TI 2025-011/III
Tinbergen Institute Discussion Paper

Score-driven time-varying parameter models with spline-based densities

Janneke van Brummelen[1]
Paolo Gorgi[2]
Siem Jan Koopman[3]

1 Vrije Universiteit Amsterdam, Tinbergen Institute
2 Vrije Universiteit Amsterdam, Tinbergen Institute
3 Vrije Universiteit Amsterdam, Tinbergen Institute

Score-driven time-varying parameter models with spline-based densities*

Janneke van Brummelen, Paolo Gorgi and Siem Jan Koopman

Vrije Universiteit Amsterdam and Tinbergen Institute

February, 2025

Abstract

We develop a score-driven time-varying parameter model where no particular parametric error distribution needs to be specified. The proposed method relies on a versatile spline-based density, which produces a score function that follows a natural cubic spline. This flexible approach nests the Gaussian density as a special case. It can also represent asymmetric and leptokurtic densities that produce outlier-robust updating functions for the time-varying parameter and are often appealing in empirical applications. As leading examples, we consider models where the time-varying parameters appear in the location or in the log-scale of the observations. The static parameter vector of the model can be estimated by means of maximum likelihood and we formally establish some of the asymptotic properties of such estimators. We illustrate the practical relevance of the proposed method in two empirical studies. We employ the location model to filter the mean of the U.S. monthly CPI inflation series and the scale model for volatility filtering of the full panel of daily stock returns from the S&P 500 index. The results show a competitive performance of the method compared to a set of competing models that are available in the existing literature.

Key words: natural cubic spline, location model, volatility model, maximum likelihood estimation, consistency

JEL classification: C13, C22.

*Corresponding author: J. van Brummelen. E-mail: `j.n.van.brummelen@vu.nl`

1. Introduction

Observation-driven models constitute a large class of dynamic models where the time-varying parameters are driven by past observations (see Cox, 1981). Examples include the seminal generalized autoregressive conditional heteroskedasticity (GARCH) model (Engle, 1982; Bollerslev, 1986) and its variants, the conditional duration model (Engle and Russell, 1998), the conditional copula model (Patton, 2006) and many other dynamic models that are widely used in economics and finance. A general and effective approach to specify the updating equation of observation-driven time-varying parameters is given by the score-driven framework (Creal et al., 2011, 2013; Harvey, 2013). We refer the reader to Artemova et al. (2022a,b) for a review of score-driven models. In score-driven models, the specification of the time-varying parameter is obtained through its score of the conditional density, which is typically selected to be a parametric density function. The use of a parametric density can be rather restrictive in practice. We consider a novel semiparametric score-driven model that features a nonparametric error distribution. The nonparametric distribution is approximated via a spline function and this leads to a flexible score that takes the form of a natural cubic spline.

Earlier semiparametric score-driven models include the conditional volatility model proposed by Blasques et al. (2016) that relies on Kernel density estimation (KDE) of the distribution of the error term. The method requires the model errors to be dependent on the score of a kernel density which in turn is constructed from these same errors. Therefore, an iterative or multi-step procedure is needed for the estimation of the model parameters. On the other hand, our approach based on a spline density does not require such an iterative estimation procedure and the model can be estimated through a single optimization of a likelihood function. Furthermore, we develop a more general class of score-driven models that can handle both location (mean) and scale (variance) but potentially also other time-varying features of the density. The dynamic quantile model of Catania and Luati (2023) and the expected shortfall model of Patton et al. (2019) are other earlier examples of semiparametric score-driven models but these consider parameter estimation based on loss functions different from the negative log likelihood, such that they do not require the specification of a parametric error density. For example, the quantile model uses the check loss function instead. Furthermore, Abdelkarim and Onour (2023) introduce a semiparametric exponentially weighted moving average model for conditional volatilities, using an approach similar to that of Blasques et al. (2016), based on KDE.

The two most common approaches to density estimation with splines are smoothing splines and interpolation splines. For smoothing splines, a knot is placed at every observation in the sample and the smoothness of the resulting density is enforced using a regularization term in the objective function, see, for example, Gu and Qiu (1993) and Gu (1993). In our score-driven model, the use of smoothing splines would lead to the same iterative procedure for parameter estimation as is needed in case KDE is used. We can circumvent a multi-step estimation approach when considering interpolation splines. In the case of interpolation splines, the number of knots must be set a-priori, while the knot positions and the corresponding ordinates can be estimated using a specifically designed procedure. The interpolation spline goes exactly through the ordinates of all knot positions in a smooth manner, depending on the order of continuity at the knots.

A challenge in spline-based density estimation is enforcing positivity of the estimated density function. This can either be achieved by adding a constraint to the optimization problem (Gu and Qiu, 1993), or by using so-called M-splines which are positive by construction (see e.g. Abrahamowicz et al., 1992), or by estimating the log density instead of the density itself. The latter approach is used by Kooperberg and Stone (1991, 1992) in their so-called logspline model. This is a density estimation method in which the log density is modeled as a natural cubic B-spline function, where the B-spline coefficients are estimated using the method of maximum likelihood (ML). Stone (1990) has shown that in case the support is in a compact interval, the ML estimator converges to the true density at an optimal rate of convergence under mild conditions, provided that the number of parameters increases with the sample size at an appropriate rate. Our proposed methodology adopts a modeling approach that is similar to this idea. In particular, we model the score function as a natural cubic spline, and use it to reconstruct the log density of the error. Since we force the score to be linear outside the outer knots, the corresponding density will have tails that decay at an exponential rate. This framework implies that the spline-based distribution will not be heavy-tailed, but it is not hard to see that any distribution can be approximated arbitrarily well if sufficiently many knots are considered. As the spline function is highly flexible, the score function can take many different shapes and forms. Hence, we can deliver a spline function with sufficient support in the tails of the density such that the filter for the time-varying parameter is effectively outlier-robust. Also, the spline can handle asymmetries in the log density.

We aim to develop the semiparametric spline-based score-driven framework and showcase its use in relevant empirical studies. For a given vector of knots, the spline parameters and remaining model parameters are estimated jointly using ML. The number of knots can be selected based on existing information criteria. We further examine the conditional location and log-scale variants of the model in detail. Consistency of the ML estimator is established for both of these models under correct specification. For the conditional location model, we also establish asymptotic normality of the ML estimator. For the log-scale model, we do not establish this, as it is too intricate to verify whether the required moment conditions hold, which is a known challenge for log-scale models, see e.g. the discussions in Nelson and Foster (1994) and Straumann and Mikosch (2006) on the exponential GARCH model. The results of Blasques et al. (2022), where high-level conditions for consistency and asymptotic normality are provided for ML estimators of score-driven models, can be used for some parts of the derivations, while other aspects of this development require tailored derivations.

We demonstrate the empirical relevance and broad applicability of our framework by presenting two illustrations in economics and finance. First, the location model is considered for the filtering of the key monthly time series of U.S. inflation. Second, the scale model is used for the volatility analysis of 456 stock returns from the S&P 500 index. The results show that our flexible spline-based framework can deliver outlier-robust filters and that the fit is in many cases considerably better than the competitive Student's t model. These findings underscore the potential of spline-based score-driven models for enhancing model flexibility and improving the modeling and analysis of macroeconomic and financial time series.

The paper is organized as follows. Section 2 introduces the spline-based density and the corresponding score-driven model. Section 3 treats parameter estimation through ML and in Section 4 the asymptotic theory for this estimator is established for location and scale models. Section 5 contains the empirical applications. Finally, Section 6 concludes.

2. The score-driven model with spline-based error density

2.1. The natural cubic spline function

We start by introducing cubic splines and the corresponding notation. The cubic spline function, as for instance treated by Poirier (1973), is constructed based on a vector of x-coordinates $\boldsymbol{t} = (t_1, \dots, t_k)$ called knots, where $k \in \mathbb{N}$ is the number of knots and where $t_1 < t_2 < \dots < t_k$, and a vector of corresponding function values or ordinates $\boldsymbol{y} = (y_1, \dots, y_k) \in \mathbb{R}^k$. Typically, a cubic spline is defined on some interval $[a, b]$ and the outer knots are set as $t_1 = a$ and $t_k = b$. Instead, here we consider the entire real line as the domain. The natural cubic spline function $h(\cdot; \boldsymbol{t}, \boldsymbol{y}) : \mathbb{R} \to \mathbb{R}$ is a twice differentiable function that goes through each (t_i, y_i)-pair, such that in each interval $(t_1, t_2]$, ..., $(t_{k-1}, t_k]$, h is a third degree polynomial and outside of $(t_1, t_k]$ the function h is linear. From now on we supress the arguments $\boldsymbol{t}$ and $\boldsymbol{y}$ in h if possible, for notational convenience. In particular, for any $z \in \mathbb{R}$, the spline function is:

$$h(z) = \begin{cases} h_0(z) & \text{if } z \leq t_1 \, , \\ h_i(z) & \text{if } t_i < z \leq t_{i+1} \, , \quad \text{for } i = 1, \ldots, k-1 \, , \\ h_k(z) & \text{if } z > t_k \, , \end{cases}$$

where h_i, for $i = 0, \ldots, k-1$, are functions of the form:

$$h_i(z) = a_i + b_i(z - t_{i+1}) + c_i(z - t_{i+1})^2 + d_i(z - t_{i+1})^3 \, ,$$

and

$$h_k(z) = a_k + b_k(z - t_k) \, ,$$

where $c_0 = d_0 = 0$, and where the real-valued coefficients a_i and b_i for $i = 0, \ldots, k$, and c_i and d_i for $i = 1, \ldots, k-1$ must be such that:

(i) $h_i(t_i) = y_i = h_{i-1}(t_i)$, for $i = 1, \ldots, k$,

(ii) $h'_i(t_i) = h'_{i-1}(t_i)$, for $i = 1, \ldots, k$,

(iii) $h''_i(t_i) = h''_{i-1}(t_i)$, for $i = 1, \ldots, k$, where it can be noted that $h''_0(t_1) = h''_k(t_k) = 0$, as h_0 and h_k are linear.

This system of $4k$ equations, with the $4(k+1) - 4 = 4k$ unkown coefficients a_i, b_i, c_i and d_i, has a unique solution that can be obtained analytically. In particular, it can be rewritten as a linear system with a tridiagonal matrix, see Poirier (1973). Such linear systems can be solved using a simple recursive algorithm, see, for instance, Alberg et al. (1967, Chapter 2). Conveniently, the coefficients are a linear function of the vector of ordinates $\boldsymbol{y}$. Also, due to the parametrization of the spline functions above, it follows immediately from condition (i) that $a_i = y_{i+1}$ for $i = 0, \ldots, k-1$ and $a_k = y_k$.

2.2. Constructing the natural cubic spline density

The next step is to construct a density function $p_\varepsilon : \mathbb{R} \to \mathbb{R}_+$ based on spline function h. For any $\varepsilon \in \mathbb{R}$ let

$$\frac{\partial \log p_\varepsilon(\varepsilon)}{\partial \varepsilon} = -h(\varepsilon; \boldsymbol{t}, \boldsymbol{y}) \, , \tag{1}$$

such that the score of the density p_ε is equal to the spline function multiplied by minus one. Then $\log p_\varepsilon(\varepsilon)$ can be obtained by calculating the anti-derivative of this function with respect to ε. As h is a piecewise polynomial function, it is straightforward to construct its anti-derivative denoted by H, since for any $z \in \mathbb{R}$:

$$H(z) = \begin{cases} H_0(z) & \text{if } z \leq t_1\,, \\ H_i(z) & \text{if } t_i < z \leq t_{i+1}\,, \quad \text{for } i = 1, \ldots, k-1\,, \\ H_k(z) & \text{if } z > t_k\,, \end{cases}$$

where for $i = 0, \ldots, k-1$:

$$H_i(z) = a_i(z - t_{i+1}) + \frac{1}{2}b_i(z - t_{i+1})^2 + \frac{1}{3}c_i(z - t_{i+1})^3 + \frac{1}{4}d_i(z - t_{i+1})^4 + e_i\,,$$

and

$$H_k(z) = a_k(z - t_k) + \frac{1}{2}b_k(z - t_k)^2 + e_k\,,$$

and where e_i for $i = 0, \ldots, k$ are some (non-unique) constants that should be chosen such that the anti-derivative H is continuous. These coefficients are again linear in the vector $\boldsymbol{y}$, as they are linear functions of the original coefficients a_i, b_i, c_i and d_i, which are themselves linear in $\boldsymbol{y}$. For numerical reasons, it is convenient to choose the e_i coefficients such that $-H(z)$ does not become too large, for instance by setting $H(0) = 0$. From (1), it follows that

$$p_\varepsilon(\varepsilon) \propto \exp(-H(\varepsilon))\,.$$

In order to obtain a density, we normalize the function $\exp(-H(\cdot))$ such that it integrates to one. To this end, we define the normalizing constant C:

$$C = \int_{-\infty}^{\infty} \exp(-H(\varepsilon))\mathrm{d}\varepsilon\,, \tag{2}$$

such that $p_\varepsilon(\varepsilon) = \exp(-H(\varepsilon))/C$. To ensure that the integral is finite, we impose that $\boldsymbol{t}$ and $\boldsymbol{y}$ are such that (i) either $b_0 > 0$ or $b_0 = 0$ and $a_0 < 0$, and (ii) either $b_k > 0$ or $b_k = 0$ and $a_k > 0$. This restrictions ensures that $\lim_{z\to\infty} H(z) = \infty$ and $\lim_{z\to-\infty} H(z) = \infty$ at a rate that is at least linear. This ensures the integrability of $\exp(-H(z))$. The normalizing constant C is not available in closed form, but can be reliably approximated numerically, as is discussed in Section E of the Supplementary Appendix.

The class of densities that we propose has a log density that is a piecewise quartic polynomial. This flexible density can, for instance, be asymmetric, multi-modal or both. Although our proposed spline density cannot have fat tails by any formal definition, it is not hard to see that if sufficiently many knots are used, any smooth density function with domain $\mathbb{R}$ can be approximated arbitrarily well. Furthermore, if the spline $h(z)$ is linear and goes through the origin, i.e. if $\boldsymbol{t} = c\boldsymbol{y}$ for some $c > 0$ and $h(0) = 0$, the distribution p_ε is a Gaussian distribution with mean zero. Thus, the Gaussian distribution is a special case in this flexible class of spline densities.

Due to the exponentially decaying tails of the density function p_ε, it is clear that $\varepsilon \sim p_\varepsilon(\varepsilon)$ has finite moments of any order. Consider the case where h is rotationally symmetric around the origin, i.e. h is a so-called odd function. Then it is clear that the expected value of $\varepsilon \sim \exp(-H(\varepsilon))/C$ is equal to zero.

In case of asymmetry, however, the expectation of $\varepsilon \sim p_\varepsilon(\varepsilon)$ is not necessarily zero. Since it is our aim to use p_ε as an innovation density with mean zero, we re-center the distribution. Let $p_{\varepsilon^*}(\varepsilon) = \exp(-H(\varepsilon))/C$ denote the uncentered version of p_ε and define

$$\mu := \mathbb{E}_{\varepsilon \sim p_{\varepsilon^*}(\varepsilon)}[\varepsilon] = \int_{-\infty}^{\infty} \varepsilon \, \frac{\exp(-H(\varepsilon))}{C} \, \mathrm{d}\varepsilon \,. \tag{3}$$

Then setting $p_\varepsilon(\varepsilon) = p_{\varepsilon^*}(\varepsilon + \mu) = \exp(-H(\varepsilon + \mu))/C$ ensures that $\mathbb{E}_{\varepsilon \sim p_\varepsilon(\varepsilon)}[\varepsilon] = 0$. This leads to $\partial \log p_\varepsilon(\varepsilon)/\partial\varepsilon = -h(\varepsilon + \mu)$. How to construct the mean μ in practice is discussed in Section E of the Supplementary Appendix.

2.3. Incorporating the spline density in a score-driven model

Let us consider the general framework of score-driven models as introduced by Creal et al. (2013) and Harvey (2013). Say that $\{x_t\}_{t\in\mathbb{Z}}$ is a sequence of observations with elements x_t which are a function of a time-varying parameter f_t and an innovation term $\varepsilon_t \sim p_\varepsilon(\varepsilon_t)$, where p_ε is the spline probability density function defined above and where the elements of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are assumed to have mean zero and are independent and identically distributed (iid) over time. In particular, let $x_t = g(f_t, \varepsilon_t; \lambda)$ for some link function $g(\cdot, \cdot; \lambda)$ that is strictly increasing in its second argument. Here λ is a parameter, which will be omitted from the notation whenever this does not lead to confusion. The time-varying parameter f_t takes values in some space $\mathcal{F} \subseteq \mathbb{R}$. Let $\{\mathcal{G}_t\}_{t\in\mathbb{Z}}$ denote the filtration containing the sigma algebras $\mathcal{G}_t = \sigma(x_t, x_{t-1}, \ldots)$. The time-varying parameter f_t is $\mathcal{G}_{t-1}$-measurable. Thus, we have that

$$\begin{aligned} x_t|\mathcal{G}_{t-1} \sim p_x(x_t|f_t) &= p_\varepsilon(g^{-1}(f_t, x_t)) \cdot \left.\frac{\partial g^{-1}(f_t, x)}{\partial x}\right|_{x=x_t} \\ &= C^{-1} \exp(-H(g^{-1}(f_t, x_t) + \mu)) \cdot \left.\frac{\partial g^{-1}(f_t, x)}{\partial x}\right|_{x=x_t}, \end{aligned} \tag{4}$$

where $g^{-1}(f, x)$ denotes the inverse of the function $g(f, \varepsilon)$ in its second argument. The score-driven update of the time-varying parameter f_t is then specified as

$$\begin{aligned} f_{t+1} &= \omega(1-\beta) + \beta f_t + \alpha s(f_t, x_t)\,, \\ s(f_t, x_t) &= S(f_t) \cdot \nabla(f_t, x_t)\,, \\ \nabla(f_t, x_t) &= \left.\frac{\partial \log p_x(x_t|f)}{\partial f}\right|_{f=f_t} \\ &= -h(g^{-1}(f_t, x_t) + \mu) \cdot \left.\frac{\partial g^{-1}(f, x_t)}{\partial f}\right|_{f=f_t} + \left.\frac{\partial \log((g^{-1})'(f, x_t))}{\partial f}\right|_{f=f_t}, \end{aligned} \tag{5}$$

with $(g^{-1})'(f, x) = \partial g^{-1}(f_t, x)/\partial x$, and where $\nabla(f_t, x_t)$ denotes the score, $S(f_t)$ denotes a positive scaling factor that may depend on f_t, and $s(f_t, x_t)$ is the scaled score. Common choices for the scaling factor $S(f_t)$ are the inverse conditional information matrix with respect to f_t, or its square root. The spline function h and the mean μ, which is defined in (3), depend on the knot vector $\boldsymbol{t}$ and corresponding ordinates $\boldsymbol{y}$, which will be treated below as part of the parameter vector. Next, we introduce two specific choices of g.

Location model. *Let $x_t = g(f_t, \varepsilon_t; \lambda) = f_t + \sigma\varepsilon_t$, where $\lambda = \sigma > 0$ is a parameter, such that $g^{-1}(f_t, x_t; \lambda) = (x_t - f_t)/\sigma$. Then the score takes the form:*

$$\nabla(f_t, x_t) = h((x_t - f_t)/\sigma + \mu)/\sigma\,.$$

Since the conditional Fisher information matrix of f_t is constant, we set $S(f) = 1$ for all $f \in \mathbb{R}$.

Scale model. *Let $x_t = g(f_t, \varepsilon_t; \lambda) = m + \exp(f_t)\varepsilon_t$, such that $g^{-1}(f_t, x_t; \lambda) = (x_t - m)\exp(-f_t)$. Here $\lambda = m$ is a parameter that represents the unconditional mean of the observations. Furthermore, $\exp(f_t)$ provides conditional time-variation in the scale. Then the score takes the form:*

$$\nabla(f_t, x_t) = h((x_t - m)\exp(-f_t) + \mu)(x_t - m)\exp(-f_t) - 1\,.$$

Since the conditional Fisher information matrix of f_t is constant, we set $S(f) = 1$ for all $f \in \mathbb{R}$.

We focus on these two link functions g throughout the remainder of the paper, but there are clearly also other interesting choices of g. For example, a scale model without exponential transformation is obtained by having $g(f, \varepsilon) = f^{1/2}\varepsilon$, for $f > 0$. Also, the introduced model can be straightforwardly extended to the case with more than one time-varying parameter. For instance, it is possible to have both a time-varying location and scale. Furthermore, if the support of ε_t is restricted to the positive part of the real line, one can consider a time-varying conditional duration model (Engle and Russell, 1998) or intensity model (Russell, 2001).

3. Parameter estimation via maximum likelihood

The spline function $h(\cdot; \boldsymbol{t}, \boldsymbol{y})$ can be parametrized in various ways. Estimating both $\boldsymbol{t}$ and $\boldsymbol{y}$ freely is not advisable, due to the high flexibility of the model. Instead, we suggest fixing the knot vector $\boldsymbol{t}$, with elements $t_1 < \ldots < t_k$, and only estimating the corresponding ordinates $\boldsymbol{y}$. We will explain below why this typically offers sufficient flexibility. Intuitively, the elements of $\boldsymbol{y}$ are more convenient to estimate than those of $\boldsymbol{t}$, because the spline coefficients linearly depend on $\boldsymbol{y}$, whereas the elements of $\boldsymbol{t}$ impact the coefficients in a nonlinear fashion. A natural choice for $\boldsymbol{t}$ is the quantile function of the standard normal distribution evaluated on some grid of values on the interval $(0, 1)$ that is symmetric around 0.5. The approach of Kooperberg and Stone (1991), which uses the order statistics of sample observations, is not feasible, because this would require having the order statistics of the residuals, which themselves depend on the spline function that determines the score.

We define the parameter vector $\boldsymbol{\theta}$ that contains all the static parameters of the model

$$\boldsymbol{\theta} = (\omega, \beta, \alpha, \boldsymbol{\gamma}^\top)^\top , \quad \text{with } \boldsymbol{\gamma} = (\lambda, \boldsymbol{\psi}^\top)^\top ,$$

where λ is a scalar and $\boldsymbol{\psi}$ is the q-dimensional vector that contains the parameters that determine $\boldsymbol{y}$, such that $\tau(\boldsymbol{\psi}) = \boldsymbol{y}$ for some continuous function $\tau : \mathbb{R}^q \to \mathbb{R}^k$. For example, we could let $\boldsymbol{\psi} = \boldsymbol{y}$, allowing for potential asymmetry. We let $\lambda \in \boldsymbol{\Lambda} \subseteq \mathbb{R}$, $\boldsymbol{\psi} \in \boldsymbol{\Psi} \subseteq \mathbb{R}^q$ and $(\omega, \beta, \alpha)^\top \in \boldsymbol{\Pi} \subseteq \mathbb{R}^3$. Also, we define the notation $\boldsymbol{\Gamma} := \boldsymbol{\Lambda} \times \boldsymbol{\Psi}$. It follows that the vector $\boldsymbol{\theta}$ takes values in a parameter space $\Theta = \boldsymbol{\Pi} \times \boldsymbol{\Gamma}$. We will impose restrictions on Θ such that the spline-based density p_ε is well-defined for any $\boldsymbol{\theta} \in \Theta$.

We assume that the observed time series $\{x_t\}_{t=1}^T$ is a subset of a sequence $\{x_t\}_{t\in\mathbb{Z}}$ that is generated from the model defined above for some parameter $\boldsymbol{\theta}_0 \in \Theta$. Our aim is to estimate this parameter $\boldsymbol{\theta}_0$ based on such a sequence of observations using ML. Since we do not observe the sequence $\{f_t\}_{t=1}^T$, we need to calculate a filtered sequence $\{\hat{f}_t(\boldsymbol{\theta})\}_{t=1}^T$ in order to be able to construct the log likelihood:

$$\hat{f}_{t+1}(\boldsymbol{\theta}) = \omega(1-\beta) + \beta \hat{f}_t(\boldsymbol{\theta}) + \alpha s(\hat{f}_t(\boldsymbol{\theta}), x_t; \boldsymbol{\gamma}) , \tag{6}$$

for $t = 1, \ldots, T-1$, initialized at some fixed real value $\hat{f}_1$. The function $s(\cdot, \cdot; \cdot) : \mathcal{F} \times \mathbb{R} \times \boldsymbol{\Gamma} \to \mathbb{R}$ is defined as s in (5), but now with the dependence on the parameters in γ explicitly stated in the notation. The ML estimator is then defined as

$$\widehat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta}),$$

$$\begin{aligned}\text{where} \quad \hat{\ell}_t(\boldsymbol{\theta}) &= \log p_x(x_t|\hat{f}_t(\boldsymbol{\theta});\boldsymbol{\gamma}) \\ &= -H(g^{-1}(\hat{f}_t(\boldsymbol{\theta}), x_t;\lambda) + \mu(\boldsymbol{\psi});\boldsymbol{\psi}) - \log(C(\boldsymbol{\psi})) \\ &\quad + \log\left(\left.\frac{\partial g^{-1}(\hat{f}_t(\boldsymbol{\theta}), x;\lambda)}{\partial x}\right|_{x=x_t}\right),\end{aligned}$$

and where $C(\boldsymbol{\psi})$ and $\mu(\boldsymbol{\psi})$ denote the normalization constant and mean defined in (2) and (3), respectively, under knot vector $\boldsymbol{t}$ and ordinate vector $\boldsymbol{y} = \tau(\boldsymbol{\psi})$. Also, we let $H(\cdot;\boldsymbol{\psi}) := H(\cdot;\boldsymbol{t},\tau(\boldsymbol{\psi}))$. Furthermore, from now on we use $h(\cdot;\boldsymbol{\psi}) := h(\cdot;\boldsymbol{t},\tau(\boldsymbol{\psi}))$ whenever convenient. Note that $p_x(x_t|f_t;\boldsymbol{\gamma})$ is simply the observation density in (4), with the dependence on the parameters in $\boldsymbol{\gamma}$ made explicit.

Scaling of the knot vector

Fixing the knot vector a priori seems restrictive, but by having a scale parameter in the model, the knots can effectively be rescaled. For instance, take the function $g(f_t,\varepsilon_t) = f_t + \sigma\varepsilon_t$. Then for any $\sigma > 0$ and $x \in \mathbb{R}$, the spline function h is such that:

$$h(x;\boldsymbol{t},\boldsymbol{y}) = h\left(\frac{x}{\sigma};\frac{\boldsymbol{t}}{\sigma},\boldsymbol{y}\right) = \frac{1}{\sigma}h\left(\frac{x}{\sigma};\frac{\boldsymbol{t}}{\sigma},\sigma\boldsymbol{y}\right),$$

where the first equality holds by definition of the natural cubic spline function and the second equality holds because the set of spline coefficients, and therefore the spline function itself, is linear in $\boldsymbol{y}$. It follows that the anti-derivative H is such that for any $\sigma > 0$ and for any $x \in \mathbb{R}$:

$$H(x;\boldsymbol{t},\boldsymbol{y}) = H\left(\frac{x}{\sigma};\frac{\boldsymbol{t}}{\sigma},\sigma\boldsymbol{y}\right), \tag{7}$$

up to some constant. Therefore, using σ, $\boldsymbol{t}$, and $\boldsymbol{y}$ is equivalent to using $d\sigma$, $\boldsymbol{t}/d$ and $d\boldsymbol{y}$ for any constant $d > 0$. Also, it can be shown that $h(x + \mu(\boldsymbol{t},\boldsymbol{y});\boldsymbol{t},\boldsymbol{y}) = h(x/\sigma + \mu(\boldsymbol{t}/\sigma,\sigma\boldsymbol{y});\boldsymbol{t}/\sigma,\sigma\boldsymbol{y})/\sigma$, where $\mu(\boldsymbol{t},\boldsymbol{y})$ denotes the value in (3) for $H(\cdot;\boldsymbol{t},\boldsymbol{y})$, so this equivalence also holds if a nonzero mean is corrected for.

Enforcing integrability of spline-based density

In practice, it is advisable to ensure that $\boldsymbol{\psi}$ is such that the spline-based density is well-defined during the numerical optimization of the log likelihood function. Given $\boldsymbol{t}$, the condition that $\boldsymbol{y}$ is such that h has positive slopes beyond the outer knots, i.e. $b_0 > 0$ and $b_k > 0$, can be reduced to two linear inequality constraints in $\boldsymbol{y}$, since we have linearity of the spline coefficients in $\boldsymbol{y}$. As it is not convenient to impose multiple inequality constraints for a particular vector of parameters, we suggest reparametrizing the spline function, where instead of estimating y_1 and y_k alongside the other elements in $\boldsymbol{y}$, we directly estimate b_0 and b_k themselves, and then reconstruct the values of y_1 and y_k based on $(y_2, \dots, y_{k-1})$ and b_0 and b_k. This reconstruction is straightforward, due to the linearity of b_0 and b_k in $\boldsymbol{y}$. The advantage of this parametrization is that $b_0 > 0$ and $b_k > 0$ can be enforced directly on these two separate parameters, instead of indirectly through functions of multiple parameters.

In order to impose $b_0 = 0$ and $a_0 < 0$ and $b_k = 0$ and $a_k > 0$, we need a different parametrization, as we have that $a_0 = y_1$ and $a_k = y_k$. In this case, it is preferable to exclude y_2 and y_{k-1} from the parameter vector and instead estimate $y_1, y_3, \dots, y_{k-2}$, and y_k, where $a_0 = y_1 > 0$ and $a_k = y_k > 0$ is imposed and where y_2 and y_{k-1} are constructed based on $y_1, y_3, \dots, y_{k-2}, y_k$ and $b_0 = b_k = 0$.

4. Asymptotic properties of maximum likelihood estimator

In this section, we derive the asymptotic theory of the maximum likelihood estimator (MLE) of the score-driven spline model for the location and the scale examples. We derive our results for the setting where the time-varying parameter and observations are stationary and ergodic (SE). Stationarity is not required for showing consistency and asymptotic normality of the MLE, see for instance Blasques et al. (2024) for an example of a score-driven location model with unit root dynamics, but as non-stationary dynamics require a more elaborate and intricate approach, we focus on the stationary setting.

For both models, we impose the following restrictions on the parameter set Θ:

Assumption 1.

(i) *$\Theta \subset \mathbb{R}^{4+q}$ is a compact set and the knot vector $\boldsymbol{t}$ is fixed and has k elements that are strictly increasing.*

(ii) *$\tau : \mathbb{R}^q \to \mathbb{R}^k$ is a continuous function such that for every $\boldsymbol{\theta} \in \Theta$, $\tau(\boldsymbol{\psi}) = \boldsymbol{y}$ is such that the coefficients of $h(\cdot; \boldsymbol{t}, \boldsymbol{y})$ are such that (i) either $b_0 > 0$ or $b_0 = 0$ and $a_0 < 0$, and (ii) either $b_k > 0$ or $b_k = 0$ and $a_k > 0$.*

(iii) *k is even, $\boldsymbol{t}$ is symmetric around 0, $\boldsymbol{\psi} = (\psi_1, \dots, \psi_q)^\top$ for $q = k/2$, and τ is such that $y_1 = -\psi_q$, $y_2 = -\psi_{q-1}$, $\dots$, $y_q = -\psi_1$, $y_{q+1} = \psi_1$, $\dots$, $y_k = \psi_q$.*

Condition (ii) ensures that the log likelihood is well-defined for any $\boldsymbol{\theta} \in \Theta$. Condition (iii) ensures that the spline function is rotationally symmetric around the origin. This condition is not strictly necessary, but allowing for asymmetry leads to more cumbersome derivations due to the inclusion of the function $\mu(\boldsymbol{\psi})$ in the log likelihood and the score. The derivations can also be easily adjusted to the case where y_1 and/or y_k in $\boldsymbol{\psi}$ are replaced by b_0 and/or b_k, in order to be able to conveniently enforce non-negative slopes beyond the outer knots, as b_0 and b_k are linear in $\boldsymbol{y}$.

For feasibility of the theoretical derivations, (iii) of Assumption 1 imposes that the knots are given a priori. We will also impose correct specification of the density and we note that this assumption, together with having a fixed number of knots, is not in line with a semiparametric setting. This is in accordance with the literature on the mixed normal distribution, where the number of mixing components is considered fixed for the purpose of deriving asymptotic properties of estimators. In practice, the number of knots can be selected using an information criterion. Furthermore, theoretical results can be straightforwardly extended to allow for potential misspecification of the density and the parametric part of the model as in Blasques et al. (2022). This requires imposing alternative assumptions on the data generating process and the parameter estimate then converges to a pseudo-true parameter as the sample size grows to infinity. For ease of exposition, we limit ourselves to the correctly specified case.

In this section, let $\|\cdot\|$ denote the L_p-norm for some $p \geq 1$ when applied to a vector, and when applied to a matrix let it denote the operator norm induced by this L_p-norm.

4.1. Location model

We start with the location model defined above, for which we put in place the following basic assumptions

Assumption 2.

(i) Let $\{x_t\}_{t\in\mathbb{Z}}$ be generated according to equations (4) and (5) under $g(f,\varepsilon;\lambda) = f + \sigma\varepsilon$ and $S(f) = 1$ for all $f \in \mathcal{F} = \mathbb{R}$, for some $\boldsymbol{\theta}_0 \in \Theta$, where $\lambda = \sigma^2$. Let Θ be such that $\sigma^2 > 0$ for any $\boldsymbol{\theta} \in \Theta$.

(ii) Let $|\beta| < 1$ for all $\boldsymbol{\theta} \in \Theta$.

Condition (i) ensures correct specification. The additional condition (ii) allows us to establish stationarity, ergodicity and bounded moments of the sequence of observations $\{x_t\}_{t\in\mathbb{Z}}$ using an application of Proposition 3.1 of Blasques et al. (2022):

Proposition 1. *Let Assumptions 1 and 2 be satisfied. Then for any* $\boldsymbol{\theta}_0 \in \Theta$*, the sequence of true time-varying components* $\{f_t\}_{t\in\mathbb{Z}}$ *and the sequence of observations* $\{x_t\}_{t\in\mathbb{Z}}$ *are SE with bounded moments of any order.*

Now we turn to the invertibility and asymptotic properties of the filter defined in (6). Filter invertibility ensures that the filter does not depend on the deterministic initialization $\hat{f}_1$ in the limit. This is important, as it ensures that the filter evaluated in $\boldsymbol{\theta}_0$ converges to the true time-varying parameter $\{f_t\}_{t\in\mathbb{Z}}$. See for instance Blasques et al. (2018) for a more thorough discussion on filter invertibility. To ensure invertibility, we need the following additional assumption:

Assumption 3. *Let the following contraction condition hold:*

$$\sup_{z\in\mathbb{R},\boldsymbol{\theta}\in\Theta} \left| \beta - \alpha \frac{h'(z;\boldsymbol{t},\boldsymbol{y})}{\sigma^2} \right| < 1\,.$$

In order to verify whether this contraction condition holds, we need to consider the derivative of the cubic spline function h, i.e. $h'(z) = \partial h(z)/\partial z$. This derivative looks as follows:

$$h'(z) = \begin{cases} h_0'(z) & \text{if } z \leq t_1\,, \\ h_i'(z) & \text{if } t_i < z \leq t_{i+1}\,,\ \text{ for } i = 1,\ldots,k-1\,, \\ h_k'(z) & \text{if } z > t_k\,, \end{cases}$$

where h_i' for $i = 0,\ldots,k-1$ is given by:

$$h_i'(z) = b_i + 2c_i(z - t_{i+1}) + 3d_i(z - t_{i+1})^2\,,$$

and

$$h_k'(z) = b_k\,,$$

where b_i, c_i and d_i are the spline coefficients of the spline $h(\cdot;\boldsymbol{t},\boldsymbol{y})$, implying that $b_0 = d_0 = 0$. By the linearity of the spline h beyond the outer knots, the derivative h' is constant beyond the outer knots. Due to the restrictions on the parameter set in Assumption 1, it follows that $h'(z)$ is uniformly bounded over $z \in \mathbb{R}$ and $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Thus, there is a nonempty parameter space Θ, for which this assumption holds.

Besides filter invertibility, the proposition below establishes the stationarity, ergodicity, and existence of bounded moments of the limit filter uniformly over the parameter space Θ. This will be useful for establishing consistency of the MLE. Furthermore, it establishes convergence of the first and second order derivatives of the process of filtered values to SE limit sequences with bounded moments of any order, which will be used for showing asymptotic normality. The first two results follow from applications of Propositions 3.2 and 3.4 of Blasques et al. (2022). For the third result, an adaptation of the latter proposition is needed, because the cubic spline function is only twice continuously differentiable.

Proposition 2. *Let Assumptions 1, 2 and 3 hold. Then, the following results hold:*

(i) The sequence $\{\hat{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges exponentially fast almost surely[2] (e.a.s.) uniformly over Θ to a unique SE limit sequence $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ as $t\to\infty$, in other words $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t(\boldsymbol{\theta})-f_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ as $t\to\infty$. Furthermore, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^n<\infty$ for any $n>0$.

(ii) The sequence $\{\partial\hat{f}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}_{t\in\mathbb{N}}$ converges e.a.s. uniformly over Θ to an SE sequence $\{\partial f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}_{t\in\mathbb{Z}}$ with $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\partial f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\|^n<\infty$ for any $n>0$.

(iii) The sequence $\{\partial^2\hat{f}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\}_{t\in\mathbb{N}}$ converges e.a.s. uniformly over Θ to an SE sequence $\{\partial^2 f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\}_{t\in\mathbb{Z}}$ with $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\partial^2 f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\|^n<\infty$ for any $n>0$.

Based on the results of Propositions 1 and 2, we can now establish strong consistency of the MLE $\widehat{\boldsymbol{\theta}}_T$ under the following additional conditions:

Assumption 4. *Let*

(i) $\alpha_0\neq 0$.

(ii) For $b,b'>0$ and $\boldsymbol{\psi},\boldsymbol{\psi}'\in\boldsymbol{\Psi}$, let $h(z/b;\boldsymbol{\psi})=h(z/b';\boldsymbol{\psi}')$ hold for every $z\in\mathbb{R}$ if and only if $(b,\boldsymbol{\psi})=(b',\boldsymbol{\psi}')$.

Conditions (i) and (ii) ensure identification of $\boldsymbol{\theta}_0$. More specifically, condition (i) rules out the degeneracy of f_t and (ii) establishes the identification of $\boldsymbol{\psi}$ and σ. A sufficient condition for condition (ii) to hold is that the spline function $h(x;\boldsymbol{t},\boldsymbol{y})$ is such that its second order derivative in x is differentiable in none of the knots. Intuitively, if h'' is such that it is differentiable at none of the knots, changing σ, will inherently change the spline function $h(x/b;\boldsymbol{\psi})$, as it will change the position of the discontinuities in the third order derivative of h. On the other hand, in the extreme case that h'' is differentiable in every knot, there is a clear lack of identification, as this means that $h(z)=az$ for some constant a, such that $H(z)=\frac{1}{2}az^2$. In that case, having $(b,\boldsymbol{\psi})$ is clearly equivalent to having $(b/\sqrt{d},d\boldsymbol{\psi})$ for any $d>0$, which implies that in this model $\boldsymbol{\psi}$ and σ are not separately

[2]The sequence $\{\hat{y}_t\}$ converges exponentially fast almost surely to the sequence $\{y_t\}$, if $c^t|\hat{y}_t-y_t|\overset{a.s.}{\to}0$ for some $c>1$ as $t\to\infty$.

identified. Even if h'' is differentiable only in a selection of the knots, an identification problem might occur, as a different scaling of the knots could lead to an identical spline function if the elements of $\boldsymbol{\psi}$ are appropriately re-scaled too. In conclusion, imposing that the third order derivative of the spline h takes different values on each segment $(-\infty, t_1)$, (t_1, t_2), ..., (t_k, ∞) is sufficient for condition (ii) of Assumption 4 to hold.

The next result provides the consistency of the MLE. We note that we cannot apply the consistency theorem of Blasques et al. (2022), because it relies on the boundedness of the log likelihood function and the score over the values of the time-varying parameter $f \in \mathcal{F}$, which does not hold for our current framework.

Theorem 1 (Consistency of MLE for location model)**.** *Let $\{x_t\}_{t=1}^T$ be a subset of consecutive observations from the sequence $\{x_t\}_{t\in\mathbb{Z}}$ for which Assumptions 1, 2, 3, and 4 are satisified. Then the MLE satisfies $\widehat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.*

Next, we establish the asymptotic normality of the MLE under the additional assumption that $\boldsymbol{\theta}_0$ lies in the interior of the parameter space Θ. Also in this case, the result of Blasques et al. (2022) cannot be used because not all required quantities are uniformly bounded in the time-varying parameter f_t in this model.

Theorem 2 (Asymptotic normality of MLE for location model)**.** *Let the assumptions of Theorem 1 be satisfied and let $\boldsymbol{\theta}_0$ be in the interior of Θ. Then, the MLE $\widehat{\boldsymbol{\theta}}_T$ satisfies $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{d}{\to} \mathcal{N}(0, \mathcal{I}^{-1})$ as $T \to \infty$, where $\mathcal{I} := -\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)]$ is the Fisher information matrix.*

4.2. Scale model

For the scale model, we impose the following specification assumption.

Assumption 5.

(i) Let $\{x_t\}_{t\in\mathbb{Z}}$ be generated according to equations (4) and (5) under $g(f, \varepsilon; \lambda) = m + \exp(f)\varepsilon$ and $S(f) = 1$ for all $f \in \mathbb{R}$, for some $\boldsymbol{\theta}_0 \in \Theta$, and let $\lambda = m$.

(ii) Let $0 < \beta < 1$ and $\alpha \geq 0$ for any $\boldsymbol{\theta} \in \Theta$, such that f_t takes values in $\mathcal{F} = [\underline{f}, \infty)$ for some $\underline{f} > -\infty$.

Condition (i) ensures correct specification and condition (ii) ensures that f_t is bounded from below by some finite value, because we know that under Assumption 1, the score function is uniformly bounded from below by some finite constant:

$$\inf_{\boldsymbol{\theta}\in\Theta, x\in\mathbb{R}, f\in\mathbb{R}} s(f, x; \boldsymbol{t}, \boldsymbol{y}) = \inf_{\boldsymbol{\theta}\in\Theta, z\in\mathbb{R}} h(z; \boldsymbol{t}, \boldsymbol{y})z - 1 = \underline{s} > -\infty\,.$$

More specifically, this holds because Assumption 1 ensures that the piecewise polynomial $h(\cdot)$ is such that $\lim_{z\to+\infty} h(z) = +\infty$ and $\lim_{z\to-\infty} h(z) = -\infty$ (or $\lim_{z\to+\infty} h(z) = c > 0$ and $\lim_{z\to-\infty} h(z) = -c < 0$ for some constant $c > 0$) for any $\boldsymbol{\theta} \in \Theta$. Due to the exponential transformation of f_t in the observation equation, the model is also well-defined without having a lower bound for f_t, but such a bound is necessary for the derivations of the theoretical results, as $\exp(-f_t)$ occurs in the score and in the log likelihood function. The proposition below establishes the stationarity and ergodicity of the observations under Assumptions 1 and 5.

Proposition 3. *Let Assumptions 1 and 5 be satisfied. Then for any $\boldsymbol{\theta}_0 \in \Theta$, the sequence of true time-varying components $\{f_t\}_{t\in\mathbb{Z}}$ and the sequence of observations $\{x_t\}_{t\in\mathbb{Z}}$ are SE, and $\{f_t\}_{t\in\mathbb{Z}}$ has bounded moments of any order.*

Given that the time-varying parameters and the observations are SE, we can now establish filter invertibility under the following contraction condition:

Assumption 6. *Let the following contraction condition hold:*

$$\mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{f\in\mathcal{F}} \left| \beta - \alpha \left[h'(z; \boldsymbol{t}, \boldsymbol{y}) z^2 + h(z; \boldsymbol{t}, \boldsymbol{y}) z \right] \Big|_{z=(x_t-m)\exp(-f)\sigma^{-1}} \right| < 0\,.$$

The function $h'(z)z^2 + h(z)z$ is either quadratic or linear in z beyond the outer knots, so it is not bounded in z, which makes an invertibility condition that is also uniform over the observations inapplicable. This complicates the derivation of bounded moments of the derivative processes $\{\partial f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}_{t\in\mathbb{Z}}$ and $\{\partial^2 f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\}_{t\in\mathbb{Z}}$, which is why we refrain from establishing asymptotic normality for this model. As we impose that $\inf_{f\in\mathcal{F}} f = \underline{f} > -\infty$, it is clear that there will be a nonempty parameter set Θ for which Assumption 6 holds. It is not straightforward to evaluate whether the contraction condition holds in practice, as it requires evaluating an expectation. See Blasques et al. (2018) for a feasible approach for verifying such a condition.

Proposition 4. *Let Assumptions 1, 5 and 6 be satisfied. Then the sequence $\{\hat{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE limit sequence $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ as $t\to\infty$ uniformly over Θ, i.e. $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ as $t\to\infty$.*

Based on the results of the proposition above and under the identification conditions in Assumption 4, we can now derive the consistency result below:

Theorem 3 (Consistency of MLE for scale model)**.** *Let $\{x_t\}_{t=1}^T$ be a subset of consecutive observations from the sequence $\{x_t\}_{t\in\mathbb{Z}}$ for which Assumptions 1, 5, 6 and 4 are satisfied. Then the MLE satisfies $\widehat{\boldsymbol{\theta}}_T \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T\to\infty$.*

5. Empirical applications

5.1. Location: U.S. monthly inflation

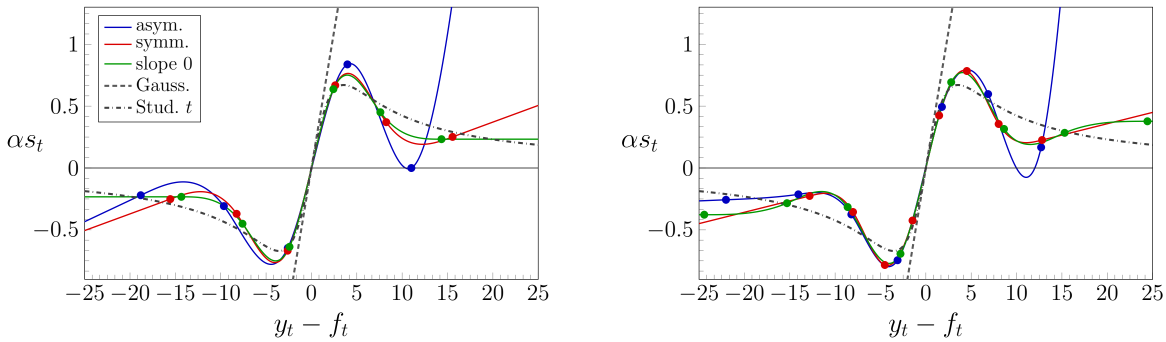
To illustrate the workings of our proposed score-driven spline location model, we apply the model to the monthly U.S. inflation from January 1950 to September 2024 (i.e. $T = 897$). In particular, we consider the percentage change in the consumer price index[3] with respect to the month before and we scale the data by a factor 10 to improve numerical stability. Our score-driven spline location model allows us to filter the conditional expectation of the observations in a flexible manner. As the inflation time series has occasional spikes, see Figure 2, a nonlinear filter will be more suitable than a linear filter.

We consider the spline location model as defined in Section 4.1 with various configurations. In particular, we apply the model with $k = 6$ and $k = 8$ knots. Having more knots does not substantially improve the fit. For the knot vector we choose $\boldsymbol{t} = 10(\Phi^{-1}(1/(k+1)), \Phi^{-1}(2/(k+1)), \ldots, \Phi^{-1}(k/(k+1)))^\top$, where Φ^{-1} denotes the quantile function of the standard normal distribution. The knot vector is scaled by 10 to enhance numerical stability, as it ensures that σ^2 has a comparable magnitude as the other parameters in this example. For both choices of k, we apply versions that are (i) asymmetric, (ii) symmetric and (iii) symmetric with the slopes beyond the outer knots fixed at zero, from now on referred to as 'slope 0'. We use the parametrization discussed at the bottom of Section 3, i.e. instead of estimating y_1 and/or y_k, we directly estimate the outer slopes b_0 and/or b_k. As a benchmark, we consider the special case where h is forced to be linear, which boils down to using a Gaussian distribution. In particular, in that case we set $\boldsymbol{y} = \boldsymbol{t}$, such that $\boldsymbol{\gamma} = \sigma$. Finally, for comparison, we consider the score-driven location model based on a Student's t distribution, see Harvey and Luati (2014). The parameters are estimated using ML. As the initial value for the filters, we use the unconditional expectation of the time-varying component, i.e. ω. Furthermore, we use a burn-in period of 12 observations to allow the filter to converge.

The parameter estimates are displayed in Table 1 alongside the log likelihood values and Akaike's Information Criterion (AIC), Takeuchi's Information Criterion (TIC), see Takeuchi (1976), and the Bayesian Information Criterion (BIC) values. The TIC is a misspecification-robust version of the AIC, that unlike the AIC does not depend on the

[3] We use 'Consumer Price Index for All Urban Consumers: All Items in U.S. City Average' [CPIAUCSL] provided by the U.S. Bureau of Labor Statistics. Data were retrieved from FRED, Federal Reserve Bank of St. Louis.

information matrix equality to be valid. The corresponding score functions are plotted in Figure 1. It stands out that all spline models have a comparable fit. The symmetric model with $k = 6$ knots and slopes of zero beyond the outer knots has the best AIC and BIC out of the spline models. Figure 1 shows that this model yields a robust filter, as values close to zero receive a relatively larger update than values far from zero. The asymmetric models lead to a better fit in terms of log likelihood, but the additional three parameters do not lead to a sufficient improvement according to the information criteria. Figure 1 shows that the left and middle parts of the asymmetric splines are similar to the symmetric splines, but the right part is notably higher in magnitude.



(a) Spline models with $k = 6$ knots.

(b) Spline models with $k = 8$ knots.

Figure 1. Score function plots multiplied by α corresponding to the parameter estimates in Table 1. The effective position of the knots and corresponding ordinates are marked with dots.

The Gaussian model has the worst fit according to the information criteria, which can be explained by the linear filter not being suitable for these data. Due to its linearity, the filter is sensitive to outliers, which also explains why the estimated persistence parameter β is lower for the Gaussian model than for the other models, as a lower β dampens the effect of such outliers. The Student's t model has a lower log likelihood value than the spline models, but due to its low number of parameters, it has the best BIC value out of all models considered, although the difference with the 'slope 0' model with $k = 6$ knots is small. On the other hand, in terms of AIC, the symmetric spline models are preferred over the Student's t model, and according to the TIC, which requires fewer assumptions to be valid, even all nonlinear spline models are preferred over the Student's t model. Figure 1 shows that the score function of this model bears resemblance to those of the symmetric score models, with the important difference that the Student's t score redescends to zero, whereas the score functions of the spline models diverge linearly or

Table 1. Maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_T$ for U.S. inflation data.

	Spline							
	$k=6$			$k=8$				
	asym.	symm.	slope 0	asym.	symm.	slope 0	Gauss.	Stud. t
ω	2.582	2.512	2.506	2.545	2.533	2.518	2.825	2.498
	(0.555)	*(0.603)*	*(0.609)*	*(0.575)*	*(0.592)*	*(0.601)*	*(0.298)*	*(0.678)*
β	0.966	0.967	0.968	0.968	0.967	0.967	0.833	0.969
	(0.012)	*(0.012)*	*(0.012)*	*(0.012)*	*(0.012)*	*(0.012)*	*(0.115)*	*(0.012)*
α	0.955	0.940	0.939	0.946	0.942	0.940	2.700	0.377
	(0.154)	*(0.151)*	*(0.151)*	*(0.153)*	*(0.151)*	*(0.150)*	*(1.000)*	*(0.059)*
σ^2	3.335	2.124	1.804	3.062	1.102	4.005	6.103	3.089
	(1.589)	*(0.555)*	*(0.331)*	*(0.353)*	*(0.371)*	*(1.053)*	*(0.527)*	*(0.246)*
ν								4.093
								(0.563)
b_0	0.121			0.010				
	(0.179)			*(0.136)*				
y_2	-0.589			-0.394				
	(0.165)			*(0.375)*				
y_3	-1.237			-0.694				
	(0.354)			*(0.246)*				
y_4	1.602	1.036	0.911	-1.379				
	(0.573)	*(0.196)*	*(0.114)*	*(0.208)*				
y_5	0.000	0.574	0.644	0.913	0.473	1.476		
	(0.298)	*(0.103)*	*(0.048)*	*(0.174)*	*(0.170)*	*(0.283)*		
y_6/b_k	2.311	0.061	0	1.105	0.874	0.672		
	(4.851)	*(0.107)*	-	*(0.158)*	*(0.121)*	*(0.194)*		
y_7				0.309	0.396	0.605		
				(0.400)	*(0.078)*	*(0.369)*		
b_8				4.065	0.021	0		
				(2.062)	*(0.042)*	-		
LLH	-1977.34	-1978.55	-1978.85	-1975.53	-1978.31	-1978.45	-2056.15	-1981.95
AIC	3974.67	3971.10	3969.70	3975.06	3972.61	3970.90	4120.30	3973.90
TIC	3976.80	3973.98	3972.00	3973.61	3975.31	3973.93	4143.34	3977.47
BIC	4022.53	4004.60	3998.41	4032.49	4010.89	4004.39	4139.45	3997.83

Standard errors reported in brackets. The columns 'slope 0' concern the symmetric spline model where the slopes beyond the outer knots are fixed to 0. LLH, AIC, TIC and BIC are the log likelihood, Akaike's Information Criterion, Takeuchi's Information Criterion and Bayesian Information Criterion, respectively.

converge to a nonzero constant.

The filtered conditional location path of the symmetric spline model with $k = 6$ knots

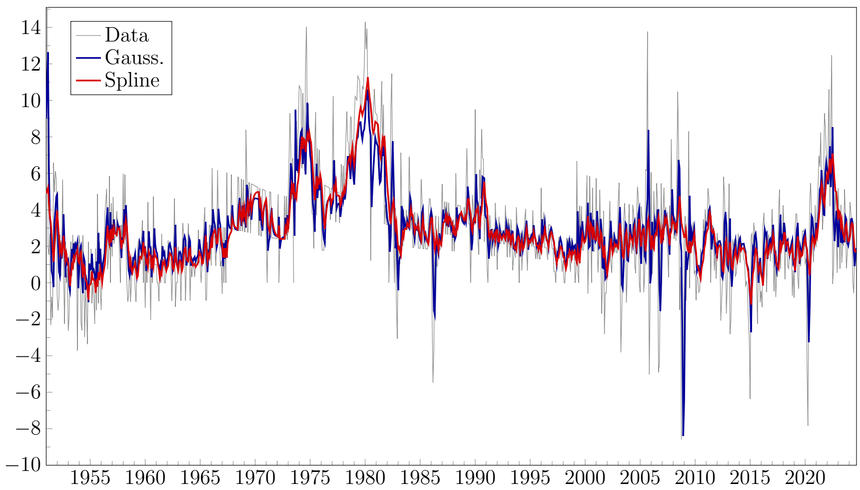


Figure 2. Filtered location for U.S. monthly inflation for the Gaussian and symmetric spline location model with $k = 6$ knots with slopes beyond outer knots fixed at zero.

is plotted alongside that of the Gaussian model in Figure 2. The spline filter's robustness is clearly visible, as occasional spikes do not heavily impact the filtered location for the spline model, whereas for the Gaussian model, there is a large distortion in the filtered path whenever there is an outlier. Depending on the magnitude of the spike, the filter needs many months to recover from such a distortion.

5.2. Scale: S&P 500 stock returns

To demonstrate the broad applicability of our proposed methodology, we apply the scale model as defined in Section 4.2, to all stocks in the Standard & Poor's (S&P) 500 index. We take daily log returns from January 3, 2013 until August 21, 2024[4], and we only include companies that were part of the S&P 500 index during this entire period. This gives us 456 companies with $T = 2,928$ observations each.

We apply the spline model with $k = 6$ knots, where we again consider versions with general h, symmetric h and symmetric h with slopes of zero outside the outer knots. We let $\boldsymbol{t} = 10(\Phi^{-1}(1/14), \Phi^{-1}(4/14), \Phi^{-1}(6/14), \Phi^{-1}(8/14), \Phi^{-1}(10/14), \Phi^{-1}(13/14))^\top$. Thus, here we do not take the standard normal quantile function of an equally spaced grid of points in $[0, 1]$, but instead the knots further away from 0.5 are more spread out. The reason is that for this financial application, we expect there to be more intricate behaviour of the score towards the tails of the distribution. We again consider the version of our model with Gaussian and Student's t errors as two benchmark models for comparison. Recall that the Gaussian model is a special case of the spline model and

[4]Data were retrieved from Yahoo! Finance using the yfinance API for Python.

bears resemblance to the seminal GARCH model of Engle (1982) and Bollerslev (1986), although here we model the conditional log volatility instead of the conditional volatility. Furthermore, the Student's t version is also known as the Beta-t-EGARCH model (Harvey and Chakravarty, 2008).

For each company under consideration, we estimate the parameters of the five models using ML. The log volatility filter is again initialized at the unconditional expectation ω. Furthermore, the first 30 observations of the sample are burned to allow the filters to converge. For the asymmetric and symmetric spline model, some of the numerical optimizations fail, possibly due to the high flexibility of these models. Thus, in some cases we might not have found the true ML estimator, and these results should therefore be taken as a lower bound of how well these models can perform. A summary of the results is presented in Table 2. The first panel shows how often each model has the highest log likelihood or the lowest AIC and BIC out of the 456 companies. It must be noted that the model with slopes of zero enforced beyond the outer knots surprisingly has the highest log likelihood for one stock. This can be explained by the fact that the asymmetric and symmetric models, a slope strictly greater than 0 is imposed, such that these models do not exactly nest the 'slope 0' model. Furthermore, the table shows that in more than 90% of the cases, the spline models have a better log likelihood than the Student's t model. The Gaussian version of the model is never selected by the information criteria. According to the AIC, the spline models are best for over 60% of the companies, while according to the BIC, this is 20%. Between the spline models, the most parsimonious model, i.e. the 'slope 0' model, is selected by the AIC in the majority of the cases, and by the BIC for more than 98% of the companies. For this reason, the last sub panel of this panel shows a direct comparison of this model and the Student's t model, where it can be seen that for just over 50% of the companies, the AIC selects the 'slope 0' model, while the BIC selects this model in about 20% of the cases. Thus, for a substantial portion of the companies, the spline model leads to better results than the competitive Student's t benchmark model, which underlines the broad applicability of this model.

The second panel of Table 2 shows the average values of the log likelihood and information criteria over all companies. The average AIC of all spline models is lower than that of the benchmark models. When considering all time series jointly, ignoring any dependence between the time series, the AIC would thus select the symmetric spline model with a slope of 0 beyond the outer knots imposed. On the other hand, the average BIC is lowest for the Student's t model. Finally, the bottom panel shows that whenever the

Table 2. Summary results of scale models applied to log returns of 456 S&P 500 stocks between January 3, 2013 and August 21, 2024.

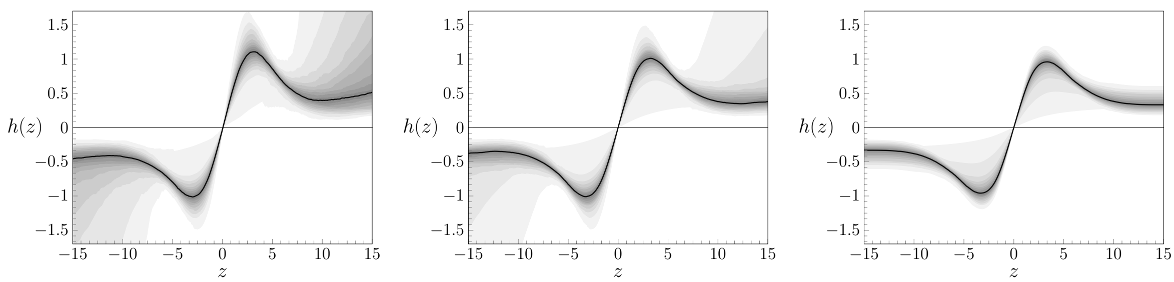
	Spline				
	asym.	symm.	slope 0	Stud. t	Gauss.
perc. best					
LLH	92.32%	0.00%	0.22%	7.46%	0.00%
AIC	26.75%	7.68%	27.85%	37.72%	0.00%
BIC	0.00%	0.66%	20.18%	79.17%	0.00%
LLH	99.78%	0.00%	0.22%	-	
AIC	32.24%	8.33%	59.43%	-	-
BIC	0.66%	1.10%	98.25%	-	-
LLH	-	-	69.52%	30.48%	-
AIC	-	-	51.75%	48.25%	-
BIC	-	-	20.39%	79.61%	-
average					
LLH	-5307.11	-5310.38	-5310.78	-5312.64	-5572.76
AIC	10634.23	10634.76	10633.56	10635.28	11153.51
BIC	10693.94	10676.57	10669.39	10665.14	11177.40
cond. avg. diff.					
LLH	-	-	3.289	1.402	-
AIC	-	-	-6.536	-3.449	-
BIC	-	-	-7.102	-7.163	-

'Slope 0' denotes the spline model where the spline is forced to have slope 0 beyond the outer knots. The first panel shows the percentage of companies for which each model is selected according to the information criteria, where each sub-panels compares different models. The middle panel shows the average criterion values. The bottom panel shows the criterion value of the 'slope 0' model minus that of the Student's t model, conditional on the former being better according to that criterion (left column) and vice versa (right column).

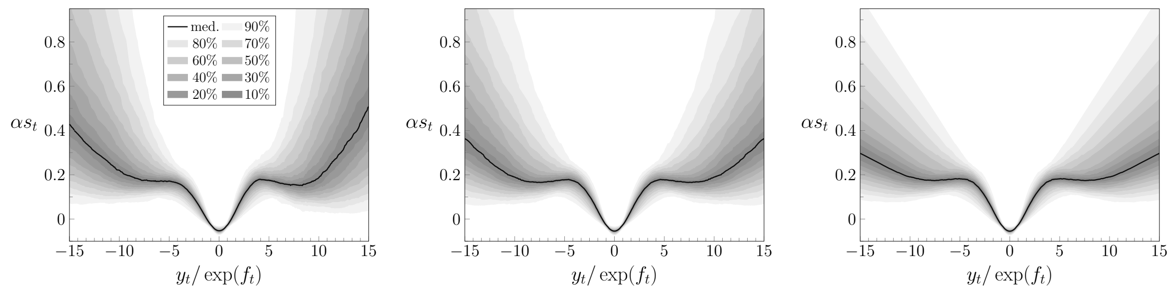
'slope 0' model has a better log likelihood value than the Student's t model, the difference in log likelihood is more than twice as high than in the remaining cases. Similarly, the difference in AIC is substantially larger if the 'slope 0' model has the lowest AIC than if the Student's t model has the lowest AIC. This demonstrates that on average, whenever the spline model has a better performance than the Student's t model in terms of log likelihood or AIC, it is considerably better, while if it is worse, it is only slightly worse.

On the other hand, for the BIC, these differences are almost identical.

To provide some insight into the forms of the estimated spline functions for the 456 companies under consideration, Figure 3 shows the median and quantiles of the estimated spline functions for the corresponding score functions, for each of the three spline models over a grid of points. It must be noted that the quantile plots do not represent actual spline and score functions of one of the companies, as the ordering of these functions per company can be different per input-value for which the quantiles are calculated. Therefore, we should be careful when interpreting these plots. The asymmetric model has a median and quantiles that are not symmetric in the origin, as the slope of the right tail tends to be slightly higher than that of the left tail. Stock returns are known to often have a negatively skewed distribution, which is reflected by the quantile plots, as the slope of the spline on the right tail is higher than the left tail, which means a thinner right tail alluding to negative skewness. The plots also show that wider spreads of the splines and corresponding score functions correspond to more flexible models. The plots appear to indicate that for the symmetric model, the outer slopes are close to zero in the majority of the cases. It is therefore not surprising that the 'slope 0' model is in many cases preferred by the AIC. The median score functions seem to indicate that many of



(a) Asymmetric spline **(b)** Symmetric spline **(c)** Slope 0 spline



(d) Asymmetric score **(e)** Symmetric score **(f)** Slope 0 score

Figure 3. Median spline values and scaled score values for each fitted spline model of 456 S&P 500 companies. Percentile regions are indicated in different shades of grey.

the score functions are outlier-robust, as their slopes are higher when the input value is close to zero relative to being far away from zero. The 'slope 0' models lead to the highest robustness, due to the linear divergence instead of quadratic divergence of the score function.

6. Conclusion

We have introduced a novel semiparametric score-driven model that relies on a spline-based density. The framework is presented in a general form, but particular attention is given to dynamic location and dynamic scale models. For a given vector of knots, parameter estimation is carried out by the method of maximum likelihood and consistency of the resulting estimator for location and scale models is established. For the location model, asymptotic normality of this estimator is also proven. The application of the location model to a monthly time series of U.S. inflation shows the ability of the model to deliver a robust filter. The more restricted model is most preferred for this data according to the information criteria. The application of the scale model to the 456 daily return series from the S&P 500 stock index leads to similar conclusions and demonstrates that the score-driven model with a spline-based density is widely applicable. Various extensions to the presented framework can be considered. First, an obvious extension is to consider multivariate models such as in Creal et al. (2014), although the numerical integration needed for the normalization constant of the density may become computationally challenging when the model dimension increases. Second, it is possible to explore the use of a bounded or double-bounded support for the density. Third, an interesting extenstion is to include multiple time-varying parameters in the model. Fourth, a regularization term to the objective function can be added in order to enforce smoothness of the spline, possibly by using a similar approach as the one of Gu and Qiu (1993) or by using P-splines as in Eilers and Marx (1996). The last extension enables the use of more parsimonious models based on a higher number of knots.

Appendix

A. Proofs of main results

Proof of Proposition 1. First we establish the stochastic properties of the true time-varying parameter process $\{f_t\}_{t\in\mathbb{Z}}$. Taking the approach of Blasques et al. (2022), define the process $\{\hat{f}_t^{\varepsilon}(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ using the following updating equation:

$$\hat{f}_{t+1}^{\varepsilon}(\boldsymbol{\theta}) = \omega(1-\beta) + \beta \hat{f}_t^{\varepsilon}(\boldsymbol{\theta}) + \alpha s(\hat{f}_t^{\varepsilon}(\boldsymbol{\theta}), g(\hat{f}_t^{\varepsilon}(\boldsymbol{\theta}), \varepsilon_t); \boldsymbol{\gamma}),$$

for $t = 1, 2, \ldots$ and for some initial value $\hat{f}_1^{\varepsilon} \in \mathcal{F}$ and where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is iid with $\varepsilon_t \sim p_{\varepsilon}(\varepsilon_t)$. We give $\hat{f}_t^{\varepsilon}$ the superscript ε to indicate that it is not a process of filtered values based on observations, but a data generating process based on iid innovations ε_t.

It follows from the form of g that $s(f, g(f, \varepsilon_t)) = h(\varepsilon_t)/\sigma$. We can directly apply Proposition 3.1 of Blasques et al. (2022). Condition (iii) of that proposition is satisfied, because for any $\boldsymbol{\theta} \in \Theta$, $\hat{f}_1 \in \mathbb{R}$ and any $n > 0$, $\mathbb{E}|s(\hat{f}_1, g(\hat{f}_1, \varepsilon_t); \boldsymbol{\gamma})|^n < \infty$, as h is a piecewise polynomial which is linear for values beyond a certain threshold, and ε_t has bounded moments of any order. Condition (iv) is also satisfied, as $\partial s(f, g(f, \varepsilon_t))/\partial f = 0$, such that the condition simplifies to $|\beta| < 1$ for any $\boldsymbol{\theta} \in \Theta$. It then follows that for any $\boldsymbol{\theta} \in \Theta$, the sequence $\{\hat{f}_t^{\varepsilon}(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges to a unique SE sequence $\{f_t^{\varepsilon}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ with $\mathbb{E}|f_t^{\varepsilon}(\boldsymbol{\theta})|^n < \infty$ for any $n > 0$. As $\boldsymbol{\theta}_0 \in \Theta$, it follows that the true $f_t = f_t^{\varepsilon}(\boldsymbol{\theta}_0)$ a.s., such that $\{f_t\}_{t\in\mathbb{Z}}$ is also SE over time with bounded moments of any order.

As the function $g(f, \varepsilon) = f + \sigma\varepsilon$ is measurable in f and ε, and $\{f_t\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are SE, it is follows from Proposition 4.3 of Krengel (1985) that $\{x_t\}_{t\in\mathbb{Z}}$ for $x_t = g(f_t^{\varepsilon}(\boldsymbol{\theta}_0), \varepsilon_t)$ is SE. As ε_t and $f_t^{\varepsilon}(\boldsymbol{\theta}_0)$ have bounded moments of any order, it is clear that this also counts for x_t, as it is the sum of these two variables.

□

Proof of Proposition 2. We can use that $s(f, x) = h((x-f)/\sigma)/\sigma$. The first result of the proposition holds because conditions (iii) and (iv) of Proposition 3.2 of Blasques et al. (2022) hold. Condition (iii) of that proposition holds, because for any $\hat{f}_1$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h((x_t - \hat{f}_1)/\sigma)/\sigma|^n < \infty$ for any $n > 0$, as h is a piecewise third degree polynomial that diverges linearly, x_t has bounded moments of any order by Proposition 1 and Θ is compact and is such that $\sigma > 0$ for any $\boldsymbol{\theta} \in \Theta$. Condition (iv) of that proposition holds due to Assumption 3, because $\partial s(f, x)/\partial f = -h'((x-f)/\sigma)/\sigma^2$, and clearly $\sup_{(f,x,\boldsymbol{\theta})\in\mathbb{R}\times\mathbb{R}\times\Theta}|\beta - \alpha h'((x-f)/\sigma)/\sigma^2| = \sup_{(z,\boldsymbol{\theta})\in\mathbb{R}\times\Theta}|\beta - \alpha h'(z)/\sigma^2|$.

Result (ii), follows from an application of the first part of Proposition 3.4 of Blasques et al. (2022). In particular, this proposition requires in addition to the conditions that were already in place, bounded moments uniformly over Θ of various quantities. Lemma B.1 shows that the required moment bounds hold for any $n > 0$, such that result (ii) of our proposition holds.

For result (iii), the second part of Proposition 3.4 of Blasques et al. (2022) cannot be directly applied, because the cubic spline function h is only twice continuously differentiable, which means that $\partial^3 s(f,x)/\partial f^3$, $\partial^3 s(f,x)/\partial f \partial \sigma^2$ and $\partial^3 s(f,x)/\partial f^2 \partial \sigma$ are not everywhere well defined. The moment bounds of these derivatives are used in the proof of Proposition 3.4 of Blasques et al. (2022) to show that $\partial^2 s(\hat{f}_t(\boldsymbol{\theta}), x_t)/\partial f \partial f$, $\partial^2 s(\hat{f}_t(\boldsymbol{\theta}), x_t)/\partial f \partial \sigma$ and $\partial^2 s(\hat{f}_t(\boldsymbol{\theta}), x_t)/\partial \sigma \partial \sigma$ converge e.a.s. to these derivatives evaluated in the limit filter $f_t(\boldsymbol{\theta})$, uniformly over Θ. Lemma B.3 shows that these convergence results still hold in this setting. The other relevant derivatives exist and they have bounded moments of any order, as shown in Lemma B.2. Thus the result of Proposition 3.4 of Blasques et al. (2022) holds for any $n > 0$, which finishes the proof. □

Proof of Theorem 1. Firstly, the MLE $\widehat{\boldsymbol{\theta}}_T$ exists, because the log likelihood function is almost surely continuous in $\boldsymbol{\theta} \in \Theta$ and Θ is compact, such that it follows from Weierstrass' theorem that a maximizer of the log likelihood function, i.e. the MLE, exists. The log likelihood function is a.s. continuous because $H(\cdot; \boldsymbol{\psi})$ is linear in $\boldsymbol{y}$, g^{-1} is continuous in σ and $\log(C(\boldsymbol{\psi}))$ is continuous in $\boldsymbol{y}$. The latter can be seen, because the expressions of $\int_{-\infty}^{t_1} \exp(-H(x))\, \mathrm{d}x$ and $\int_{t_k}^{\infty} \exp(-H(x))\, \mathrm{d}x$ derived in Section E.1 in the Supplementary Appendix are clearly continuous functions of $\boldsymbol{y}$. Furthermore, as $\exp(-H(x; \boldsymbol{\psi}))$ is continuous in $(x, \boldsymbol{\psi})$ over the compact set $[t_1, t_k] \times \boldsymbol{\Psi}$, implying it is uniformly continuous, it follows that $\int_{t_1}^{t_k} \exp(-H(x; \boldsymbol{\psi}))\, \mathrm{d}x$ is also continuous in $\boldsymbol{y}$. Furthermore, the score function $s(f, x; \boldsymbol{\gamma})$ is also clearly continuous in all arguments, which implies the filtered values $\hat{f}_t(\boldsymbol{\theta})$ are almost surely continuous in $\boldsymbol{\theta}$.

Due to the compactness of the parameter space Θ, consistency can be shown by verifying the conditions below, as can be shown by standard arguments, see for instance Wald (1949).

(C1) $\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})] \right| \overset{a.s.}{\to} 0$ as $T \to \infty$, where $\ell_t(\boldsymbol{\theta}) = \log p_x(x_t | f_t(\boldsymbol{\theta}); \boldsymbol{\gamma})$.

(C2) The true parameter $\boldsymbol{\theta}_0$ is the unique maximizer of the limit criterion: $\mathbb{E}[\ell_t(\boldsymbol{\theta})] < \mathbb{E}[\ell_t(\boldsymbol{\theta}_0)]$ for any $\boldsymbol{\theta} \in \Theta, \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

To show that condition **(C1)** holds, it is helpful to notice that by the triangle inequality:

$$\left|\frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t(\boldsymbol{\theta})-\mathbb{E}[\ell_t(\boldsymbol{\theta})]\right|\leq\left|\frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t(\boldsymbol{\theta})-\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta})\right|+\left|\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta})-\mathbb{E}[\ell_t(\boldsymbol{\theta})]\right|.$$

It is shown in Lemmas B.4 and B.5 that both terms on the right hand side converge to zero almost surely as $T\to\infty$ uniformly over $\boldsymbol{\theta}\in\Theta$. Finally, Lemma B.6 shows that condition **(C2)** holds. □

Proof of Theorem 2. Let us follow the approach of the proof of Theorem 3.1 of Gorgi and Koopman (2021), in which first the asymptotic normality of

$$\tilde{\boldsymbol{\theta}}_T=\arg\max_{\boldsymbol{\theta}\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}),$$

for $\ell_t(\boldsymbol{\theta})=\log p_x(x_t|f_t(\boldsymbol{\theta});\boldsymbol{\gamma})$ is shown, and then the final result follows by showing that $\tilde{\boldsymbol{\theta}}_T$ and $\widehat{\boldsymbol{\theta}}_T$ have the same limiting distribution.

We start by noting that $\ell_t(\boldsymbol{\theta})=\log p_x(x_t|f_t(\boldsymbol{\theta});\boldsymbol{\gamma})=-H((x_t-f_t(\boldsymbol{\theta}))/\sigma;\boldsymbol{\psi})-\log C(\boldsymbol{\psi})-0.5\log\sigma^2$ is a.s. twice continuously differentiable in $\boldsymbol{\theta}$. This is the case because the function $H(\cdot;\boldsymbol{\psi})$ is three times continuously differentiable, as h is a cubic spline. Furthermore, it is linear in $\boldsymbol{y}$, and therefore it is clearly differentiable in $\boldsymbol{y}$. That the first and second order derivatives of the limit filter $f_t(\boldsymbol{\theta})$ exist, was shown in Proposition 2. Finally, the normalization constant $C(\boldsymbol{\psi})$ is twice continuously differentiable, as the integrals beyond the outer knots given in Section E.1 in the Supplementary Appendix are clearly twice continuously differentiable in $\boldsymbol{y}$, and the remaining part is an integral of the function $\exp(-H(z;\boldsymbol{\psi}))$ on a compact domain, where the integrand is continuous in $z\in[t_1,t_k]$ and twice continuously differentiable in $\boldsymbol{y}$. Thus, the integral and derivative can be swapped by Leibniz rule.

The derivatives of $\ell_t(\boldsymbol{\theta})$ are based on the observations $\{x_t\}$ and the limit processes $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{\partial f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}_{t\in\mathbb{Z}}$ and $\{\partial^2 f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\}_{t\in\mathbb{Z}}$, which are SE by Proposition 1 and 2. Thus, as the derivatives of ℓ_t are continuous in these processes, they are themselves SE by Krengel (1985, Proposition 4.3). From the proof of Theorem 1, it follows that $\tilde{\boldsymbol{\theta}}_T\overset{a.s.}{\to}\boldsymbol{\theta}_0$ where we assume now that $\boldsymbol{\theta}_0$ lies in the interior of Θ. Then the following conditions can be shown to imply $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\overset{d}{\to}\mathcal{N}(0,-\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)]^{-1})$ as $T\to\infty$:

(A1) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\ell_t''(\boldsymbol{\theta})\|<\infty$,

(A2) $-\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)]$ is positive definite,

(A3) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, -\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)]^{-1})$ as $T \to \infty$,

Lemmas B.7, B.8 and B.9 show that these three conditions hold, respectively. To show that $\tilde{\boldsymbol{\theta}}_T$ and $\widehat{\boldsymbol{\theta}}_T$ have the same asymptotic distribution, it suffices to show that

$$\sqrt{T}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}\ell_t'(\boldsymbol{\theta}) - \frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t'(\boldsymbol{\theta})\right\| \xrightarrow{a.s.} 0 \qquad \text{as} \quad T\to\infty\,,$$

which holds by Lemma B.10. □

Proof of Proposition 3. We use the same approach as for Propostion 1. The first result follows from an application of Proposition 3.1 of Blasques et al. (2022). We have that $s(f^\varepsilon, g(f^\varepsilon, \varepsilon_t); \boldsymbol{\gamma}) = h(\varepsilon_t)\varepsilon_t - 1$. Thus, condition (iii) of that proposition holds for any $n > 0$, as ε_t has bounded moments of any order and h is a piecewise third degree polynomial. Furthermore, as $\partial s(f^\varepsilon, g(f^\varepsilon, \varepsilon_t); \boldsymbol{\gamma})/\partial f^\varepsilon = 0$, condition (iv) of that proposition simplifies to $|\beta| < 1$, which holds for any $\boldsymbol{\theta} \in \Theta$ as $0 < \beta < 1$, which is imposed in Assumption 5. Thus, it follows from Proposition 3.1 of Blasques et al. (2022) that for any $\hat{f}_1^\varepsilon$ and for any $\boldsymbol{\theta} \in \Theta$, $\{\hat{f}_t^\varepsilon(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE limit sequence $\{f_t^\varepsilon(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ where the elements $f_t^\varepsilon(\boldsymbol{\theta})$ have bounded moments of any order. As $\boldsymbol{\theta}_0 \in \Theta$, it thus follows that the true sequence of time-varying parameters $\{f_t\}_{t\in\mathbb{Z}}$ is SE with bounded moments of any order. Finally, $\{x_t\}_{t\in\mathbb{Z}}$ is SE by the same argument as in Proposition 1. □

Proof of Proposition 4. The convergence result follows from an application of the first part of Proposition 3.2 of Blasques et al. (2022). It follows from Proposition 3 that $\{x_t\}_{t\in\mathbb{Z}}$ is an SE sequence. We have that $s(f, x; \boldsymbol{\gamma}) = h((x-m)\exp(-f))(x-m)\exp(-f) - 1$, so for any $\hat{f}_1 \in \mathcal{F}$, we have $\mathbb{E}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\log^+|s(\hat{f}_1, x_t; \boldsymbol{\gamma})| < \infty$ as h is a piecewise cubic spline, $\boldsymbol{\Gamma}$ is compact and x_t has bounded moments of any order by Proposition 3. Thus, condition (i) of Proposition 3.2 of Blasques et al. (2022) is satisfied. Because

$$\begin{aligned}\frac{\partial s(f, x; \boldsymbol{\gamma})}{\partial f} &= -h'((x-m)\exp(-f))(x-m)^2\exp(-2f)\\ &\quad - h((x-m)\exp(-f))(x-m)\exp(-f)\,,\end{aligned}$$

the condition in Assumption 6 ensures that condition (ii) of that proposition also holds. Thus, indeed the sequence $\{\hat{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE limit sequence $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ as $t\to\infty$ uniformly over Θ.

□

Proof of Theorem 3. The existence result holds by the same argumentation as for Theorem 1.

Using the proof of Theorem 4.1 of Blasques et al. (2018), which is similar to Theorem 4.1 of Straumann and Mikosch (2006), we know that the following three conditions are sufficient for showing consistency.

(C1) $\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t(\boldsymbol{\theta})-\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta})\right| \overset{a.s.}{\to} 0$ as $T\to\infty$, where we have $\ell_t(\boldsymbol{\theta}) = \log p_x(x_t|f_t(\boldsymbol{\theta});\boldsymbol{\gamma})$.

(C2) $\sup_{(x,f,\boldsymbol{\theta})\in\mathbb{R}\times\mathcal{F}\times\Theta}\log p_x(x|f;\boldsymbol{\gamma})<\infty$ and $\mathbb{E}|\ell_1(\boldsymbol{\theta}_0)|<\infty$,

(C3) The true parameter $\boldsymbol{\theta}_0$ is the unique maximizer of the limit criterion: $\mathbb{E}[\ell_t(\boldsymbol{\theta})] < \mathbb{E}[\ell_t(\boldsymbol{\theta}_0)]$ for any $\boldsymbol{\theta}\in\Theta, \boldsymbol{\theta}\neq\boldsymbol{\theta}_0$.

Conditions **(C1)**, **(C2)** and **(C3)** hold by Lemmas B.11, B.12 and B.13, respectively.

□

B. Lemmas

The proofs of the Lemmas below are given in Section C of the Supplementary Appendix.

Lemma B.1. *Let the conditions of Proposition 2 be satisfied. Then for any $n>0$:*

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|s(f_t(\boldsymbol{\theta}),x_t)|^n<\infty\,,\qquad \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial s(f_t(\boldsymbol{\theta}),x_t)}{\partial\boldsymbol{\psi}}\right\|^n<\infty\,,$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial s(f_t(\boldsymbol{\theta}),x_t)}{\partial\sigma^2}\right|^n<\infty\,,\qquad \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left|\frac{\partial s(f,x_t)}{\partial f}\right|^n<\infty\,,$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left|\frac{\partial^2 s(f,x_t)}{\partial f\partial f}\right|^n<\infty\,,\qquad \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left\|\frac{\partial^2 s(f,x_t)}{\partial f\partial\boldsymbol{\psi}}\right\|^n<\infty\,,$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left|\frac{\partial^2 s(f,x_t)}{\partial f\partial\sigma^2}\right|^n<\infty\,.$$

Lemma B.2. *Let the conditions of Proposition 2 be satisfied. Then for any $n>0$:*

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 s(f_t(\boldsymbol{\theta}),x_t)}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top}\right\|^n<\infty\,,\qquad \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^2 s(f_t(\boldsymbol{\theta}),x_t)}{\partial\sigma^2\partial\sigma^2}\right|^n<\infty\,,$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 s(f_t(\boldsymbol{\theta}),x_t)}{\partial\boldsymbol{\psi}\partial\sigma^2}\right\|^n<\infty\,,\qquad \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left\|\frac{\partial^3 s(f,x_t)}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\partial f}\right\|^n<\infty\,,$$
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left\|\frac{\partial^3 s(f,x_t)}{\partial\boldsymbol{\psi}\partial\sigma^2\partial f}\right\|^n<\infty\,,\qquad \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f\in\mathbb{R}}\left\|\frac{\partial^3 s(f,x_t)}{\partial\boldsymbol{\psi}\partial f\partial f}\right\|^n<\infty\,.$$

Lemma B.3. *Let the conditions of Proposition 2 be satisfied. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^2 s(f,x_t)}{\partial f \partial f}\bigg|_{f=\hat{f}_t(\boldsymbol{\theta})} - \frac{\partial^2 s(f,x_t)}{\partial f \partial f}\bigg|_{f=f_t(\boldsymbol{\theta})}\right| \overset{e.a.s.}{\to} 0 \qquad \text{as } t\to\infty\,,$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^2 s(f,x_t)}{\partial \sigma^2 \partial f}\bigg|_{f=\hat{f}_t(\boldsymbol{\theta})} - \frac{\partial^2 s(f,x_t)}{\partial \sigma^2 \partial f}\bigg|_{f=f_t(\boldsymbol{\theta})}\right| \overset{e.a.s.}{\to} 0 \qquad \text{as } t\to\infty\,,$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^2 s(\hat{f}_t(\boldsymbol{\theta}),x_t)}{\partial \sigma^2 \partial \sigma^2} - \frac{\partial^2 s(f_t(\boldsymbol{\theta}),x_t)}{\partial \sigma^2 \partial \sigma^2}\right| \overset{e.a.s.}{\to} 0 \qquad \text{as } t\to\infty\,.$$

Lemma B.4. *Let the conditions of Theorem 1 hold. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t(\boldsymbol{\theta}) - \frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta})\right| \overset{a.s.}{\to} 0 \qquad \text{as } T\to\infty\,.$$

Lemma B.5. *Let the conditions of Theorem 1 hold. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}) - \mathbb{E}[\ell_t(\boldsymbol{\theta})]\right| \overset{a.s.}{\to} 0 \qquad \text{as } T\to\infty\,.$$

Lemma B.6. *Let the conditions of Theorem 1 hold. Then the true parameter $\boldsymbol{\theta}_0$ is the unique maximizer of the limit criterion:*

$$\mathbb{E}[\ell_t(\boldsymbol{\theta})] < \mathbb{E}[\ell_t(\boldsymbol{\theta}_0)] \quad \text{for any } \boldsymbol{\theta}\in\Theta, \boldsymbol{\theta}\neq\boldsymbol{\theta}_0\,.$$

Lemma B.7. *Let the conditions of Theorem 2 hold. Then*

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\ell_t''(\boldsymbol{\theta})\| < \infty\,.$$

Lemma B.8. *Let the conditions of Theorem 2 hold. Then $-\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)]$ is positive definite.*

Lemma B.9. *Let the conditions of Theorem 2 hold. Then*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}_0) \overset{d}{\to} \mathcal{N}(0, -\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)]^{-1}) \qquad \text{as } T\to\infty\,.$$

Lemma B.10. *Let the conditions of Theorem 2 hold. Then*

$$\sqrt{T}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{T}\sum_{t=1}^{T}\ell_t'(\boldsymbol{\theta}) - \frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t'(\boldsymbol{\theta})\right\| \overset{a.s.}{\to} 0 \qquad \text{as } T\to\infty\,.$$

Lemma B.11. *Let the conditions of Theorem 3 hold. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t(\boldsymbol{\theta})-\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\theta})\right| \overset{a.s.}{\to} 0 \qquad \textit{as } T\to\infty\,.$$

Lemma B.12. *Let the conditions of Theorem 3 hold. Then* $\sup_{(x,f,\boldsymbol{\theta})\in\mathbb{R}\times\mathcal{F}\times\Theta}\log p_x(x|f;\boldsymbol{\gamma})<\infty$ *and* $\mathbb{E}|\ell_1(\boldsymbol{\theta}_0)|<\infty$ *.*

Lemma B.13. *Let the conditions of Theorem 3 hold. Then the true parameter* $\boldsymbol{\theta}_0$ *is the unique maximizer of the limit criterion:*

$$\mathbb{E}[\ell_t(\boldsymbol{\theta})]<\mathbb{E}[\ell_t(\boldsymbol{\theta}_0)] \quad \textit{for any } \boldsymbol{\theta}\in\Theta, \boldsymbol{\theta}\neq\boldsymbol{\theta}_0\,.$$

References

Abdelkarim, R. A. and Onour, I. A. (2023). Semiparametric score-driven exponentially weighted moving average model. In *The International Conference On Global Economic Revolutions*, pages 210–221. Springer.

Abrahamowicz, M., Clampl, A., and Ramsay, J. O. (1992). Nonparametric density estimation for censored survival data: regression-spline approach. *Canadian Journal of Statistics*, 20(2):171–185.

Alberg, J., Nilson, E., and Walsh, J. (1967). The theory of splines and their applications. *New York: Academic Press.*

Artemova, M., Blasques, F., van Brummelen, J., and Koopman, S. J. (2022a). Score-driven models: Methodology and Theory. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Artemova, M., Blasques, F., van Brummelen, J., and Koopman, S. J. (2022b). Score-Driven Models: Methods and Applications. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Billingsley, P. (1999). *Convergence of Probability Measures.* John Wiley & Sons, 2nd edition.

Blasques, F., Gorgi, P., Koopman, S. J., and Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, 12(1):1019–1052.

Blasques, F., Ji, J., and Lucas, A. (2016). Semiparametric score driven volatility models. *Computational Statistics & Data Analysis*, 100:58–69.

Blasques, F., van Brummelen, J., Gorgi, P., and Koopman, S. J. (2024). Maximum likelihood estimation for non-stationary location models with mixture of normal distributions. *Journal of Econometrics*, 238(1):105575.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Catania, L. and Luati, A. (2023). Semiparametric modeling of multiple quantiles. *Journal of Econometrics*, 237(2):105365.

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

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

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

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

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127–1162.

Gorgi, P. and Koopman, S. J. (2021). Beta observation-driven models with exogenous regressors: a joint analysis of realized correlation and leverage effects. *Journal of Econometrics*. forthcoming.

Gu, C. (1993). Smoothing spline density estimation: A dimensionless automatic algorithm. *Journal of the American Statistical Association*, 88(422):495–504.

Gu, C. and Qiu, C. (1993). Smoothing spline density estimation: Theory. *The Annals of Statistics*, 21(1):217–234.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Cambridge University Press.

Harvey, A. C. and Chakravarty, T. (2008). Beta-t-(E)GARCH. *University of Cambridge, Faculty of Economics, Working paper CWPE 08340.*

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

Kooperberg, C. and Stone, C. J. (1991). A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12(3):327–347.

Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1(4):301–328.

Krengel, U. (1985). *Ergodic theorems*, volume 6 of *De Gruyter Studies in Mathematics*. de Gruyter, Berlin.

Nelson, D. B. and Foster, D. P. (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica*, 62:1–41.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and Value-at-Risk). *Journal of Econometrics*, 211(2):388–413.

Poirier, D. J. (1973). Piecewise regression using cubic splines. *Journal of the American Statistical Association*, 68(343):515–524.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33(5):659–680.

Russell, J. R. (2001). Econometric modeling of multivariate irregularly-spaced high-frequency data. Technical report, University of Chicago.

Stone, C. J. (1990). Large-sample inference for log-spline models. *The Annals of Statistics*, 18(2):717–741.

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

Takeuchi, K. (1976). The distribution of information statistics and the criterion of goodness of fit of models. *Mathematical Science*, 153:12–18.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601.

Supplementary Appendix of

Score-driven time-varying parameter models with spline-based densities

C. Proofs of lemmas

Proof of Lemma B.1. See Section D.1 for the expressions of the derivatives.

That for any $n > 0$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |s(f_t(\boldsymbol{\theta}), x_t)|^n < \infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\partial s(f_t(\boldsymbol{\theta}), x_t)/\partial\boldsymbol{\psi}\|^n < \infty$, and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\partial s(f_t(\boldsymbol{\theta}), x_t)/\partial\sigma^2|^n < \infty$, follows from the assumptions on Θ and the fact that $s(f_t(\boldsymbol{\theta}), x_t)$, $\partial s(f_t(\boldsymbol{\theta}), x_t)/\partial\boldsymbol{\psi}$ and $\partial s(f_t(\boldsymbol{\theta}), x_t)/\partial\sigma^2$ are all at most third order piecewise polynomials evaluated in $x_t - f_t(\boldsymbol{\theta})$, and x_t and $f_t(\boldsymbol{\theta})$ are known to have bounded moments of any order by Propositions 1 and 2(i). In particular, due to the compactness of Θ, the conditions on $\boldsymbol{\psi}$ that ensure the spline-based density is well-defined and $\sigma > 0$ for any $\boldsymbol{\theta} \in \Theta$, the expectation of the supremum over Θ of these quantities will be finite.

For the remaining expectations, so those with a supremum over $f \in \mathbb{R}$, it can be argued that the derivatives are uniformly bounded over $\boldsymbol{\theta} \in \Theta$, $f \in \mathbb{R}$ and $x \in \mathbb{R}$. In particular, it is clear that $\sup_{z\in\mathbb{R},\boldsymbol{\psi}\in\boldsymbol{\Psi}} |h'(z;\boldsymbol{\psi})| < D$ for some constant $D < \infty$. Furthermore, the second order derivative of h is not only such that $\sup_{z\in\mathbb{R},\boldsymbol{\psi}\in\boldsymbol{\Psi}} |h''(z;\boldsymbol{\psi})| < D$ for some $D < \infty$, but also has value zero outside the outer knots, due to h being linear outside the outer knots. It follows that also $\sup_{z\in\mathbb{R},\boldsymbol{\psi}\in\boldsymbol{\Psi}} |zh''(z;\boldsymbol{\psi})| < D$ for some $D < \infty$. By again taking into account the assumptions on Θ, it is clear that the derivatives are indeed uniformly bounded over f and $\boldsymbol{\theta}$, and therefore have bounded moments of any order. □

Proof of Lemma B.2. See Section D.1 for the expressions of the derivatives.

For the expectations without a supremum over $f \in \mathbb{R}$ and the expectations with a supremum over $f \in \mathbb{R}$, exactly the same two respective argumentations can be used as in the proof of Lemma B.1. Here we also need to use that $\sup_{z\in\mathbb{R},\boldsymbol{\psi}\in\boldsymbol{\Psi}} |z^2h''(z;\boldsymbol{\psi})| < D$ for some constant $D < \infty$, which holds because $h''(z;\boldsymbol{\psi}) = 0$ whenever $z < t_1$ or $z > t_k$. □

Proof of Lemma B.3. The derivative

$$\frac{\partial^2 s(f, x_t)}{\partial f \partial f} = \frac{1}{\sigma^3} h''\left(\frac{x - f}{\sigma}; \boldsymbol{\psi}\right),$$

is a continuous piecewise linear function of $x - f$. Due to the compactness of Θ and the other conditions in place, it is clear that the function h'' is Lipschitz continuous uniformly over $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, i.e. there exists a constant K such that $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |h''(z_1;\boldsymbol{\psi}) - h''(z_2;\boldsymbol{\psi})| < K|z_1 - z_2|$ for any $z_1, z_2 \in \mathbb{R}$. Therefore,

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta} & \left| \frac{\partial^2 s(f,x_t)}{\partial f \partial f}\bigg|_{f=\hat{f}_t(\boldsymbol{\theta})} - \frac{\partial^2 s(f,x_t)}{\partial f \partial f}\bigg|_{f=f_t(\boldsymbol{\theta})} \right| \\
&= \sup_{\boldsymbol{\theta}\in\Theta} \left| \frac{1}{\sigma^3} h''\left(\frac{x_t - \hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right) - \frac{1}{\sigma^3} h''\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right) \right| \\
&\leq \sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{\sigma^3} K \sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{\sigma} \sup_{\boldsymbol{\theta}\in\Theta} \left| f_t(\boldsymbol{\theta}) - \hat{f}_t(\boldsymbol{\theta}) \right| \overset{e.a.s}{\to} 0\,,
\end{aligned}$$

as $t \to \infty$, because $\sigma > 0$ for any $\boldsymbol{\theta} \in \Theta$, Θ is compact and $\hat{f}_t(\boldsymbol{\theta})$ converges e.a.s. to $f_t(\boldsymbol{\theta})$ uniformly over Θ by Proposition 2(i).

For the second derivative under consideration,

$$\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial \sigma^2 \partial f} = \frac{1}{\sigma^4} h'\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right) - \frac{x-f}{\sigma^5} h''\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right),$$

we will show the e.a.s. convergence of the two terms separately, which implies convergence of the total expression by the subadditivity of the sup-norm. The first term is continuously differentiable in f, so an application of the mean value theorem gives:

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta} & \left| \frac{1}{\sigma^4} h'\left(\frac{x_t - \hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right) - \frac{1}{\sigma^4} h'\left(\frac{x_t - \hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right) \right| \\
&\leq \sup_{f\in\mathbb{R},\boldsymbol{\theta}\in\Theta} \left| \frac{1}{\sigma^5} h''\left(\frac{x_t - f}{\sigma};\boldsymbol{\psi}\right) \right| |\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})| \\
&\leq D|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0\,,
\end{aligned}$$

as $t \to \infty$, where $D = \sup_{\boldsymbol{\theta},z} |h''(z;\boldsymbol{\psi})/\sigma^5|$, which is finite because h'' is continuous and is zero for any $z < t_1$ and $z > t_k$ by construction and Θ is compact with $\sigma > 0$.

For the remaining term, let $z_t(\boldsymbol{\theta}) = (x_t - f_t(\boldsymbol{\theta}))/\sigma$ and $\hat{z}_t(\boldsymbol{\theta}) = (x_t - \hat{f}_t(\boldsymbol{\theta}))/\sigma$. Then,

$$\sup_{\boldsymbol{\theta}\in\Theta} \left| \frac{\hat{z}_t(\boldsymbol{\theta})}{\sigma^4} h''\left(\hat{z}_t(\boldsymbol{\theta});\boldsymbol{\psi}\right) - \frac{z_t(\boldsymbol{\theta})}{\sigma^4} h''\left(z_t(\boldsymbol{\theta});\boldsymbol{\psi}\right) \right| \overset{e.a.s.}{\to} 0\,,$$

as $t \to \infty$ by Lemma TA.14 of Blasques et al. (2022), because clearly $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ and $\sup_{\boldsymbol{\theta}\in\Theta} |h''\left(\hat{z}_t(\boldsymbol{\theta});\boldsymbol{\psi}\right) - h''\left(z_t(\boldsymbol{\theta});\boldsymbol{\psi}\right)| \overset{e.a.s.}{\to} 0$, by the uniform Lipschitz continuity of h'', and $z_t(\boldsymbol{\theta})$ and $h''\left(z_t(\boldsymbol{\theta});\boldsymbol{\psi}\right)$ have a bounded $\log^+$-moment uniformly over $\boldsymbol{\theta} \in \Theta$.

Finally, given that

$$\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial \sigma^2 \partial \sigma^2} = \frac{3}{4\sigma^5} h\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right) + \frac{5(x-f)}{4\sigma^6} h'\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right),$$

the final convergence result follows from an application of the mean value theorem, the e.a.s. convergence of $\hat{f}_t(\boldsymbol{\theta})$ to $f_t(\boldsymbol{\theta})$ uniformly over Θ, and the observation that the derivative of the expression with respect to f is bounded:

$$\sup_{x\in\mathbb{R}, f\in\mathbb{R}, \boldsymbol{\theta}\in\Theta} \left| \frac{3}{4\sigma^6} h'\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right) + \frac{5(x-f)}{4\sigma^7} h''\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right) \right| = \tilde{D},$$

for some finite constant $\tilde{D}$, due to $h'(z)$ and $zh''(z)$ being bounded in z and Θ being compact. □

Proof of Lemma B.4. To prove the result holds, it is sufficient to show that $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{\ell}_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$ by Lemma 2.1 of Straumann and Mikosch (2006). The expression of $\ell_t(\boldsymbol{\theta})$ is given in (D.1). Only the first term, $-H((x_t - f_t(\boldsymbol{\theta}))/\sigma; \boldsymbol{\psi})$, contains $f_t(\boldsymbol{\theta})$, so we can disregard the other terms. By construction, $H(\cdot;\boldsymbol{\psi})$ is a continuously differentiable function. Thus, if we let $\hat{z}_t(\boldsymbol{\theta}) := (x_t - \hat{f}_t(\boldsymbol{\theta}))/\sigma$ and $z_t(\boldsymbol{\theta}) := (x_t - f_t(\boldsymbol{\theta}))/\sigma$ we can apply the mean value theorem to obtain:

$$|H(\hat{z}_t(\boldsymbol{\theta});\boldsymbol{\psi}) - H(z_t(\boldsymbol{\theta});\boldsymbol{\psi})| \le |h(z_t^*(\boldsymbol{\theta}))| \, |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| ,$$

for some processes $\{\hat{z}_t(\boldsymbol{\theta})\}$ and $\{z_t(\boldsymbol{\theta})\}$, and where $z_t^*(\boldsymbol{\theta})$ is on the line segment between $\hat{z}_t(\boldsymbol{\theta})$ and $z_t(\boldsymbol{\theta})$. As h is also continuously differentiable, we can apply the mean value theorem to the first factor on the right-hand side, which gives

$$\begin{aligned} |h(z_t^*(\boldsymbol{\theta});\boldsymbol{\psi})| &\le |h(z_t(\boldsymbol{\theta});\boldsymbol{\psi})| + |h'(z_t^{**}(\boldsymbol{\theta});\boldsymbol{\psi})| \, |z_t^*(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| \\ &\le |h(z_t(\boldsymbol{\theta});\boldsymbol{\psi})| + \sup_{z\in\mathbb{R}} |h'(z;\boldsymbol{\psi})| \, |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| , \end{aligned}$$

for some value $z_t^{**}(\boldsymbol{\theta})$ on the line segment between $z_t^*(\boldsymbol{\theta})$ and $z_t(\boldsymbol{\theta})$, and where we use for the second inequality that $z_t^*(\boldsymbol{\theta})$ is on the line segment between $z_t(\boldsymbol{\theta})$ and $\hat{z}_t(\boldsymbol{\theta})$. It therefore follows that:

$$\begin{aligned} &\sup_{\boldsymbol{\theta}\in\Theta} |H(\hat{z}_t(\boldsymbol{\theta});\boldsymbol{\psi}) - H(z_t(\boldsymbol{\theta});\boldsymbol{\psi})| \\ &\qquad \le \left(\sup_{\boldsymbol{\theta}\in\Theta} |h(z_t(\boldsymbol{\theta});\boldsymbol{\psi})| + \sup_{\boldsymbol{\theta}\in\Theta} \sup_{z\in\mathbb{R}} |h'(z;\boldsymbol{\psi})| \sup_{\boldsymbol{\theta}\in\Theta} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| \right) \sup_{\boldsymbol{\theta}\in\Theta} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| , \end{aligned}$$

which converges e.a.s. to zero as $t \to \infty$ as long as $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$ as $t \to \infty$ and $z_{tt\in\mathbb{Z}}$ is stationary with a bounded $\log^+$-moment. This follows because Θ is compact and $\sup_{\boldsymbol{\theta}\in\Theta} \sup_{z\in\mathbb{R}} |h'(z;\boldsymbol{\psi})| = D$ for some finite constant D due to the boundedness of h' for every $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. The result then follows from Lemma 2.1 of Straumann and

Mikosch (2006), which says that if $\xi_t \overset{e.a.s.}{\to} 0$ as $t \to \infty$ and $\{v_t\}$ is strictly stationary with $\mathbb{E}\log^+|v_t| < \infty$, then $v_t\xi_t \overset{e.a.s.}{\to} 0$ as $t \to \infty$.

Now it remains to be shown that $\{\hat{z}_t\}$ converges to $\{z_t\}$ e.a.s. and $\{z_t\}$ has the required properties. We have that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}|\hat{z}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})| &= \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{x_t - \hat{f}_t(\boldsymbol{\theta})}{\sigma} - \frac{x_t - f_t(\boldsymbol{\theta})}{\sigma}\right| \\ &\le \sup_{\boldsymbol{\theta}\in\Theta}\frac{1}{\sigma}\cdot\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})\right| \overset{e.a.s.}{\to} 0\,,\end{aligned}$$

as $t \to \infty$, due to the compactness of Θ and $\sigma^2 > 0$ for any $\boldsymbol{\theta} \in \Theta$ by assumptions, and the convergence result of Proposition 2(i). Finally, $\{z_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}} = \{(x_t - f_t(\boldsymbol{\theta}))/\sigma\}_{t\in\mathbb{Z}}$ is clearly SE by Krengel (1985, Proposition 4.3) and it has bounded moments of any order due to x_t and $f_t(\boldsymbol{\theta})$ having bounded moments of any order by Propositions 1 and 2. This concludes the proof.

□

Proof of Lemma B.5. See Equation (D.1) for an expression of ℓ_t. The elements $\ell_t(\cdot)$ are continuous functions from Θ to $\mathbb{R}$. Furthermore, $\{\ell_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE for every $\boldsymbol{\theta} \in \Theta$ by Krengel (1985, Proposition 4.3), as it is a continuous function of the elements of the SE sequences $\{x_t\}_{t\in\mathbb{Z}}$ and $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$.

Just as in Lemma TA.6 of Blasques et al. (2022), we apply the ergodic theorem for separable Banach spaces of Rao (1962) to $\{\ell_t(\cdot)\}_{t\in\mathbb{Z}}$, which requires us to show $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\ell_t(\boldsymbol{\theta})| < \infty$. As the last two terms of $\ell_t(\boldsymbol{\theta})$ are deterministic and time invariant, we only have to consider the first term here. We can see immediately that

$$E\sup_{\boldsymbol{\theta}\in\Theta}\left|H\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)\right| < \infty\,,$$

because *(i)* Θ is compact, *(ii)* $\sigma > 0$ for any $\boldsymbol{\theta} \in \Theta$, and *(iii)* $\mathbb{E}|x_t|^n < \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^n < \infty$ for any $n > 0$ by Propositions 1 and 2, and H is a continuous piecewise quartic polynomial in $x_t - f_t(\boldsymbol{\theta})$. Thus, we can apply the ergodic theorem and the result follows.

□

Proof of Lemma B.6. It follows from the proof of Lemma B.5 that $\mathbb{E}[\ell_t(\boldsymbol{\theta})]$ exists and is bounded for any $\boldsymbol{\theta} \in \Theta$. It can be shown straightforwardly, see for instance Blasques et al. (2018, Theorem 4.1), that under correct specification, in case $\ell_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}_0)$ holds almost surely if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, then $\boldsymbol{\theta}_0$ is the unique maximizer of $\mathbb{E}[\ell_t(\boldsymbol{\theta})]$.

For ease of exposition, we introduce the notation $\ell(f_t(\boldsymbol{\theta}), \boldsymbol{\gamma}) := \ell_t(\boldsymbol{\theta})$. We start by showing that $\ell(f_t(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0) = \ell(f_t(\boldsymbol{\theta}), \boldsymbol{\gamma})$ a.s. if and only if $f_t(\boldsymbol{\theta}_0) = f_t(\boldsymbol{\theta})$ a.s. and $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. Notice that $f_t(\boldsymbol{\theta}_0)$ is the true time-varying parameter, as we know from Proposition 2 that $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is the unique limit sequence of $\{\hat{f}_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{N}}$, which implies that $x_t = f_t(\boldsymbol{\theta}_0) + \sigma_0\varepsilon_t$.

First, we notice that because each vector $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ leads to a unique cubic spline function $h(\cdot;\boldsymbol{\psi})$, it also leads to a unique anti-derivative $H(\cdot;\boldsymbol{\psi})$. Despite to the anti-derivative H for a given $\boldsymbol{\psi}$ only being unique up to an additive constant, it is clear that for $a, a' \in \mathbb{R}$ and $\boldsymbol{\psi}, \boldsymbol{\psi}' \in \boldsymbol{\Psi}$, we have $a + H(z;\boldsymbol{\psi}) = a' + H(z;\boldsymbol{\psi}')$ for all $z \in \mathbb{R}$ if and only if $(a, \boldsymbol{\psi}) = (a', \boldsymbol{\psi}')$.

Building on this, we notice that for any $a, a', c, c' \in \mathbb{R}$, $b, b' > 0$, and $\boldsymbol{\psi}, \boldsymbol{\psi}' \in \boldsymbol{\Psi}$, $a + H((z-c)/b;\boldsymbol{\psi}) = a' + H((z-c')/b';\boldsymbol{\psi}')$ for any $z \in \mathbb{R}$ if and only if $(a, b, c, \boldsymbol{\psi}) = (a', b', c', \boldsymbol{\psi}')$. This follows because the division by b essentially leads to a scaling of the knots $\boldsymbol{t}$, while c effectively yields a shift of the knots, as it can be seen that $h((z-c)/b;\boldsymbol{t},\boldsymbol{y}) = h(z-c;b\boldsymbol{t},\boldsymbol{y}) = h(z;b\boldsymbol{t}+c,\boldsymbol{y})$. Due to the rotational symmetry around the origin that is imposed on the spline h, it is clear that a shift of the knots will always lead to an inherently different spline function, which implies that we must have $c = c'$ for the equality above to hold. Furthermore, from the second part of Assumption 4, it follows that $h((z-c)/b;\boldsymbol{\psi}) = h((z-c)/b';\boldsymbol{\psi}')$ for any $z \in \mathbb{R}$ if and only if $(b, \boldsymbol{\psi}) = (b, \boldsymbol{\psi}')$. This concludes the proof for the statement above.

Therefore, if $\varepsilon_1 \sim p_\varepsilon(\varepsilon_1;\boldsymbol{\psi})$, where $p_\varepsilon(\cdot;\boldsymbol{\psi})$ denotes the spline-based density for parameter $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, then $a' + H((\varepsilon_1 - c')/b';\boldsymbol{\psi}') = a'' + H((\varepsilon_1 - c'')/b'';\boldsymbol{\psi}'')$ a.s. if and only if $(a', b', c', \boldsymbol{\psi}') = (a'', b'', c'', \boldsymbol{\psi}'')$, as ε_1 is an absolutely continuous random variable with support on the entire real line. Thus, as $\ell(f_t(\boldsymbol{\theta}), \boldsymbol{\gamma}) = \ell(f_t(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0)$ a.s. if and only if

$$
\begin{aligned}
- H\left(\frac{f_t(\boldsymbol{\theta}_0) - f_t(\boldsymbol{\theta}) + \sigma_0\varepsilon_t}{\sigma};\boldsymbol{\psi}\right) - \log C(\boldsymbol{\psi}) - \frac{1}{2}\log\sigma^2 \\
= -H(\varepsilon_t;\boldsymbol{\psi}_0) - \log C(\boldsymbol{\psi}_0) - \frac{1}{2}\log\sigma_0^2\,,
\end{aligned}
$$

holds almost surely, it follows that $\ell(f_t(\boldsymbol{\theta}), \boldsymbol{\gamma}) = \ell(f_t(\boldsymbol{\theta}_0), \boldsymbol{\gamma}_0)$ a.s. if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ and $f_t(\boldsymbol{\theta}_0) = f_t(\boldsymbol{\theta})$ a.s.. To show that $\ell_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}_0)$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, it now only remains to be shown that given $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, $f_t(\boldsymbol{\theta}) = f_t(\boldsymbol{\theta}_0)$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Because $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE for any $\boldsymbol{\theta} \in \Theta$, then if $f_t(\boldsymbol{\theta}) = f_t(\boldsymbol{\theta}_0)$ a.s. for some $t \in \mathbb{Z}$, this equality holds almost surely for any $t \in \mathbb{Z}$. Let us assume that $\boldsymbol{\theta} \in \Theta$ is such that $f_t(\boldsymbol{\theta}) = f_t(\boldsymbol{\theta}_0)$

a.s., then it follows from the form of the updating equation that

$$f_{t+1}(\boldsymbol{\theta}_0) - f_{t+1}(\boldsymbol{\theta}) = \omega_0(1-\beta_0) - \omega(1-\beta) + (\beta_0 - \beta) f_t(\boldsymbol{\theta}_0) + (\alpha_0 - \alpha)\frac{1}{\sigma_0} h\left(\varepsilon_t; \boldsymbol{\psi}_0\right),$$

where $\varepsilon_t \sim p_\varepsilon(\varepsilon_t; \boldsymbol{\psi}_0)$. We start by proving that to have $f_t(\boldsymbol{\theta}_0) = f_t(\boldsymbol{\theta})$ a.s., we must have $\omega(1-\beta) = \omega_0(1-\beta_0)$. We use a proof by contradiction. Say $\omega(1-\beta) \neq \omega_0(1-\beta_0)$ and $f_{t+1}(\boldsymbol{\theta}_0) - f_{t+1}(\boldsymbol{\theta}) = 0$ a.s., then we must have $(\beta_0 - \beta) f_t(\boldsymbol{\theta}_0) + (\alpha_0 - \alpha)\frac{1}{\sigma_0} h\left(\varepsilon_t; \boldsymbol{\psi}_0\right) = \omega(1-\beta) - \omega_0(1-\beta_0) \neq 0$ a.s. As the first and second term of the expression on the left hand side are independent, this equality can only hold a.s. if both terms are constants and at least one is different from zero. However, neither of these terms can be nonzero constants, because $f_t(\boldsymbol{\theta}_0)$ is non-deterministic due to $\alpha_0 \neq 0$ by Assumption 4 and $h(\varepsilon_t; \boldsymbol{\psi}_0)$ is also clearly non-deterministic under the conditions imposed on $\boldsymbol{\Psi}$. Therefore, we must have $\omega(1-\beta) = \omega_0(1-\beta_0)$. Secondly, we show that $\beta = \beta_0$ is needed to have $f_t(\boldsymbol{\theta}_0) = f_t(\boldsymbol{\theta})$ a.s.. If by contradiction we say that $\beta \neq \beta_0$, then to have $f_{t+1}(\boldsymbol{\theta}_0) - f_{t+1}(\boldsymbol{\theta}) = 0$ a.s., we must have $(\alpha_0 - \alpha)\frac{1}{\sigma_0} h(\varepsilon; \boldsymbol{\psi}_0) = (\beta - \beta_0) f_t(\boldsymbol{\theta}_0) \neq 0$ a.s., which is ruled out by $h(\varepsilon; \boldsymbol{\psi}_0)$ and $f_t(\boldsymbol{\theta}_0)$ being independent and non-deterministic. Therefore, we must also have $\beta = \beta_0$, which together with $\omega(1-\beta) = \omega_0(1-\beta_0)$ and the assumption that $|\beta_0| < 1$, implies that we must have $\omega = \omega_0$. Finally, we must clearly also have $\alpha = \alpha_0$, because this is the only way to have $(\alpha_0 - \alpha)\frac{1}{\sigma_0} h\left(\varepsilon_t; \boldsymbol{\psi}_0\right) = 0$ a.s.. This finishes the proof.

□

Proof of Lemma B.7. The expression of $\ell_t''(\boldsymbol{\theta}) = \partial^2 \ell_t(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ is given in Section D.2. By the sub-additivity of the norm $\sup_{\boldsymbol{\theta}\in\Theta} \|\cdot\|$, it suffices to show that the expectation of this norm for each term in this expression is bounded. That all these expectations are bounded follows readily from the following reasons: *(i)* Θ is a compact set such that for any $\boldsymbol{\theta} \in \Theta$, $\sigma > 0$ and $\boldsymbol{\psi}$ is such that $\exp(-H(x; \boldsymbol{\psi}))$ is integrable, *(ii)* by Proposition 1 we have $\mathbb{E}|x_t|^n < \infty$ and by Proposition we have 2 $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta})| < \infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\partial f_t(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}|^n < \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\partial^2 f_t(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top|^n < \infty$ for any $n > 0$, and *(iii)* $h(\cdot; \boldsymbol{\psi})$ being a piecewise cubic function and $h'(\cdot; \boldsymbol{\psi})$ being a piecewise quadratic function for any $\boldsymbol{y} \in \mathbb{R}^k$, and *(iv)* the assumptions on $\boldsymbol{\Psi}$ that ensure that for any $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, $H(x; \boldsymbol{\psi})$ diverges at least linearly as $|x| \to \infty$, which implies that $\inf_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} C(\boldsymbol{\psi}) > 0$ and $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \|\partial C(\boldsymbol{\psi}) / \partial\boldsymbol{\psi}\| < \infty$ and $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \|\partial^2 C(\boldsymbol{\psi}) / \partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top\| < \infty$. For the terms that contain products of random variables, the finiteness of the expectation of the supremum follows from the sub-multiplactivity of the norm and the Cauchy-Schwarz inequalty.

□

Proof of Lemma B.8. We use an approach similar to the proof of Lemma A.5 of Gorgi and Koopman (2021). Due to the correct specification assumption and the existence of a finite moment of the second order derivative of the log likelihood function shown in Lemma B.7, we can use the Fisher information matrix equality $-\mathbb{E}[\ell_t''(\boldsymbol{\theta}_0)] = \mathbb{E}[\ell_t'(\boldsymbol{\theta}_0)\ell_t'(\boldsymbol{\theta}_0)^\top]$. See Equation D.2 for the expression of $\ell_t(\boldsymbol{\theta})$. Because $\mathbb{E}[\ell_t'(\boldsymbol{\theta}_0)\ell_t'(\boldsymbol{\theta}_0)^\top]$ is positive semi-definite by construction, we only have to show that it is full rank, which can be achieved by proving that

$$\boldsymbol{v}^\top \ell_t'(\boldsymbol{\theta}_0) = \boldsymbol{v}^\top \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ -\frac{1}{2\sigma_0^2}\left(1 - \varepsilon_t h\left(\varepsilon_t; \boldsymbol{\psi}_0\right)\right) \\ -\frac{\partial H(\varepsilon_t;\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}} - \frac{1}{C(\boldsymbol{\psi}_0)}\frac{\partial C(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}} \end{pmatrix} + \begin{pmatrix} \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \omega} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \beta} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \alpha} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \sigma^2} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}} \end{pmatrix} \frac{1}{\sigma_0} h\left(\varepsilon_t; \boldsymbol{\psi}_0\right) \right] = 0 \qquad \text{a.s.}, \tag{C.1}$$

if and only if $\boldsymbol{v} = 0$, where $\boldsymbol{v} \in \mathbb{R}^{4+q}$. Furthermore, we use that $f_t(\boldsymbol{\theta}_0)$ is almost surely equal to the true value of f_t in $x_t = f_t + \sigma_0 \varepsilon_t$ due to the correct specification assumption and due to $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ being the unique limit sequence of $\{\hat{f}_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{N}}$ for any $\boldsymbol{\theta} \in \Theta$ by Proposition 2.

Let us split the vector $\boldsymbol{v}$ into $\boldsymbol{v} = (\boldsymbol{v}_1^\top, v_2, \boldsymbol{v}_3^\top)^\top$, where $\boldsymbol{v}_1 \in \mathbb{R}^3$, $v_2 \in \mathbb{R}$ and $\boldsymbol{v}_3 \in \mathbb{R}^q$. Then, $\boldsymbol{v}^\top \ell_t(\boldsymbol{\theta}_0) = 0$ a.s. for $\boldsymbol{v} \neq 0$ only holds in case either *(i)* $v_2 = 0$ and $\boldsymbol{v}_3 \neq 0$, *(ii)* $v_2 \neq 0$ and $\boldsymbol{v}_3 \neq 0$, or *(iii)* $\boldsymbol{v}_1 \neq 0$, $v_2 = 0$ and $\boldsymbol{v}_3 = 0$. We will now prove that neither of these options are possible here. Because $h(\varepsilon_t; \boldsymbol{\psi}_0) \neq 0$ a.s., with probability one we have that (C.1) implies

$$\boldsymbol{v}_1^\top \begin{pmatrix} \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \omega} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \beta} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \alpha} \end{pmatrix} + \begin{pmatrix} v_2 \\ \boldsymbol{v}_3 \end{pmatrix}^\top \begin{pmatrix} \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \sigma^2} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}} \end{pmatrix} = \frac{\sigma_0}{h(\varepsilon_t; \boldsymbol{\psi}_0)} \begin{pmatrix} v_2 \\ \boldsymbol{v}_3 \end{pmatrix}^\top \begin{pmatrix} \frac{1}{2\sigma_0^2}\left(1 - \varepsilon_t h\left(\varepsilon_t; \boldsymbol{\psi}_0\right)\right) \\ \frac{\partial H(\varepsilon_t;\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}} + \frac{1}{C(\boldsymbol{\psi}_0)}\frac{\partial C(\boldsymbol{\psi}_0)}{\partial \boldsymbol{\psi}} \end{pmatrix}. \tag{C.2}$$

The terms on the left-hand side are $\mathcal{G}_{t-1}$-measurable, while those on the right-hand side are not. Thus, this equality can only hold almost surely if both sides are equal to some constant almost surely. We first show that this is not possible under option *(i)*; $v_2 = 0$ and $\boldsymbol{v}_3 \neq 0$. In that case, we must have that $\boldsymbol{v}_3^\top(\partial H(\varepsilon_t; \boldsymbol{\psi}_0)/\partial \boldsymbol{\psi} + C(\boldsymbol{\psi}_0)^{-1}\partial C(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi}) = 0$ a.s., because $(h(\varepsilon_t; \boldsymbol{\psi}_0))^{-1}$ is non-degenerate. As $C(\boldsymbol{\psi}_0)^{-1}\partial C(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi}$ is constant, it suffices to show that $\boldsymbol{v}_3^\top \partial H(\varepsilon_t; \boldsymbol{\psi}_0)/\partial \boldsymbol{\psi}$ is non-degenerate whenever $\boldsymbol{v}_3 \neq 0$. Due to the linear dependence of the cubic spline coefficients on the vector of ordinates $\boldsymbol{y}$, we can write for any $\boldsymbol{\psi} \in \mathbb{R}^q$ and any $z \in \mathbb{R}$:

$$H(z;\boldsymbol{\psi}) = \boldsymbol{\psi}^\top \begin{pmatrix} H(z;\boldsymbol{e}_1) \\ H(z;\boldsymbol{e}_2) \\ \vdots \\ H(z;\boldsymbol{e}_q) \end{pmatrix} = \boldsymbol{\psi}^\top \frac{\partial H(z;\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}, \tag{C.3}$$

where $\boldsymbol{e}_i$ for $i = 1, \ldots, k$ denotes a q-dimensional vector with a 1 on entry i and zeroes elsewhere. As $H(\varepsilon_t;\boldsymbol{\psi})$ is clearly a non-degenerate function of ε_t for any $\boldsymbol{\psi} \neq 0$, it is immediately clear that option *(i)* is ruled out.

We can now rule out option *(ii)* by showing that there exists no nontrivial linear combination of $\varepsilon_t h(\varepsilon_t;\boldsymbol{\psi}_0)$ and $\partial H(\varepsilon_t;\boldsymbol{\psi}_0)/\partial\boldsymbol{\psi}$ that is almost surely equal to a constant. Because by (C.3) $H(\varepsilon_t;\boldsymbol{\psi}) = \boldsymbol{\psi}^\top \partial H(\varepsilon_t;\boldsymbol{\psi})/\partial\boldsymbol{\psi}$, where the derivative does not depend on $\boldsymbol{\psi}$, it suffices to show that there do not exist $\tilde{\boldsymbol{\psi}} \in \mathbb{R}^q$ and $B \in \mathbb{R}$ such that

$$\varepsilon_t h(\varepsilon_t;\boldsymbol{\psi}_0) = H(\varepsilon_t;\tilde{\boldsymbol{\psi}}) + B \quad \text{a.s.}$$

Both functions $zh(z;\boldsymbol{\psi}_0)$ and $H(z;\tilde{\boldsymbol{\psi}})$ are piecewise quartic polynomials in $z \in \mathbb{R}$, but they are inherently different, as the former is only twice continuously differentiable under the assumptions on $\boldsymbol{\Psi}$, while the latter is three times continuously differentiable. In particular, their derivatives are different, as $h(z;\tilde{\boldsymbol{\psi}})$ is a cubic spline function by construction while, $h(z;\boldsymbol{\psi}_0) + zh'(z;\boldsymbol{\psi}_0)$ is not, as $\boldsymbol{\Psi}$ is such that h'' is non-differentiable at the knots $\boldsymbol{t}$ under the current assumptions. Therefore, given that the density function p_ε is nonzero on $\mathbb{R}$, there exist no $\tilde{\boldsymbol{\psi}} \in \mathbb{R}^q$ and $B \in \mathbb{R}$ such that the equality above holds almost surely. Thus, to have that the right-hand side of (C.2) is equal to a constant almost surely, we must have $v_2 = 0$ and $\boldsymbol{v}_3 = 0$.

For case *(iii)*, we would need

$$\frac{1}{\sigma_0} h(\varepsilon_t;\boldsymbol{\psi}_0)\boldsymbol{v}_1^\top \begin{pmatrix} \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\omega} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\beta} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\alpha} \end{pmatrix} = 0 \quad \text{a.s.}\,,$$

for $\boldsymbol{v}_1 \neq 0$. First of all, we know that $h(\varepsilon_t;\boldsymbol{\psi}_0)/\sigma_0$ is non-degenerate. Thus, for the equality to hold almost surely, we must have that the inner product next to it, is equal to zero almost surely. Looking at the expressions of the equations that describe these derivative processes:

$$\frac{\partial f_{t+1}(\boldsymbol{\theta}_0)}{\partial\omega} = B_t \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\omega} + 1 - \beta_0\,,$$

$$\frac{\partial f_{t+1}(\boldsymbol{\theta}_0)}{\partial \beta} = B_t \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \beta} - \omega_0 + \beta_0 f_t(\boldsymbol{\theta}_0)\,,$$
$$\frac{\partial f_{t+1}(\boldsymbol{\theta}_0)}{\partial \alpha} = B_t \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial \alpha} + s(f_t(\boldsymbol{\theta}_0), x_t)\,,$$
$$\text{where} \quad B_t = \beta_0 + \alpha_0 \left.\frac{\partial s(f, x_t)}{\partial f}\right|_{f=f_t(\boldsymbol{\theta}_0)}\,,$$

we see that $\partial f_t(\boldsymbol{\theta}_0)/\partial\omega$, $\partial f_t(\boldsymbol{\theta}_0)/\partial\beta$, and $\partial f_t(\boldsymbol{\theta}_0)/\partial\alpha$, are elements of autoregressive processes that are identical, except for their 'innovations' being equal to $1-\beta_0$, $-\omega_0+f_{t-1}(\boldsymbol{\theta}_0)$ and $h(\varepsilon_{t-1};\boldsymbol{\psi}_0)/\sigma_0$, respectively. Clearly, none of the three processes is degenerate and their 'innovations' take different values with probability one, as 1 is a constant and the remaining two 'innovations' are continuous random variables that are independent of each other. It follows that case (iii) is also ruled out, which concludes the proof.

□

Proof of Lemma B.9. We apply the central limit theorem of Billingsley (1999) for SE martingale difference sequences to $\{\ell_t'(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$, which requires us to show that this sequence is an SE martingale difference sequence with $\mathbb{E}|\ell_t'(\boldsymbol{\theta}_0)|^2 < \infty$. An application of Krengel (1985, Proposition 4.3) tells us that $\{\ell_t'(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE, as $\ell_t'(\boldsymbol{\theta}_0)$ is a continuous function of $f_t(\boldsymbol{\theta}_0)$, $\partial f_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ and x_t, which are elements of SE sequences by Propositions 1 and 2. That $\mathbb{E}|\ell_t'(\boldsymbol{\theta}_0)|^2 < \infty$ can be seen straightforwardly from the expression of $\ell_t'(\boldsymbol{\theta}_0)$ that can be found in (C.1), as it is known that ε_t and $\partial f_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ have bounded moments of any order, and h and H are piecwise polynomials. Finally, $\{\ell_t'(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is a martingale difference sequence, because using the notation from the proof of Lemma B.8, we can write

$$\mathbb{E}\left[\ell_t'(\boldsymbol{\theta}_0)|\mathcal{G}_{t-1}\right] = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\frac{1}{2\sigma_0^2}\left(1-\mathbb{E}[\varepsilon_t h\left(\varepsilon_t;\boldsymbol{\psi}_0\right)]\right) \\ -\mathbb{E}\left[\frac{\partial H(\varepsilon_t;\boldsymbol{\psi}_0)}{\partial\boldsymbol{\psi}}\right] - \frac{1}{C(\boldsymbol{\psi}_0)}\frac{\partial C(\boldsymbol{\psi}_0)}{\partial\boldsymbol{\psi}} \end{pmatrix} + \begin{pmatrix} \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\omega} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\beta} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\alpha} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\sigma^2} \\ \frac{\partial f_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\psi}} \end{pmatrix} \frac{1}{\sigma_0}\mathbb{E}[h(\varepsilon_t;\boldsymbol{\psi}_0)] = 0\,,$$

where we use that the derivative $\partial f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is $\mathcal{G}_{t-1}$ measurable and that $\mathbb{E}[h(\varepsilon_t;\boldsymbol{\psi}_0)] = 0$, because h is rotationally symmetric around zero by construction, which implies that p_ε is a symmetric distribution. Furthermore, $\mathbb{E}[\varepsilon_t h(\varepsilon_t;\boldsymbol{\psi}_0)] = 1$ follows from:

$$\mathbb{E}[\varepsilon_t h(\varepsilon_t;\boldsymbol{\psi}_0)] = \int_{-\infty}^{\infty} x h(x;\boldsymbol{\psi}_0) p_\varepsilon(x)\,\mathrm{d}x = -\int_{-\infty}^{\infty} x \frac{\partial \log p_\varepsilon(x)}{\partial x} p_\varepsilon(x)\,\mathrm{d}x$$

$$= -\int_{-\infty}^{\infty} x p_\varepsilon'(x)\,\mathrm{d}x = -\left[x p_\varepsilon(x) - \mathbb{P}_{X\sim p_\varepsilon}(X \leq x)\right]_{-\infty}^{\infty} = 1\,.$$

Finally, for $i = 1, \ldots, q$, it can be seen using the expressions in Section D.2 that

$$\mathbb{E}\left[\frac{\partial H(\varepsilon_t; \boldsymbol{\psi}_0)}{\partial \psi_i}\right] = \int_{-\infty}^{\infty} H(x; \boldsymbol{e}_i)\frac{\exp(-H(x;\boldsymbol{\psi}_0))}{C(\boldsymbol{\psi}_0)}\,\mathrm{d}x = -\frac{1}{C(\boldsymbol{\psi}_0)}\frac{\partial C(\boldsymbol{\psi}_0)}{\partial \psi_i}\,.$$

Thus, $\{\ell_t'(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is indeed a martingale difference sequence and the central limit theorem can be applied, which together with the Fisher information matrix equality, which was argued to hold in the proof of Lemma B.8, concludes the proof.

□

Proof of Lemma B.10. By Lemma 2.1 of Straumann and Mikosch (2006) and the sub-additivity of the sup-norm, it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\hat{\ell}_t'(\boldsymbol{\theta}) - \ell_t'(\boldsymbol{\theta})\| \overset{e.a.s.}{\to} 0 \quad \text{as } t\to\infty\,.$$

By considering the expression of the first order derivative in (D.2), and using that $\inf_{\boldsymbol{\theta}\in\Theta} 1/\sigma > 0$ under the imposed assumptions, it is clear that the following convergence results are sufficient for this result to hold:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\| h\left(\frac{x_t - \hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)\frac{\partial \hat{f}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - h\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)\frac{\partial f_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \overset{e.a.s.}{\to} 0\,,$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left| h\left(\frac{x_t - \hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)(x_t - \hat{f}_t(\boldsymbol{\theta})) - h\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)(x_t - f_t(\boldsymbol{\theta}))\right| \overset{e.a.s.}{\to} 0\,,$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\| \left.\frac{\partial H(z;\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{z=(x_t-\hat{f}_t(\boldsymbol{\theta}))/\sigma} - \left.\frac{\partial H(z;\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{z=(x_t-f_t(\boldsymbol{\theta}))/\sigma}\right\| \overset{e.a.s.}{\to} 0\,, \tag{C.4}$$

as $t\to\infty$. The first two results follow from applications of Corollary TA.16 of Blasques et al. (2022), which tells us that if $\{\hat{z}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ and $\{\hat{y}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converge e.a.s. to SE limit sequences $\{z_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{y_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ as $t\to\infty$, respectively, uniformly over Θ, then in case $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\log^+|z_t(\boldsymbol{\theta})| < \infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\log^+|y_t(\boldsymbol{\theta})| < \infty$, $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{z}_t(\boldsymbol{\theta})\hat{y}_t(\boldsymbol{\theta}) - z_t(\boldsymbol{\theta})y_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ as $t\to\infty$. It is not hard to see that this result can be extended to the case where either $\{\hat{z}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ or $\{\hat{y}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ is a vector process and the absolute value is replaced by a norm $\|\cdot\|$. Thus, here it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})\right| \overset{e.a.s.}{\to} 0\,,$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \hat{f}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \frac{\partial f_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \overset{e.a.s.}{\to} 0\,,$$

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|h\left(\frac{x_t-\hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)-h\left(\frac{x_t-f_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)\right| \stackrel{e.a.s.}{\to} 0\,,$$

as $t\to\infty$ and that the processes $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, $\{\partial f_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}_{t\in\mathbb{Z}}$ and $\{h((x_t - f_t(\boldsymbol{\theta})/\sigma;\boldsymbol{\psi}))\}_{t\in\mathbb{Z}}$ are SE and have a finite $\log^+$-moment uniformly over Θ. For the first two processes, all the required results follow readily from Proposition 2. For the last process, the stationarity and ergodicity follows from Krengel (1985, Proposition 4.3), as it is a continuous function of SE sequences. Furthermore, the finite log^+ moment follows from the assumptions on Θ, the fact that x_t and $f_t(\boldsymbol{\theta})$ have been shown to have finite moments of any order uniformly over Θ and h being a piecewise third degree polynomial. Finally, for the convergence result, we can apply the mean value theorem, as $h(\cdot;\boldsymbol{\psi})$ is continuously differentiable and has a uniformly bounded derivative under the current assumptions:

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}&\left|h\left(\frac{x_t-\hat{f}_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)-h\left(\frac{x_t-f_t(\boldsymbol{\theta})}{\sigma};\boldsymbol{\psi}\right)\right|\\ &\leq \underbrace{\sup_{\boldsymbol{\theta}\in\Theta}\sup_{z\in\mathbb{R}}\left|\frac{h'(z;\boldsymbol{\psi})}{\sigma}\right|}_{=D<\infty}\cdot\sup_{\boldsymbol{\theta}\in\Theta}\left|\hat{f}_t(\boldsymbol{\theta})-f_t(\boldsymbol{\theta})\right| \stackrel{e.a.s.}{\to} 0\,.\end{aligned}$$

Now it only remains to be shown that (C.4) holds, which can be shown using the same approach as is used in the proof of Lemma B.4, because $\partial H(z;\boldsymbol{\psi})/\partial\psi_i = H(z;\boldsymbol{e}_i)$ for $i=1,\ldots,q$.

□

Proof of Lemma B.11. Using the same argument as in the proof of Lemma B.4, it is sufficient to show that $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{\ell}_t(\boldsymbol{\theta})-\ell_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$ as $t\to\infty$. The expression of $\ell_t(\boldsymbol{\theta})$ is given by

$$\ell_t(\boldsymbol{\theta}) = -H\left(\frac{x_t-m}{\exp(f_t(\boldsymbol{\theta}))};\boldsymbol{\psi}\right)-\log C(\boldsymbol{\psi})-f_t(\boldsymbol{\theta})\,. \tag{C.5}$$

As only the first and last term depend on $f_t(\boldsymbol{\theta})$, the middle term can be disregarded. For the final term we have

$$\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t(\boldsymbol{\theta})-f_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty\,,$$

which follows directly from Proposition 3. Thus, it only remains to be shown that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|H\left(\frac{x_t-m}{\exp(\hat{f}_t(\boldsymbol{\theta}))};\boldsymbol{\psi}\right)-H\left(\frac{x_t-m}{\exp(f_t(\boldsymbol{\theta}))};\boldsymbol{\psi}\right)\right| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty\,,$$

By arguments used in the proof of Lemma B.4, it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{x_t - m}{\exp(\hat{f}_t(\boldsymbol{\theta}))} - \frac{x_t - m}{\exp(f_t(\boldsymbol{\theta}))}\right| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t \to \infty\,,$$

and that $\{(x_t - m)/\exp(f_t(\boldsymbol{\theta}))\}_{t\in\mathbb{Z}}$ is SE and has a bounded $\log^+$-moment uniformly over Θ. The convergence result holds by an application of the mean value theorem:

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{x_t - m}{\exp(\hat{f}_t(\boldsymbol{\theta}))} - \frac{x_t - m}{\exp(f_t(\boldsymbol{\theta}))}\right| &\leq \sup_{\boldsymbol{\theta}\in\Theta, f\in\mathcal{F}}\left|\frac{x_t - m}{\exp(f)}\right| \sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})| \\ &\leq \frac{|x_t| + \bar{m}}{\exp(\underline{f})} \sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})|\end{aligned}$$

as we know $f_t(\boldsymbol{\theta}) \in \mathcal{F} = [\underline{f}, \infty)$ for some $\underline{f} > -\infty$, and where $\bar{m} = \sup_{\boldsymbol{\theta}\in\Theta} m < \infty$. It follows from Proposition 3 that $\{|x_t|\}_{t\in\mathbb{Z}}$ is SE and also:

$$\begin{aligned}\mathbb{E}(\log^+|x_t|) &= \mathbb{E}(\log^+|m_0 + \exp(f_t)\varepsilon_t|) \\ &\leq 2\log 2 + \log^+|m_0| + \mathbb{E}(\log^+|\exp(f_t)\varepsilon_t|) \\ &\leq 2\log 2 + \log^+|m_0| + \mathbb{E}(\log^+\exp(f_t)) + \mathbb{E}(\log^+|\varepsilon_t|) < \infty\end{aligned}$$

where the inequalities follow from Lemma 2.2 of Straumann and Mikosch (2006). The final expression is finite, because $\mathbb{E}(\log^+\exp(f_t)) = \mathbb{E}(\mathbb{1}_{f_t>0}f_t) \leq \mathbb{E}|f_t| < \infty$, as f_t has bounded moments of any order by Proposition 3, and $\mathbb{E}(\log^+|\varepsilon_t|) < \infty$ as ε_t has bounded moments of any order.

Finally, that $\{(x_t - m)/\exp(f_t(\boldsymbol{\theta}))\}_{t\in\mathbb{Z}}$ is SE follows from Krengel (1985, Proposition 4.3) and that the elements of this sequence have a bounded $\log^+$-moment uniformly over Θ holds by another application of Lemma 2.2 of Straumann and Mikosch (2006), and the fact that $f_t(\boldsymbol{\theta})$ is bounded from below by $\underline{f} > -\infty$. □

Proof of Lemma B.12. The log likelihood function $\log p_x(x|f;\boldsymbol{\gamma})$ is uniformly bounded from above over $x \in \mathbb{R}$, $f \in \mathcal{F}$ and $\boldsymbol{\theta} \in \Theta$, as can be seen from its expression in (C.5), as we have that Θ is compact, $f_t(\boldsymbol{\theta}) \geq \underline{f} > -\infty$ a.s., $\boldsymbol{\psi}$ is such that $H(z)$ diverges to $+\infty$ as $|z| \to \infty$, which together with $H(z;\boldsymbol{\psi})$ being finite on $\mathbb{R}$, implies that $H(z;\boldsymbol{\psi})$ is uniformly bounded from below. Also, $C(\boldsymbol{\psi})$ is finite for any $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ under the imposed conditions. Therefore, $\log p_x(x|f;\boldsymbol{\gamma})$ is uniformly bounded from above, implying that $\mathbb{E}\ell_t(\boldsymbol{\theta}) < \infty$ exists for any $\boldsymbol{\theta}$, although it can be equal to $-\infty$.

The second result holds because

$$\mathbb{E}|\ell_t(\boldsymbol{\theta}_0)| \leq \mathbb{E}\left|H\left(\frac{x_t - m_0}{\exp(f_t(\boldsymbol{\theta}_0))};\boldsymbol{\psi}_0\right)\right| + |\log C(\boldsymbol{\psi}_0)| + \mathbb{E}|f_t(\boldsymbol{\theta}_0)|$$

$$= \mathbb{E}\left|H\left(\varepsilon_t;\boldsymbol{\psi}_0\right)\right| + |\log C(\boldsymbol{\psi}_0)| + \mathbb{E}|f_t(\boldsymbol{\theta}_0)| < \infty\,,$$

where we use that $f_t(\boldsymbol{\theta}_0) = f_t$ a.s., due to $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ being a unique limit sequence. Each individual term in the expression on the right-hand side is finite, due to $f_t(\boldsymbol{\theta}_0)$ having bounded moments of any order by Proposition 4 and H being a piecewise quartic polynomial evaluated in ε_t, which has bounded moments of any order due to the exponential decay of the tails of its distribution. Furthermore, clearly $C(\boldsymbol{\psi}_0) < \infty$ under the assumptions that are in place. □

Proof of Lemma B.13. We know from Lemma B.12 that $\ell_t(\boldsymbol{\theta})$ is bounded from above, implying it is integrable, such that $\mathbb{E}\ell_t(\boldsymbol{\theta})$ exists for any $\boldsymbol{\theta} \in \Theta$. That lemma also tells us that $\mathbb{E}|\ell_t(\boldsymbol{\theta}_0)| < \infty$. Clearly, whenever $\boldsymbol{\theta} \in \Theta$ is such that $\mathbb{E}\ell_t(\boldsymbol{\theta}) = -\infty$, then $\mathbb{E}\ell_t(\boldsymbol{\theta}) < \mathbb{E}\ell_t(\boldsymbol{\theta}_0)$, so we just have to consider vectors $\boldsymbol{\theta} \in \Theta$ for which $\mathbb{E}\ell_t(\boldsymbol{\theta})$ is finite.

As for Lemma B.6, we prove the claim by showing that $\ell_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}_0)$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

We have argued in the proof of Lemma B.6 that if $\varepsilon_1 \sim p_\varepsilon(\varepsilon_1;\boldsymbol{\psi})$, then $a' + H((\varepsilon_1 - c')/b';\boldsymbol{\psi}') = a'' + H((\varepsilon_1 - c'')/b'';\boldsymbol{\psi}'')$ a.s. if and only if $(a', b', c', \boldsymbol{\psi}') = (a'', b'', c'', \boldsymbol{\psi}'')$. Thus, because $\ell_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}_0)$ a.s. if and only if

$$-H\left(\frac{\exp(f_t(\boldsymbol{\theta}_0))\varepsilon_t + m_0 - m}{\exp(f_t(\boldsymbol{\theta}))};\boldsymbol{\psi}\right) - \log C(\boldsymbol{\psi}) - f_t(\boldsymbol{\theta})$$
$$= -H\left(\varepsilon_t;\boldsymbol{\psi}_0\right) - \log C(\boldsymbol{\psi}_0) - f_t(\boldsymbol{\theta}_0)$$

a.s., it is clear that $\ell_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}_0)$ holds almost surely if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, $m_0 = m$ and $f_t(\boldsymbol{\theta}) = f_t(\boldsymbol{\theta}_0)$ a.s. So it remains to be shown that given $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ and $m = m_0$, $f_t(\boldsymbol{\theta}) = f_t(\boldsymbol{\theta}_0)$ a.s. holds if and only if $(\omega, \beta, \alpha) = (\omega_0, \beta_0, \alpha_0)$. Because $\{f_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE for any $\boldsymbol{\theta} \in \Theta$, then if we have $f_t(\boldsymbol{\theta}) = f_t(\boldsymbol{\theta}_0)$ almost surely, this must hold for any $t \in \mathbb{Z}$. Say that $\boldsymbol{\theta} \in \Theta$ is such that the equality holds almost surely, then the update equation of $f_t(\boldsymbol{\theta})$ implies that:

$$f_{t+1}(\boldsymbol{\theta}_0) - f_{t+1}(\boldsymbol{\theta})$$
$$= \omega_0(1-\beta_0) - \omega(1-\beta) + (\beta_0 - \beta)f_t(\boldsymbol{\theta}_0) + (\alpha_0 - \alpha)(h(\varepsilon_t;\boldsymbol{\psi}_0)\varepsilon_t - 1) \quad \text{a.s.}$$

To have that the left-hand side is equal to zero almost surely, we must have $\omega_0(1-\beta_0) = \omega(1-\beta)$, because otherwise, the equality implies that $(\beta_0-\beta)f_t(\boldsymbol{\theta}_0)+(\alpha_0-\alpha)(h(\varepsilon_t;\boldsymbol{\psi}_0)\varepsilon_t - 1) = \omega(1-\beta) - \omega_0(1-\beta_0) \neq 0$, which is ruled out, because the left-hand side cannot be equal to a nonzero constant due to $f_t(\boldsymbol{\theta}_0)$ being independent of $h(\varepsilon_t)\varepsilon_t$, and $f_t(\boldsymbol{\theta}_0)$

being non-degenerate due to $\alpha_0 \neq 0$ and $h(\varepsilon_t)\varepsilon_t$ clearly being non-degenerate under the maintained assumptions. Next, we also need $\beta_0 = \beta$ for the equality above to hold a.s., because in case $\beta_0 \neq \beta$, we would need $(\alpha_0 - \alpha)(h(\varepsilon_t; \boldsymbol{\psi}_0)\varepsilon_t - 1) = (\beta - \beta_0) f_t(\boldsymbol{\theta}_0)$, which is not possible due to $h(\varepsilon_t)\varepsilon_t$ and $(\beta - \beta_0) f_t(\boldsymbol{\theta}_0)$ being independent and non-degenerate. It then follows from $\omega_0(1 - \beta_0) = \omega(1 - \beta)$ and the assumption that $\beta_0 < 1$ that $\omega = \omega_0$. Finally, we must clearly also have $\alpha = \alpha_0$, as there is no other way of having $(\alpha_0 - \alpha)(h(\varepsilon_t; \boldsymbol{\psi}_0)\varepsilon_t - 1) = 0$ a.s.. This concludes the proof.

□

D. Derivatives location model

D.1. Derivatives of score function location model

In this section we provide the first, second and selected third order derivatives of $s(f, x_t; \boldsymbol{\gamma})$ with respect to $\boldsymbol{\theta}$, so effectively $\boldsymbol{\gamma} = (\sigma^2, \boldsymbol{\psi})$, and f, necessary for deriving the properties of the derivatives of $f_t(\boldsymbol{\theta})$. Recall that here $\boldsymbol{\theta} = (\omega, \beta, \alpha, \sigma^2, \boldsymbol{\psi})^\top$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_q)$ for $q = k/2$, and where $\tau(\boldsymbol{\psi})$ is such that $y_1 = -\psi_k$, ..., $y_q = -\psi_1$, $y_{q+1} = \psi_1$, ..., $y_k = \psi_k$. We have that for any $f \in \mathbb{R}$, $x \in \mathbb{R}$ and $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$:

$$s(f, x; \boldsymbol{\gamma}) = \frac{1}{\sigma} h\left(\frac{x - f}{\sigma}; \boldsymbol{\psi}\right).$$

For the first order derivatives we will use that the nonzero elements of the derivative $\partial h(z; \boldsymbol{\psi})/\partial \boldsymbol{\theta}$ are given by:

$$\frac{\partial h\left(z; \boldsymbol{\psi}\right)}{\partial \psi_i} = h(z; \boldsymbol{e}_i)\,,$$

for any $z \in \mathbb{R}$ and $i = 1, \ldots, q$, and where $\boldsymbol{e}_j$ for $j = 1, \ldots, q$ denotes a q-dimensional vector with a 1 on entry j and zeroes elsewhere. This follows from the linearity of the function $h(z; \boldsymbol{\psi})$ in $\boldsymbol{\psi}$. It follows that for any $i = 1, \ldots, q$

$$\begin{aligned}
\frac{\partial s(f, x; \boldsymbol{\gamma})}{\partial f} &= -\frac{1}{\sigma^2} h'\left(\frac{x - f}{\sigma}; \boldsymbol{\psi}\right)\,, \\
\frac{\partial s(f, x; \boldsymbol{\gamma})}{\partial \sigma^2} &= -\frac{1}{2\sigma^3}\left[h\left(\frac{x - f}{\sigma}; \boldsymbol{\psi}\right) + h'\left(\frac{x - f}{\sigma}; \boldsymbol{\psi}\right)\frac{x - f}{\sigma}\right]\,, \\
\frac{\partial s(f, x; \boldsymbol{\gamma})}{\partial \psi_i} &= \frac{1}{\sigma} h\left(\frac{x - f}{\sigma}; \boldsymbol{e}_i\right)\,,
\end{aligned}$$

where derivative $h'(z; \boldsymbol{\psi})$ denotes the derivative of the spline function h with respect to z, which is a second degree piecewise polynomial.

For second order derivatives of s, we will use that by the linearity of $h(z;\boldsymbol{\psi})$ in $\boldsymbol{y}$, the second order derivative of h with respect to the static parameters is a matrix of zeroes:

$$\frac{\partial^2 h\left(z;\boldsymbol{\psi}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \mathbf{0}_{4+q\times 4+q}\,.$$

Furthermore, the nonzero elements of $\partial h'(z;\boldsymbol{\psi})/\partial\boldsymbol{\theta}$ are given by

$$\frac{\partial h'(z;\boldsymbol{\psi})}{\partial\psi_i} = h'(z;\boldsymbol{e}_i)\,,$$

for any $z\in\mathbb{R}$ and $i=1,\ldots,q$. Thus, we have for any $i=1,\ldots,q$ and $j=1,\ldots,q$

$$\begin{aligned}
\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial f\partial f} &= \frac{1}{\sigma^3}h''\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right)\\
\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial\sigma^2\partial f} &= \frac{1}{\sigma^4}h'\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right) - \frac{x-f}{\sigma^5}h''\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right)\,,\\
\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial\psi_i\partial f} &= -\frac{1}{\sigma^2}h'\left(\frac{x-f}{\sigma};\boldsymbol{e}_i\right)\,,\\
\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial\sigma^2\partial\sigma^2} &= \frac{3}{4\sigma^5}h\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right) + \frac{5(x-f)}{4\sigma^6}h'\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right)\\
&\quad + \frac{(x-f)^2}{4\sigma^7}h''\left(\frac{x-f}{\sigma};\boldsymbol{\psi}\right)\,,\\
\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial\sigma^2\partial\psi_i} &= -\frac{1}{2\sigma^3}\left[h\left(\frac{x-f}{\sigma};\boldsymbol{e}_i\right)\right.\\
&\qquad\qquad \left. + h'\left(\frac{x-f}{\sigma};\boldsymbol{e}_i\right)\frac{x-f}{\sigma}\right]\,,\\
\frac{\partial^2 s(f,x;\boldsymbol{\gamma})}{\partial\psi_i\partial\psi_j} &= 0\,.
\end{aligned}$$

Notice that $h''(z;\boldsymbol{\psi})$ is a continuous piecewise linear function and can be straightforwardly computed given $h'(z;\boldsymbol{\psi})$.

Finally, we compute the following three selected third order derivatives for $i=1,\ldots,q$ and $j=1,\ldots,q$:

$$\begin{aligned}
\frac{\partial^3 s(f,x;\boldsymbol{\gamma})}{\partial\psi_i\partial\psi_j\partial f} &= 0\,,\\
\frac{\partial^3 s(f,x;\boldsymbol{\gamma})}{\partial\sigma^2\partial\psi_i\partial f} &= \frac{1}{\sigma^4}h'\left(\frac{x-f}{\sigma};\boldsymbol{e}_i\right) - \frac{x-f}{\sigma^5}h''\left(\frac{x-f}{\sigma};\boldsymbol{e}_i\right)\,,\\
\frac{\partial^3 s(f,x;\boldsymbol{\gamma})}{\partial\psi_i\partial f\partial f} &= \frac{1}{\sigma^3}h''\left(\frac{x-f}{\sigma};\boldsymbol{e}_i\right)\,.
\end{aligned}$$

D.2. Derivatives of log likelihood location model

In this section we give the first and second order derivatives of the log likelihood function of the location model defined in Section 4.1, with respect to $\boldsymbol{\theta}$ where $\boldsymbol{\theta}=(\omega,\beta,\alpha,\sigma^2,\boldsymbol{\psi})^\top$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_q)$ for $q = k/2$, and where $\tau(\boldsymbol{\psi})$ is such that $y_1 = -\psi_k, \ldots, y_q = -\psi_1$, $y_{q+1} = \psi_1, \ldots, y_k = \psi_k$. For any $\boldsymbol{\theta} \in \Theta$, the log likelihood function reads:

$$\frac{1}{T}\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta})\,, \quad \text{where} \quad \ell_t(\boldsymbol{\theta}) = -H\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma}; \boldsymbol{\psi}\right) - \log C(\boldsymbol{\psi}) - \frac{1}{2}\log\sigma^2\,. \tag{D.1}$$

The first order derivative of $\ell_t(\boldsymbol{\theta})$ is given by:

$$\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = h\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma}; \boldsymbol{\psi}\right)\left[\frac{1}{\sigma}\frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - (x_t - f_t(\boldsymbol{\theta}))\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}}\right] - \left.\frac{\partial H\,(z;\boldsymbol{\psi})}{\partial \boldsymbol{\theta}}\right|_{z=(x_t - f_t(\boldsymbol{\theta}))/\sigma}$$
$$- \frac{1}{C(\boldsymbol{\psi})}\frac{\partial C(\boldsymbol{\psi})}{\partial \boldsymbol{\theta}} - \frac{1}{2\sigma^2}\frac{\partial \sigma^2}{\partial \boldsymbol{\theta}}\,, \tag{D.2}$$

where the nonzero elements of $\partial H(z;\boldsymbol{\psi})/\partial\boldsymbol{\theta}$, $\partial C(\boldsymbol{\psi})/\partial\boldsymbol{\theta}$, $\partial\sigma^{-1}/\partial\boldsymbol{\theta}$ and $\partial\sigma^2/\partial\sigma^2$ are given by

$$\begin{aligned}
\frac{\partial H(z;\boldsymbol{\psi})}{\partial \psi_i} &= H(z;\boldsymbol{e}_i)\,,\\
\frac{\partial C(\boldsymbol{\psi})}{\partial \psi_i} &= \int_{-\infty}^{\infty} \frac{\partial \exp(-H(x;\boldsymbol{\psi}))}{\partial \boldsymbol{\pi}_i}\,\mathrm{d}x\\
&= \int_{-\infty}^{\infty} -\exp(-H(x;\boldsymbol{\psi}))H(x;\boldsymbol{e}_i)\,\mathrm{d}x\,,\\
\frac{\partial \sigma^{-1}}{\partial \sigma^2} &= -\frac{1}{2\sigma^3}\,,\\
\frac{\partial \sigma^2}{\partial \sigma^2} &= 1\,,
\end{aligned}$$

for $i = 1, \ldots, q$ and any $z \in \mathbb{R}$. The form of $\partial H(z;\boldsymbol{\psi})/\partial\psi_i$ follows from the linearity of the spline coefficients in $\boldsymbol{y}$. We will now argue why the derivative can be taken into the integral in the calculation of $\partial C(\boldsymbol{\psi})/\partial\psi_i$. Because Θ is compact, there must be some real positive value $t^* > 0$ such that for some $\boldsymbol{\psi}^* \in \boldsymbol{\Psi}$ and every $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ and corresponding $\boldsymbol{y}$ and $\boldsymbol{y}^*$, respectively, $|\partial \exp(-H(x;\boldsymbol{\psi}^*))/\partial\psi_i| \geq \partial\exp(-H(x;\boldsymbol{\psi}))/\partial\psi_i$ for every $x > t^*$ or $x < -t^*$. Within $[-t^*, t^*]$, the derivative can be taken into the integral by Leibniz rule and outside of this interval, the derivative can be taken into the integral by an application of the dominated convergence theorem, as $|\partial\exp(-H(x;\boldsymbol{\psi}^*))/\partial\psi_i|$ is clearly integrable under the maintained assumptions.

For the second order derivative, we obtain

$$\begin{aligned}
\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = -h'&\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma}; \boldsymbol{\psi}\right)\\
&\left[\frac{1}{\sigma}\frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - (x_t - f_t(\boldsymbol{\theta}))\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}}\right]\left[\frac{1}{\sigma}\frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - (x_t - f_t(\boldsymbol{\theta}))\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}^\top}\right]\\
+ h&\left(\frac{x_t - f_t(\boldsymbol{\theta})}{\sigma}; \boldsymbol{\psi}\right)\left[\frac{1}{\sigma}\frac{\partial^2 f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}}\frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right.
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad - (x_t - f_t(\boldsymbol{\theta}))\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Bigg] \\
&+ \left[\frac{1}{\sigma}\frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - (x_t - f_t(\boldsymbol{\theta}))\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}}\right] \frac{\partial h(z;\boldsymbol{\psi})}{\partial \boldsymbol{\theta}^\top}\bigg|_{z=(x_t - f_t(\boldsymbol{\theta}))/\sigma} \\
&+ \frac{\partial h(z;\boldsymbol{\psi})}{\partial \boldsymbol{\theta}}\bigg|_{z=(x_t - f_t(\boldsymbol{\theta}))/\sigma} \left[\frac{1}{\sigma}\frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} - (x_t - f_t(\boldsymbol{\theta}))\frac{\partial \sigma^{-1}}{\partial \boldsymbol{\theta}^\top}\right] \\
&- \frac{1}{C(\boldsymbol{\psi})}\frac{\partial^2 C(\boldsymbol{\psi})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} + \frac{1}{C(\boldsymbol{\psi})^2}\left(\frac{\partial C(\boldsymbol{\psi})}{\partial \boldsymbol{\theta}}\right)^2 - \frac{1}{2}\frac{\partial \sigma^2}{\partial \boldsymbol{\theta}}\frac{\partial \sigma^{-2}}{\partial \boldsymbol{\theta}^\top},
\end{aligned}
$$

where we use that

$$
\frac{\partial^2 H(z;\boldsymbol{\psi})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \mathbf{0}_{4+q\times 4+q}, \quad \text{and} \quad \frac{\partial^2 \sigma^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \mathbf{0}_{4+q\times 4+q},
$$

because $H(z;\boldsymbol{\psi})$ is linear in $\boldsymbol{y}$ and because $\partial\sigma^2/\partial\sigma^2 = 1$. The expressions of the nonzero elements of $\partial h(z;\boldsymbol{\psi})/\partial\boldsymbol{\theta}$ are given in the previous subsection. Finally, the nonzero elements of $\partial^2\sigma^{-1}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$, $\partial\sigma^{-2}/\partial\sigma^2$ and $\partial^2 C(\boldsymbol{\psi})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are given by

$$
\begin{aligned}
\frac{\partial^2 \sigma^{-1}}{\partial \sigma^2 \partial \sigma^2} &= \frac{3}{4\sigma^5}, \\
\frac{\partial \sigma^{-2}}{\partial \sigma^2} &= -\frac{1}{\sigma^4}, \\
\frac{\partial^2 C(\boldsymbol{\psi})}{\partial \psi_i \partial \psi_j} &= \int_{-\infty}^{\infty} \exp(-H(x;\boldsymbol{\psi}))H(x;\boldsymbol{e}_i)H(x;\boldsymbol{t},\boldsymbol{e}_j)\,\mathrm{d}x,
\end{aligned}
$$

for any $i,j = 1,\dots,q$, where we use the same argumentation as for $\partial C(\boldsymbol{\psi})/\partial\boldsymbol{\psi}$ for taking the derivative into the integral.

E. Calculation of moments

Let us derive the moments of the spline probability distribution p_ε. For any non-negative integer n, we have that:

$$
\begin{aligned}
C\mathbb{E}[\varepsilon_t^n] &= \int_{-\infty}^{\infty} \varepsilon^n \exp(-H(\varepsilon;\boldsymbol{t},\boldsymbol{y}))\mathrm{d}\varepsilon \\
&= \int_{-\infty}^{t_1} \varepsilon^n \exp(-H(\varepsilon;\boldsymbol{t},\boldsymbol{y}))\mathrm{d}\varepsilon + \int_{t_1}^{t_k} \varepsilon^n \exp(-H(\varepsilon;\boldsymbol{t},\boldsymbol{y}))\mathrm{d}\varepsilon \\
&\quad + \int_{t_k}^{\infty} \varepsilon^n \exp(-H(\varepsilon;\boldsymbol{t},\boldsymbol{y}))\mathrm{d}\varepsilon,
\end{aligned} \tag{E.1}
$$

where C denotes the integrating constant defined in (2). We will discuss how this constant can be calculated in Section E.1. The middle integral can be reliably evaluated numerically, because it is a smooth function that is integrated over a bounded integration range. In particular, we suggest using some quadrature routine between each pair of subsequent knots (t_i, t_{i+1}) for $i = 1,\dots,k-1$, in order to obtain a good approximation.

The first and last integrals can be simplified analytically. Note that beyond the outer knots, the natural cubic spline h is linear and therefore its anti-derivative is quadratic. For the tail integrals to be finite, the assumption on the coefficients in Assumption 1 must hold. Because the coefficients of the piecewise polynomial are linear in the y-values $\boldsymbol{y}$, the conditions $b_0 > 0$ and $b_k > 0$ can be imposed straightforwardly using a linear constraint on the elements of $\boldsymbol{y}$.

To evaluate the tail integrals, we have to distinguish between the case where $b_0 > 0$ ($b_k > 0$) and where $b_0 = 0$ and $a_0 < 0$ ($b_k = 0$ and $a_k > 0$). Let us start with the first integral on the right-hand side of (E.1) for the case $b_0 > 0$:

$$\begin{aligned}
\int_{-\infty}^{t_1} \varepsilon^n \exp(-H(\varepsilon; \boldsymbol{t}, \boldsymbol{y}))\mathrm{d}\varepsilon &= \int_{-\infty}^{t_1} \varepsilon^n \exp(-H_0(\varepsilon; \boldsymbol{t}, \boldsymbol{y}))\mathrm{d}\varepsilon \\
&= \exp(-e_0) \int_{-\infty}^{t_1} \varepsilon^n \exp\left(-a_0(\varepsilon - t_1) - \frac{1}{2}b_0(\varepsilon - t_1)^2\right) \mathrm{d}\varepsilon \\
&= \exp(-e_0) \int_{-\infty}^{0} (y + t_1)^n \exp\left(-a_0 y - \frac{1}{2}b_0 y^2\right) \mathrm{d}y \\
&= \sqrt{2\pi b_0^{-1}} \exp\left(-e_0 + \frac{1}{2}a_0^2 b_0^{-1}\right) \int_{-\infty}^{0} (y + t_1)^n \underbrace{\frac{1}{\sqrt{2\pi b_0^{-1}}} \exp\left(-\frac{(y + a_0 b_0^{-1})^2}{2b_0^{-1}}\right)}_{\text{pdf of } \mathcal{N}(-a_0 b_0^{-1}, b_0^{-1})} \mathrm{d}y \\
&= \tilde{\sigma}_0 \frac{\Phi(\tilde{\beta}_0)}{\varphi(\tilde{\beta}_0)} \exp(-e_0) \int_{-\infty}^{0} (y + t_1)^n \underbrace{\frac{1}{\Phi(\tilde{\beta}_0)\sqrt{2\pi}\tilde{\sigma}_0} \exp\left(-\frac{(y - \tilde{\mu}_0)^2}{2\tilde{\sigma}_0^2}\right)}_{\text{pdf of } tr\mathcal{N}(\tilde{\mu}_0, \tilde{\sigma}_0^2, -\infty, 0)} \mathrm{d}y \\
&= \tilde{\sigma}_0 \frac{\Phi(\tilde{\beta}_0)}{\varphi(\tilde{\beta}_0)} \exp(-e_0) \mathbb{E}_{Z \sim tr\mathcal{N}(\tilde{\mu}_0, \tilde{\sigma}_0^2, -\infty, 0)} \left[(Z + t_1)^n\right] ,
\end{aligned}$$

where we use the notation $\tilde{\mu}_0 = -a_0 b_0^{-1}$, $\tilde{\sigma}_0^2 = b_0^{-1}$ and $\tilde{\beta}_0 = -\tilde{\mu}_0/\tilde{\sigma}_0 = a_0\sqrt{b_0^{-1}}$, where we use the change of variable $y = \varepsilon - t_1$ in the third equality and where $tr\mathcal{N}(\mu, \sigma^2, a, b)$ for $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and $-\infty \leq a < b \leq \infty$, denotes the truncated normal distribution with mean μ, variance σ^2 and truncation on the interval (a, b). Furthermore, $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and $\varphi(\cdot)$ denotes its probability density function. For the last integral in (E.1), under the restriction $b_k > 0$, a similar derivation gives:

$$\int_{t_k}^{\infty} \varepsilon^n \exp(-H(\varepsilon; \boldsymbol{t}, \boldsymbol{y}))\mathrm{d}\varepsilon = \tilde{\sigma}_k \frac{\Phi(-\tilde{\beta}_k)}{\varphi(\tilde{\beta}_k)} \exp(-e_k) \mathbb{E}_{Z \sim tr\mathcal{N}(\tilde{\mu}_k, \tilde{\sigma}_k^2, 0, \infty)} \left[(Z + t_k)^n\right] ,$$

where $\tilde{\mu}_k = -a_k b_k^{-1}$, $\tilde{\sigma}_k^2 = b_k^{-1}$ and $\tilde{\beta}_k = -\tilde{\mu}_k/\tilde{\sigma}_k = a_k\sqrt{b_k^{-1}}$.

For the case $b_0 = 0$ and $a_0 < 0$, the first integral on the right-hand side of (E.1) is equal to

$$\begin{aligned}\int_{-\infty}^{t_1} \varepsilon^n \exp(-H(\varepsilon;\boldsymbol{t},\boldsymbol{y}))\mathrm{d}\varepsilon &= \int_{-\infty}^{t_1} \varepsilon^n \exp(-H_0(\varepsilon;\boldsymbol{t},\boldsymbol{y}))\mathrm{d}\varepsilon \\ &= \exp(-e_0)\int_{-\infty}^{t_1} \varepsilon^n \ \exp\left(-a_0(\varepsilon - t_1)\right)\mathrm{d}\varepsilon \\ &= \exp(-e_0)\int_{0}^{\infty} (t_1 - y)^n \ \exp\left(-(-a_0)y\right)\mathrm{d}y \\ &= -\frac{\exp(-e_0)}{a_0}\int_{0}^{\infty} (t_1 - y)^n \ \underbrace{(-a_0)\exp\left(-(-a_0)y\right)}_{\text{pdf of Exp}(\lambda)\text{ with }\lambda=-a_0}\mathrm{d}y \\ &= -\frac{\exp(-e_0)}{a_0}\mathbb{E}_{Z\sim\text{Exp}(-a_0)}\left[(t_1 - Z)^n\right] ,\end{aligned}$$

where we use the change of variable $y = t_1 - \varepsilon$ in the third equality and where $\text{Exp}(\lambda)$ denotes the exponential distriubtion with parameter $\lambda > 0$. The last integral in (E.1), under the restriction $b_k = 0$ and $a_k > 0$, can be shown to be equal to $\frac{\exp(-e_k)}{a_k}\mathbb{E}_{Z\sim\text{Exp}(a_k)}\left[(Z + t_k)^n\right]$ using a similar derivation.

In summary, we have found that for any integer n,

$$\begin{aligned}\int_{-\infty}^{t_1} \varepsilon^n \exp(-H(\varepsilon))\mathrm{d}\varepsilon &= \begin{cases} \tilde{\sigma}_0\frac{\Phi(\tilde{\beta}_0)}{\varphi(\tilde{\beta}_0)}\exp(-e_0)\mathbb{E}_{Z\sim tr\mathcal{N}(\tilde{\mu}_0,\tilde{\sigma}_0^2,-\infty,0)}\left[(Z+t_1)^n\right], & \text{if } b_0 > 0\,, \\ -\frac{\exp(-e_0)}{a_0}\mathbb{E}_{Z\sim\text{Exp}(-a_0)}\left[(t_1 - Z)^n\right], & \text{if } b_0 = 0 \text{ and } a_0 < 0\,, \end{cases} \\ \int_{t_k}^{\infty} \varepsilon^n \exp(-H(\varepsilon))\mathrm{d}\varepsilon &= \begin{cases} \tilde{\sigma}_k\frac{\Phi(-\tilde{\beta}_k)}{\varphi(\tilde{\beta}_k)}\exp(-e_k)\mathbb{E}_{Z\sim tr\mathcal{N}(\tilde{\mu}_k,\tilde{\sigma}_k^2,0,\infty)}\left[(Z+t_k)^n\right], & \text{if } b_k > 0\,, \\ \frac{\exp(-e_k)}{a_k}\mathbb{E}_{Z\sim\text{Exp}(a_k)}\left[(Z + t_k)^n\right], & \text{if } b_k = 0 \text{ and } a_k > 0\,, \end{cases}\end{aligned} \tag{E.2}$$

where $\tilde{\mu}_0 = -a_0 b_0^{-1}$, $\tilde{\sigma}_0^2 = b_0^{-1}$ and $\tilde{\beta}_0 = -\tilde{\mu}_0/\tilde{\sigma}_0 = a_0\sqrt{b_0^{-1}}$ and where $\tilde{\mu}_k = -a_k b_k^{-1}$, $\tilde{\sigma}_k^2 = b_k^{-1}$ and $\tilde{\beta}_k = -\tilde{\mu}_k/\tilde{\sigma}_k = a_k\sqrt{b_k^{-1}}$.

For the first and second moment we can use that if $Z \sim tr\mathcal{N}(\tilde{\mu}_0, \tilde{\sigma}_0^2, -\infty, 0)$:

$$\mathbb{E}\left[Z\right] = \tilde{\mu}_0 - \frac{\varphi(\tilde{\beta}_0)}{\Phi(\tilde{\beta}_0)}\tilde{\sigma}_0\,, \quad \text{and} \quad \mathbb{V}\text{ar}\left(Z\right) = \tilde{\sigma}_0^2\left[1 - \frac{\tilde{\beta}_0\varphi(\tilde{\beta}_0)}{\Phi(\tilde{\beta}_0)} - \left(\frac{\varphi(\tilde{\beta}_0)}{\Phi(\tilde{\beta}_0)}\right)^2\right],$$

$$\text{such that} \quad \mathbb{E}\left[Z^2\right] = \tilde{\sigma}_0^2\left[1 + \tilde{\beta}_0^2 + \tilde{\beta}_0\frac{\varphi(\tilde{\beta}_0)}{\Phi(\tilde{\beta}_0)}\right],$$

and if $Z \sim tr\mathcal{N}(\tilde{\mu}_k, \tilde{\sigma}_k^2, 0, \infty)$:

$$\mathbb{E}\left[Z\right] = \tilde{\mu}_k + \frac{\varphi(\tilde{\beta}_k)}{\Phi(-\tilde{\beta}_k)}\tilde{\sigma}_k\,, \quad \text{and} \quad \mathbb{V}\text{ar}\left(Z\right) = \tilde{\sigma}_k^2\left[1 + \frac{\tilde{\beta}_k\varphi(\tilde{\beta}_k)}{\Phi(-\tilde{\beta}_k)} - \left(\frac{\varphi(\tilde{\beta}_k)}{\Phi(-\tilde{\beta}_k)}\right)^2\right],$$

$$\text{such that} \quad \mathbb{E}\left[Z^2\right] = \tilde{\sigma}_k^2\left[1 + \tilde{\beta}_k^2 - \tilde{\beta}_k\frac{\varphi(\tilde{\beta}_k)}{\Phi(-\tilde{\beta}_k)}\right].$$

E.1. Normalizing constant

We can compute the normalizing constant C using (E.1) evaluated in $n = 0$. The middle integral on the right-hand side can be approximated numerically, as discussed above. Using (E.2), it is clear that:

$$\int_{-\infty}^{t_1} \exp(-H(\varepsilon))\mathrm{d}\varepsilon = \begin{cases} \tilde{\sigma}_0 \frac{\Phi(\tilde{\beta}_0)}{\varphi(\tilde{\beta}_0)} \exp(-e_0)\,, & \text{if } b_0 > 0\,, \\ -\frac{\exp(-e_0)}{a_0}\,, & \text{if } b_0 = 0 \text{ and } a_0 < 0\,, \end{cases}$$

$$\text{and} \qquad \int_{t_k}^{\infty} \exp(-H(\varepsilon))\mathrm{d}\varepsilon = \begin{cases} \tilde{\sigma}_k \frac{\Phi(-\tilde{\beta}_k)}{\varphi(\tilde{\beta}_k)} \exp(-e_k)\,, & \text{if } b_k > 0\,, \\ \frac{\exp(-e_k)}{a_k}\,, & \text{if } b_k = 0 \text{ and } a_k > 0\,. \end{cases}$$

E.2. Expectation

Let us derive the mean μ of ε_t for a given spline density:

$$\mu = \mathbb{E}[\varepsilon_t] = \frac{1}{C} \int_{-\infty}^{\infty} \varepsilon \ \exp(-H(\varepsilon; \boldsymbol{t}, \boldsymbol{y}))\mathrm{d}\varepsilon\,,$$

where C denotes the normalizing constant. Based on (E.2) for $n = 1$ and the form of the expectation of a truncated normal distribution, we have that:

$$\int_{-\infty}^{t_1} \varepsilon \exp(-H(\varepsilon))\mathrm{d}\varepsilon = \begin{cases} \exp(-e_0)\tilde{\sigma}_0 \left[\frac{\Phi(\tilde{\beta}_0)}{\varphi(\tilde{\beta}_0)}(t_1 + \tilde{\mu}_0) - \tilde{\sigma}_0\right]\,, & \text{if } b_0 > 0 \\ -\frac{\exp(-e_0)}{a_0}(t_1 + a_0^{-1})\,, & \text{if } b_0 = 0 \text{ and } a_0 < 0\,, \end{cases}$$

$$\text{and} \qquad \int_{t_k}^{\infty} \varepsilon \exp(-H(\varepsilon))\mathrm{d}\varepsilon = \begin{cases} \exp(-e_k)\tilde{\sigma}_k \left[\frac{\Phi(-\tilde{\beta}_k)}{\varphi(\tilde{\beta}_k)}(t_k + \tilde{\mu}_k) + \tilde{\sigma}_k\right]\,, & \text{if } b_k > 0\,, \\ \frac{\exp(-e_k)}{a_k}(t_k + a_k^{-1})\,, & \text{if } b_k = 0 \text{ and } a_k > 0\,. \end{cases}$$

E.3. Variance

Now let us turn to the calculation of $\mathbb{E}[\varepsilon_t^2]$, such that we can calculate the variance of ε_t.

$$\mathbb{E}[\varepsilon_t^2] = \frac{1}{C} \int_{-\infty}^{\infty} \varepsilon^2 \ \exp(-H(\varepsilon; \boldsymbol{t}, \boldsymbol{y}))\mathrm{d}\varepsilon\,.$$

From (E.2) for $n = 2$ it follows that:

$$\int_{-\infty}^{t_1} \varepsilon^2 \exp(-H(\varepsilon))\mathrm{d}\varepsilon = \begin{cases} \exp(-e_0)\tilde{\sigma}_0 \left[\frac{\Phi(\tilde{\beta}_0)}{\varphi(\tilde{\beta}_0)}\left([t_1 + \tilde{\mu}_0]^2 + \tilde{\sigma}_0^2\right) - \tilde{\sigma}_0\left(\tilde{\mu}_0 + 2t_1\right)\right]\,, & \\ & \text{if } b_0 > 0, \\ -\frac{\exp(-e_0)}{a_0}\left[t_1^2 + t_1\frac{2}{a_0} + \frac{2}{a_0^2}\right]\,, & \text{if } b_0 = 0 \text{ and } a_0 < 0\,, \end{cases}$$

$$\int_{t_k}^{\infty} \varepsilon^2 \exp(-H(\varepsilon)) \mathrm{d}\varepsilon = \begin{cases} \exp(-e_k)\tilde{\sigma}_k \left[\frac{\Phi(-\tilde{\beta}_k)}{\varphi(\tilde{\beta}_k)} \left([t_k + \tilde{\mu}_k]^2 + \tilde{\sigma}_k^2 \right) + \tilde{\sigma}_k \left(\tilde{\mu}_k + 2t_k \right) \right], & \\ & \text{if } b_k > 0, \\ \frac{\exp(-e_k)}{a_k} \left[t_k^2 + t_k \frac{2}{a_k} + \frac{2}{a_k^2} \right], & \text{if } b_k = 0 \text{ and } a_k > 0\,, \end{cases}$$

Based on the values of $\mu = \mathbb{E}[\varepsilon_t]$ and $\mathbb{E}[\varepsilon_t^2]$, we can calculate $\sigma^2 = \mathbb{V}\mathrm{ar}(\varepsilon_t) = \mathbb{E}[\varepsilon_t^2] - \mu^2$.