then, using this identity together with (F.12) and (F.43), we have*

$$\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\text{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) = \frac{\partial}{\partial \boldsymbol{\gamma}^\top}\Bigg\{w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\left(\frac{\partial\,\text{vec}\,\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\big)^\top}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_{n_\gamma}\right)\text{vec}\,\Big(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\Big)$$
$$+ \text{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\boldsymbol{e}_i(\boldsymbol{\gamma})\frac{\partial w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right)\Bigg\}$$
$$= \left(\frac{\partial\,\text{vec}\,\big(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\big)^\top}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_{n_\gamma}\right)\boldsymbol{\mathcal{P}}_{1,i}^{\theta\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \boldsymbol{\mathcal{P}}_{2,i}^{\theta\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu}), \tag{F.47}$$

*where, using* $\dfrac{\partial \boldsymbol{e}_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} = -\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top$,

$$\boldsymbol{\mathcal{P}}_{1,i}^{\theta\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu}) = w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})\left(-\Big(\boldsymbol{I}_N \otimes \text{vec}\,\big(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\big)\Big)\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top + \big(\boldsymbol{e}_i(\boldsymbol{\gamma}) \otimes \boldsymbol{I}_{n_\gamma N}\big)\frac{\partial\,\text{vec}\,\big(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\big)}{\partial \boldsymbol{\gamma}^\top}\right)$$
$$+ \text{vec}\,\Big(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\Big)\frac{\partial w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}, \tag{F.48}$$

$$\boldsymbol{\mathcal{P}}_{2,i}^{\theta\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu}) = \left(\left(\frac{\partial w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\boldsymbol{e}_i(\boldsymbol{\gamma})^\top\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right) \otimes \boldsymbol{I}_{n_\gamma}\right)\frac{\partial\,\text{vec}\,\big(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\big)}{\partial \boldsymbol{\gamma}^\top} + \Big(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\Big)$$
$$\times \left(\big(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{e}_i(\boldsymbol{\gamma})\big)\frac{\partial^2 w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top \partial \boldsymbol{\theta}} - \left(\frac{\partial w_i(\boldsymbol{\gamma}_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_N\right)\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\right). \tag{F.49}$$

*All relevant partial derivatives from (F.46) to (F.49) are presented in the preceding results.*

***(vii)*** *Similarly, we have*

$$\frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) = \left(\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top \otimes \boldsymbol{I}_{n_\gamma n_\theta}\right)\boldsymbol{\mathcal{S}}_1^{\theta\theta}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \boldsymbol{\mathcal{S}}_2^{\theta}(\boldsymbol{\gamma},\boldsymbol{\nu})\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}$$
$$+ \left(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right)\frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) + \left(\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right)^\top \otimes \boldsymbol{I}_{n_\gamma}\right)\frac{\partial \operatorname{vec}\left(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right)}{\partial \boldsymbol{\theta}^\top}, \quad \text{(F.50)}$$

*where* $\left\|\boldsymbol{\mathcal{S}}_1^{\theta\theta}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| \leq C$ *and* $\left\|\boldsymbol{\mathcal{S}}_2^{\theta}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| \leq C$ *(as defined in* (F.46)*), provided that Assumption AN1 is satisfied and given the equivalence of norms in finite dimensional spaces. Furthermore,*

$$\frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top}\right) = \left(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)\boldsymbol{\mathcal{P}}_{1,i}^{\theta\theta}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \boldsymbol{\mathcal{P}}_{2,i}^{\theta\theta}(\boldsymbol{\gamma},\boldsymbol{\nu}), \quad \text{(F.51)}$$

*where*

$$\boldsymbol{\mathcal{P}}_{1,i}^{\theta\theta}(\boldsymbol{\gamma},\boldsymbol{\nu}) = w_i(\gamma_i,\boldsymbol{\nu})\frac{\partial}{\partial \boldsymbol{\theta}^\top} \operatorname{vec}\left(\frac{\partial \operatorname{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial \boldsymbol{\theta}^\top}\right) + \operatorname{vec}\left(\frac{\partial \operatorname{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial \boldsymbol{\theta}^\top}\right)\frac{\partial w_i(\gamma_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}^\top},$$
$$\boldsymbol{\mathcal{P}}_{2,i}^{\theta\theta}(\boldsymbol{\gamma},\boldsymbol{\nu}) = \left(\boldsymbol{I}_{n_\theta} \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1/2}\boldsymbol{u}_i(\boldsymbol{\nu}_2,\boldsymbol{\gamma})\right)\left(\frac{\partial^2 w_i(\gamma_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\right)^\top$$
$$+ \left(\frac{\partial w_i(\gamma_i,\boldsymbol{\nu})}{\partial \boldsymbol{\theta}}\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)\frac{\partial \operatorname{vec}\left(\boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right)}{\partial \boldsymbol{\theta}^\top}.$$

*All relevant partial derivatives have been provided in the preceding results.*

***(viii)*** *By applying* (F.19) *to* (F.44)*, we have*

$$\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{s}_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right) = \left(\nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})^\top \otimes \boldsymbol{I}_{n_\gamma^2}\right)\boldsymbol{\mathcal{S}}_1^{\gamma\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu}) + \boldsymbol{\mathcal{S}}_2^{\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu})\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}$$
$$+ \left(\boldsymbol{I}_{n_\gamma} \otimes \boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right)\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right) + \left(\left(\frac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right)^\top \otimes \boldsymbol{I}_{n_\gamma}\right)\frac{\partial \operatorname{vec}\left(\boldsymbol{S}(\boldsymbol{\gamma},\boldsymbol{\nu})\right)}{\partial \boldsymbol{\gamma}^\top}, \quad \text{(F.52)}$$

*where* $\left\|\boldsymbol{\mathcal{S}}_1^{\gamma\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| \leq C$ *and* $\left\|\boldsymbol{\mathcal{S}}_2^{\gamma}(\boldsymbol{\gamma},\boldsymbol{\nu})\right\| \leq C$ *under Assumption AN1 and norm equivalence in finite dimensional spaces. We omit the details. The construction of* $\dfrac{\partial \nabla_i(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}$ *is derived in* (F.45).

*It remains to consider:*

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \nabla_i(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right) \\
&= \frac{\partial}{\partial \boldsymbol{\gamma}^\top}\left\{ w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \operatorname{vec}\left(\left(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \otimes \boldsymbol{I}_{n_\gamma}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)}{\partial \boldsymbol{\gamma}^\top}\right)\right\} \\
&\quad - \frac{\partial}{\partial \boldsymbol{\gamma}^\top}\left\{ w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\right)\right\} \\
&\quad + \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma}) \frac{\partial w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}\right) - \sum_{j=1}^N \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}\right) \\
&=: \boldsymbol{\mathcal{P}}_{1,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu}) - \boldsymbol{\mathcal{P}}_{2,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu}) + \boldsymbol{\mathcal{P}}_{3,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu}) - \boldsymbol{\mathcal{P}}_{4,i}^{\gamma\gamma}(\boldsymbol{\gamma}),
\end{aligned} \tag{F.53}$$

*following from* (F.45)*. By* (F.19) *and* $\frac{\partial \boldsymbol{e}_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} = -\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top$, $\boldsymbol{\mathcal{P}}_{1,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu})$ *can be written as*

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_{1,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \operatorname{vec}\left(\left(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \otimes \boldsymbol{I}_{n_\gamma}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)}{\partial \boldsymbol{\gamma}^\top}\right) \frac{\partial w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} \\
&\quad + w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})\Big(\left(\boldsymbol{e}_i(\boldsymbol{\gamma})^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \otimes \boldsymbol{I}_{n_\gamma^2}\right) \boldsymbol{\mathcal{R}}_{1,i}(\boldsymbol{\gamma}) - \boldsymbol{\mathcal{R}}_{2,i}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\Big),
\end{aligned}$$

*where* $\frac{\partial w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top}$ *and* $\frac{\partial \operatorname{vec}(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}^\top}$ *are provided in Part (iii-3) and below* (F.45)*, respectively. Moreover,*

$$\left\|\boldsymbol{\mathcal{R}}_{1,i}(\boldsymbol{\gamma})\right\| \leq \left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top} \operatorname{vec}\left(\frac{\partial \operatorname{vec}(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}^\top}\right)\right\|_F \leq C + C\left\|\boldsymbol{e}_i(\boldsymbol{\gamma})\right\|, \tag{F.54}$$

*provided that Assumptions A1 and AN2 hold, and* $\left\|\boldsymbol{\mathcal{R}}_{2,i}(\boldsymbol{\gamma})\right\| \leq \left\|\frac{\partial \operatorname{vec}(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}^\top}\right\|_F$*. The second term* $\boldsymbol{\mathcal{P}}_{2,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu})$ *in* (F.53) *can be written as*

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_{2,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\right) \frac{\partial w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}^\top} + w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu}) \\
&\times \left(\left(\boldsymbol{I}_{n_\gamma} \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top\right)}{\partial \boldsymbol{\gamma}^\top} + \left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \otimes \boldsymbol{I}_{n_\gamma}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)}{\partial \boldsymbol{\gamma}^\top}\right).
\end{aligned}$$

*Finally, by* $\frac{\partial \boldsymbol{e}_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} = -\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top$*, we have*

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_{3,i}^{\gamma\gamma}(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \left(\frac{\partial w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}} \boldsymbol{e}_i(\boldsymbol{\gamma})^\top \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \otimes \boldsymbol{I}_{n_\gamma}\right) \frac{\partial \operatorname{vec}\left(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})\right)}{\partial \boldsymbol{\gamma}^\top} - \left(\frac{\partial w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}}\right. \\
&\quad \left.\otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1}\right) \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma})^\top + \left(\boldsymbol{I}_{n_\gamma} \otimes \boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}) \boldsymbol{\Lambda}(\boldsymbol{\nu}_2)^{-1} \boldsymbol{e}_i(\boldsymbol{\gamma})\right) \left(\frac{\partial^2 w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top}\right)^\top,
\end{aligned}$$

*where* $\dfrac{\partial^2 w_i(\boldsymbol{\gamma}_i, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top}$ *is provided in* (F.41)*, and finally,* $\big\|\boldsymbol{\mathcal{P}}_{4,i}^{\gamma\gamma}(\boldsymbol{\gamma})\big\| \le C$ *if Assumption AN2 holds.*

***Proof of Lemma F.7.*** The results are derived by repeatedly applying Lemma F.6 and basic linear algebra. The key steps are outlined in the lemma, and the remaining minor (but cumbersome) details are omitted. Here, we only provide further details on the second inequality in (F.54). Let $e_{i,j}(\boldsymbol{\gamma})$ be the $j_{\text{th}}$ element of $\boldsymbol{e}_i(\boldsymbol{\gamma})$. We have

$$\begin{aligned}\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \operatorname{vec}(\boldsymbol{\Phi}_i^{\mu,\sigma}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}^\top}\right)\right\|_F &\le \sum_{j=1}^N \left\{\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t}_j;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}\right)\right\|_F\right.\\ &+\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}e_{i,j}(\boldsymbol{\gamma})\right)\right\|_F + \left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})\boldsymbol{\Phi}^\mu(\boldsymbol{t}_j;\boldsymbol{\gamma})^\top\right)\right\|_F\\ &+\left.\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})^\top e_{i,j}(\boldsymbol{\gamma})\right)\right\|_F\right\},\end{aligned}$$

where $\sum_{j=1}^N\left(\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t}_j;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}\right)\right\|_F + \left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}e_{i,j}(\boldsymbol{\gamma})\right)\right\|_F\right) \le C + C\big\|\boldsymbol{e}_i(\boldsymbol{\gamma})\big\|$ by the norm equivalence, and Assumptions A1 and AN2. Note that $\dfrac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})\boldsymbol{\Phi}^\mu(\boldsymbol{t};\boldsymbol{\gamma})^\top\right) = \left(\boldsymbol{\Phi}^\mu(\boldsymbol{t};\boldsymbol{\gamma})\otimes \boldsymbol{I}_{n_\gamma}\right)\dfrac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top} + \left(\boldsymbol{I}_{n_\gamma}\otimes \boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})\right)\dfrac{\partial \boldsymbol{\Phi}^\mu(\boldsymbol{t};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}$ and

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})^\top e_{i,j}(\boldsymbol{\gamma})\right) &= \operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})^\top\right)\frac{\partial e_{i,j}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}\\ &+ e_{i,j}(\boldsymbol{\gamma})\Big(\left(\boldsymbol{I}_{n_\gamma}\otimes \boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})\right) + \left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})\otimes \boldsymbol{I}_{n_\gamma}\right)\Big)\frac{\partial \boldsymbol{\Phi}^\sigma(\boldsymbol{t};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^\top}.\end{aligned}$$

Following the same argument, we obtain $\sum_{j=1}^N\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})\boldsymbol{\Phi}^\mu(\boldsymbol{t}_j;\boldsymbol{\gamma})^\top\right)\right\|_F \le C$ and $\sum_{j=1}^N\left\|\frac{\partial}{\partial \boldsymbol{\gamma}^\top}\operatorname{vec}\left(\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})\boldsymbol{\Phi}^\sigma(\boldsymbol{t}_j;\boldsymbol{\gamma})^\top e_{i,j}(\boldsymbol{\gamma})\right)\right\|_F \le C\big\|\boldsymbol{e}_i(\boldsymbol{\gamma})\big\|$ if Assumptions A1 and AN2 hold. To sum up, under Assumptions A1 and AN2 hold, we have $\big\|\boldsymbol{\mathcal{R}}_{1,i}(\boldsymbol{\gamma})\big\| \le C + C\big\|\boldsymbol{e}_i(\boldsymbol{\gamma})\big\|$. □

# G More empirical results

## G.1 Volatility curves of stock returns using fGARCH

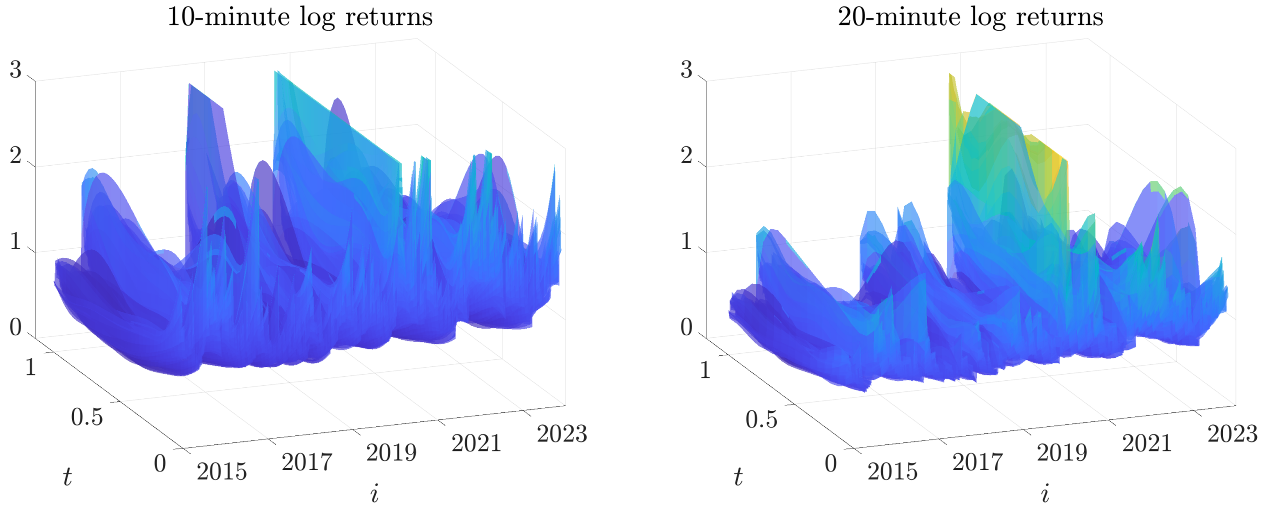


**Figure G1:** Fitted intraday volatility curves for PFE by fGARCH from 2 January 2015 up to 29 December 2023: The fitted volatility levels are annualized using $6 \cdot 6 \cdot \sqrt{252}$ for 10-minute intraday log returns and $6 \cdot 3 \cdot \sqrt{252}$ for 20-minute intraday log returns, ensuring they are displayed on a comparable scale. The left figure shows the fitted curves for 10-minute log returns, while the right figure represents the curves for 20-minute log returns.

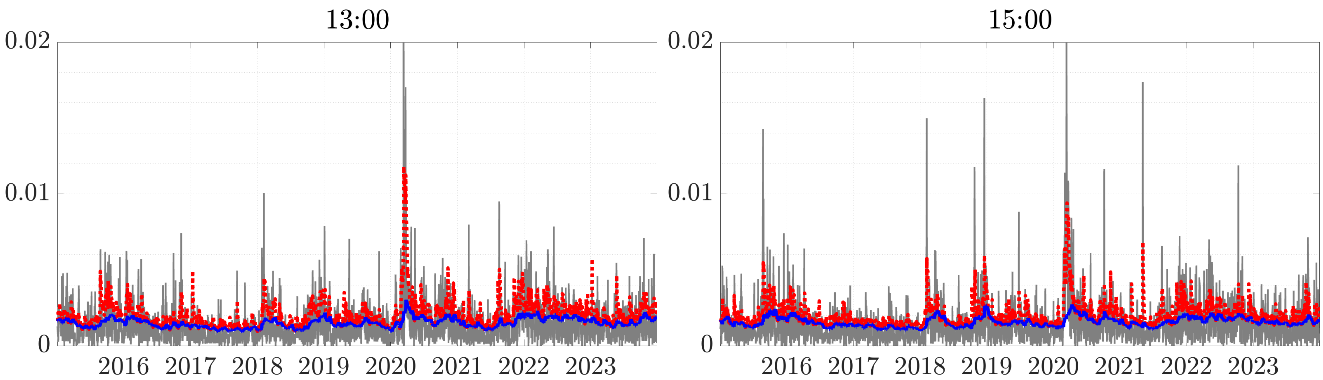


**Figure G2:** Fitted (nonannualized) volatility paths for PFE at 13:00 and 15:00 using fGAS (blue solid curves) and fGARCH (red dotted curves) from January 2, 2015 to December 29, 2023, based on 20-minute intraday log returns. Gray solid lines represent the absolute values of 20-minute intraday log returns.

## G.2 Further details on $PM_{2.5}$ concentration

**Table G.1:** Indices of the selected sensors

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2542 | 5942 | 16237 | 16999 | 38805 | 56229 | 89581 | 101209 | 126489 | 162877 |
| 2546 | 6540 | 16445 | 17017 | 42095 | 56507 | 90479 | 101481 | 129823 | 165897 |
| 2556 | 6694 | 16789 | 17615 | 43753 | 57949 | 92593 | 101597 | 130345 | 166359 |
| 3064 | 7302 | 16833 | 21947 | 43891 | 60929 | 93379 | 101603 | 133001 | 168033 |
| 4033 | 9942 | 16849 | 23335 | 44187 | 77083 | 94351 | 103240 | 134054 | 169575 |
| 4127 | 10116 | 16929 | 23809 | 47165 | 77429 | 95689 | 104076 | 134192 | 169749 |
| 4331 | 13027 | 16959 | 23837 | 48039 | 83339 | 95847 | 118937 | 138380 | 171777 |
| 4406 | 13029 | 16961 | 23999 | 49233 | 83983 | 96479 | 120429 | 138412 | 172203 |
| 4454 | 14319 | 16971 | 32277 | 49607 | 86175 | 98487 | 120785 | 138454 | 173569 |
| 5428 | 14371 | 16973 | 37173 | 52367 | 87233 | 99633 | 121331 | 150828 | 174847 |
| 5892 | 14857 | 16993 | 37599 | 55513 | 87351 | 99663 | 126107 | 156537 | |

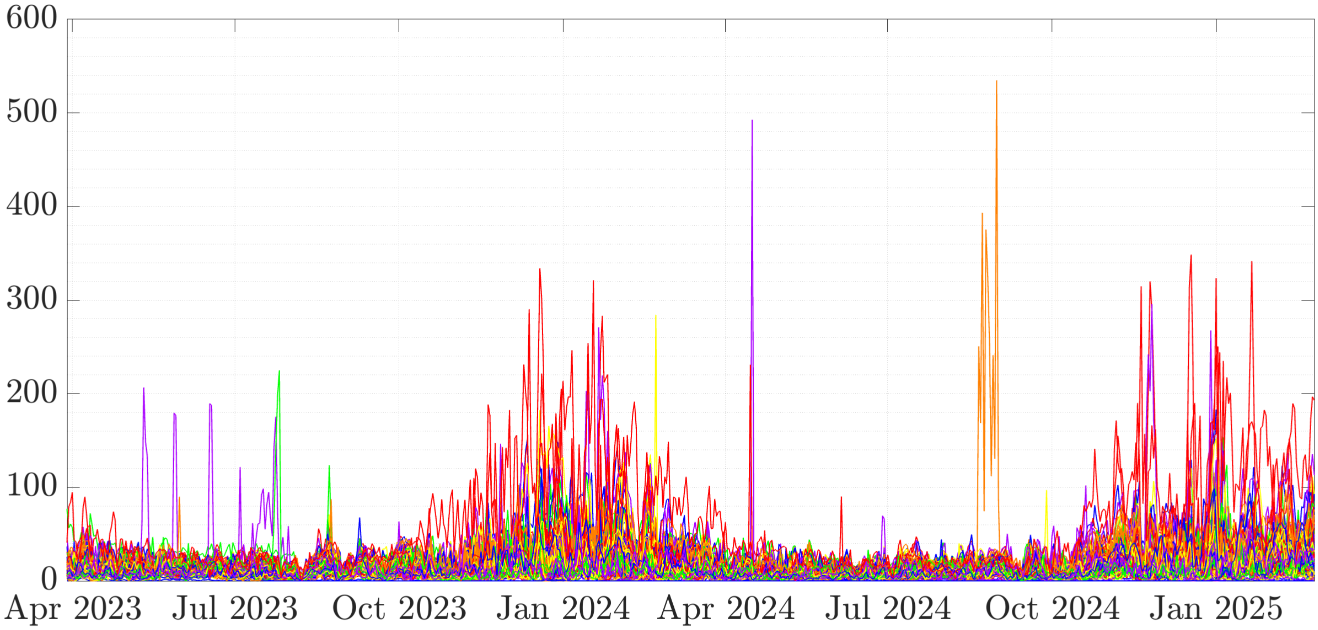


**Figure G3:** Time series plots of daily average $PM_{2.5}$ emissions, with each curve representing the data recorded by a sensor over time.

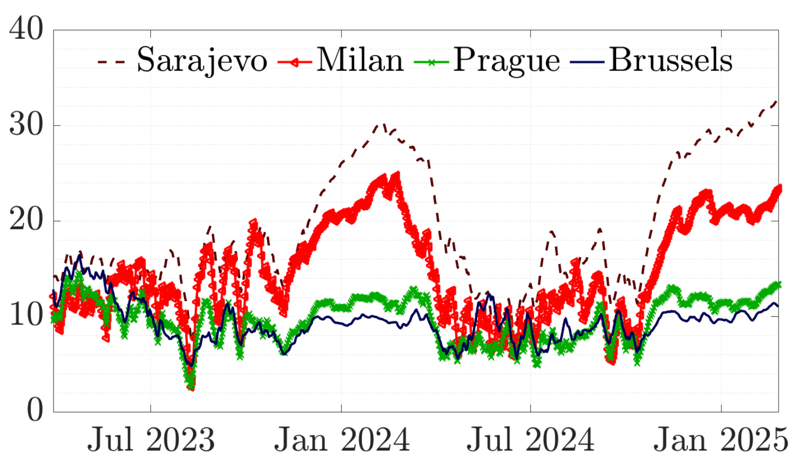


**Figure G4:** Predicted $PM_{2.5}$ levels for four selected sensors

# References in Online Appendix


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.

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

Fang, K.-T., S. Kotz, and K. W. Ng (2018). *Symmetric Multivariate and Related Distributions.* Boca Raton: CRC Press, Taylor & Francis Group.

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

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

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

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis* (2nd ed.). Cambridge: Cambridge University Press.

Klenke, A. (2020). *Probability Theory: A Comprehensive Course* (3rd ed.). Universitext. Cham: Springer.

Krengel, U. (1985). *Ergodic Theorems*, Volume 6. Berlin–New York: Walter de Gruyter.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Berlin: Springer.

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.

Song, D. and A. K. Gupta (1997). $L_p$-norm uniform distribution. *Proceedings of the American Mathematical Society 125*(2), 595–601.

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), 2449–2495.

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

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