Journal of Time Series Analysis

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Empirical-Process Limit Theory and Filter Approximation Bounds for Score-Driven Time Series Models

Enzo D'Innocenzo

Department of Economics, Università di Bologna, Bologna, Italy

Correspondence: Enzo D'Innocenzo (enzo.dinnocenzo2@unibo.it)

Received: 16 February 2025 | **Revised:** 21 November 2025 | **Accepted:** 15 December 2025

Keywords: approximation error | empirical processes | goodness-of-fit problem | invertible models | Orlicz norm | score-driven models

ABSTRACT

This article examines the filtering and approximation-theoretic properties of score-driven time series models. Under specific Lipschitz-type and tail conditions, new results are derived, leading to maximal and deviation inequalities for the filtering approximation error using empirical process theory. This approach allows the study of the asymptotic behavior of the empirical distribution function and empirical process of the approximated noise, extending the results of Francq and Zakoïan (2022) for generalized autoregressive conditional heteroskedasticity models. For general score-driven models, however, it is proven that the asymptotic distribution of the empirical process of the approximated noise is model-dependent and influenced by the estimation of model parameters. This contrasts with well-known results for linear and some nonlinear time series models, and is mainly due to the fact that the finite-dimensional static model parameters affect both the noise density and the score-driven process. The goodness-of-fit problem is then considered, and an application of these results is demonstrated with the Beta-t-GARCH(1,1) model, a popular score-driven time series model.

JEL Classification: C10, C12, C13, C22

MSC2020 Classification: 60, 62

1 | Introduction

Score-driven models are a broad class of nonlinear time series models that have proven highly effective in modeling parametric distributions with time-varying parameters. Since their introduction by Creal et al. (2011, 2013) and Harvey (2013), these models have gained widespread popularity, with applications extending beyond economics and finance to diverse fields such as neuroscience and public health.[1]

At their core, score-driven models leverage the score of the conditional observation density, that is, the first derivative of the log-likelihood function with respect to the parameter of interest, to drive the evolution of the model's parameters. From a theoretical standpoint, Harvey (2013) laid the foundation for the asymptotic theory of these models in the context of time-varying location and scale models with heavy-tailed distributions. Later, Blasques et al. (2022) extended this framework to a more general class of univariate score-driven models.

Further theoretical advancements include the work of Blasques et al. (2015), who explored the information-theoretic properties of score-driven models through the Kullback-Leibler (KL) divergence. Other optimality results appear in Blasques et al. (2019), with more recent contributions by Gorgi et al. (2024) and Beutner et al. (2023). The stochastic properties of score-driven models as data-generating processes were first studied by Blasques et al. (2014), while Blasques et al. (2018) established conditions for the existence of moments and invertibility of the filtering equations. Both works build on the theory of stochastic recurrence equations (SRE), leveraging Lipschitz and contraction conditions from Bougerol (1993) to ensure strict stationarity and ergodicity of score-driven processes, as well as filter invertibility. See also Straumann and Mikosch (2006) for a discussion of these topics in the context of nonlinear GARCH models.

This article makes two primary contributions to the literature on score-driven models. First, I establish new theoretical results based on a novel set of Lipschitz-type and tail conditions that refine control of the (potentially highly nonlinear) filter recursion. Unlike earlier studies, the analysis is formulated in Orlicz norms rather than classical moment bounds, yielding greater flexibility. Crucially, an entropy-driven invertibility argument combined with an explicit uniform deviation bound ensures that all subsequent results hold under sub-exponential, or, more generally, any finite $p > 2$ noise tails, and without the global one-step contraction imposed in Assumption 4.4-(*ii*) of Blasques et al. (2022). The overall strategy is reminiscent of the weak-dependence framework developed by Doukhan and Wintenberger (2008) for Markov chains with infinite memory.

A key advantage of this approach is its ability to capture the tail behavior of score-driven models precisely. This leads to sharp maximal and deviation inequalities, which in turn allow for a rigorous analysis of filtering approximation errors. These results facilitate the study of the asymptotic properties of the empirical distribution function and the empirical process of the approximated noise, drawing on empirical process theory (see Dehling et al. (2002) and the recent monograph by van der Vaart and Wellner (2023)). To the best of my knowledge, these techniques have not been applied in the score-driven literature before.

As a second contribution, I derive an explicit formula for the variance of the empirical process of the approximated noise, revealing an important property of score-driven models. Specifically, I show that the asymptotic distribution of this empirical process is both asymptotically model dependent (AMD), as it depends on the distribution of the unobserved noise process and the true model parameter vector, and asymptotically estimation dependent (AED), as it is influenced by parameter estimation. The AMD property arises because the finite-dimensional model parameters directly affect the noise density, shaping the dynamics of the score-driven process. To my knowledge, this is a novel result that provides new insights into the intricate structure of score-driven models.

Similar results have appeared in the literature. Berkes and Horváth (2003) obtained a functional central-limit theorem for the empirical process of squared pseudo-residuals in GARCH models, and related empirical-process and strong-approximation results for ARCH/GARCH and augmented GARCH sequences can be found in Hörmann (2008), Berkes et al. (2008), Berkes et al. (2009), and Berkes et al. (2011). Francq and Zakoïan (2022) extended that analysis to residuals of general semi-parametric conditional location-scale models. Under their baseline assumptions for pure conditional scale models, the residual empirical process is asymptotically model free (AMF) and AED, while an AMD component appears only when the extra stability-by-scaling condition fails or when a conditional mean is introduced. In the fully parametric score-driven framework studied here, the same parameter vector governs both the noise density and the filter recursion, so the empirical process is AMD and AED for all parameter values. The result therefore complements that of Francq and Zakoïan (2022). Remark 6 gives the full assumption-by-assumption comparison, and Section 4.1 shows how the two theories coincide in Gaussian GARCH(1,1).

The AMD and AED properties also contradict earlier results for classical linear time series models. For example, Boldin (1983) and Koul (1992) showed that the empirical process of residuals in autoregressive (AR) and moving-average (MA) models is not AMD. Consequently, standard asymptotic results for goodness-of-fit testing, such as those based on the Kolmogorov–Smirnov statistic, are inapplicable to general score-driven models, as the test's null distribution is neither distribution-free nor estimation-free.

In nonlinear time series models, Koul and Ling (2006) circumvented the AMD issue for ARMA models with GARCH noise process by introducing weighted residual empirical processes. However, this approach does not extend to general score-driven models, where the asymptotic structure of the empirical process is significantly more complex. Nevertheless, by leveraging filtering approximation errors and the asymptotic behavior of the empirical distribution function, I propose a potential solution to the goodness-of-fit testing problem.

The remainder of this article is organized as follows. In Section 2, I introduce the general score-driven time series model and discuss the related problem of filter invertibility and filtering approximation error. In Section 3, I derive maximal and deviation inequalities for the general invertible score-driven filter that characterize the rate of convergence of the unstable filtering recursion to the corresponding unique stationary solution. Building upon these results, in Section 4 I study the asymptotic behavior of the empirical distribution function and the empirical process of the approximated noise. Here it is also proved that, by making some appropriate distributional assumptions, these asymptotic results collapse to those retrieved by Francq and Zakoïan (2022) for the standard GARCH(1,1) model. Finally, in Section 5 the goodness-of-fit problem is considered, and in Section 6 I consider a popular score-driven time series model, namely, the Beta-*t*-GARCH(1,1) model, to show how the conditions considered in this article can be actually applied in practice. Although the Beta-*t*-GARCH(1,1) model is driven by Student's *t* heavy-tailed noise, the score recursion keeps the one-step prediction error under an exponentially decaying (sub-Gaussian) tail. This yields probability bounds that are much tighter than the polynomial bounds usually obtained for heavy-tailed models, so that all maximal and deviation inequalities, consistency results, and limit theorems derived in the article applies with correspondingly sharper rates. The conclusions are drawn in Section 7. The Supporting Information gathers all the proofs and mathematical details.

2 | The General Score-Driven Model

Given a time series $\{y_t\}_{t\in\mathbb{Z}}$ with associated filtration $\mathcal{Y}_t = \sigma\{y_t, y_{t-1}, y_{t-2}, \dots\}$, the general score-driven model is specified by the following dynamic equations:

$$y_t = g(f_t(\theta_0), \epsilon_t), \quad \epsilon_t \overset{IID}{\sim} p_\epsilon(\epsilon_t, \lambda_0), \tag{1}$$

where $g(\cdot,\cdot)$ is a differentiable link function that is strictly increasing in its second argument, and p_ϵ is the noise density indexed by the static parameter $\lambda_0 \in \Lambda \subset \mathbb{R}^{\dim(\lambda)}$. The parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^{\dim(\theta)}$, where Θ is a compact parameter space. Here, $f_t(\theta_0)$ is a stochastic time-varying parameter that indexes the predictive conditional density p_y given the entire past of the process $\mathcal{Y}_{t-1}$. Throughout the article, I work with univariate $f_t(\theta_0)$ taking values in a space $\mathcal{F} \subset \mathbb{R}$, so that $f_t(\theta_0) \in \mathcal{F}$, and the process satisfies the SRE

$$f_{t+1}(\theta_0) = \phi\big(f_t(\theta_0), y_t, \theta_0\big) := \omega_0 + \alpha_0 s_t(f_t(\theta_0), \lambda_0) + \beta_0 f_t(\theta_0), \tag{2}$$

where the driving force of the process is given by the scaled-score function:

$$s_t(f_t(\theta_0), \lambda_0) = S(f_t(\theta_0), y_t, \lambda_0)\nabla(f_t(\theta_0), y_t, \lambda_0), \tag{3}$$

and where $\theta_0 = (\omega_0, \alpha_0, \beta_0, \lambda_0^\top)^\top$ is the true static parameter vector. The function $\phi(\cdot,\cdot,\cdot)$ is continuous (typically Lipschitz) from $\mathcal{F} \times \mathbb{R} \times \Theta$ into $\mathcal{F}$, differentiable with respect to its first coordinate. The term

$$\nabla(f_t(\theta_0), y_t, \lambda_0) = \frac{\partial \log p_y(y_t | f, \lambda_0, \mathcal{Y}_{t-1})}{\partial f}\Big|_{f=f_t(\theta_0)},$$

represents the score function of the log-conditional density with respect to f, while $S(f_t(\theta_0), y_t, \lambda_0)$ is a univariate scaling function. Together, these terms define the scaled-score function $s_t(f_t(\theta_0), \lambda_0)$, which plays a fundamental role in the model's dynamics.

Remark 1. As seen in the recurrence equations (1) to (3), a key feature of score-driven models is that the density p_ϵ of the noise process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ depends on the unknown parameter vector $\lambda_0 \in \Lambda$. This dependence is crucial because λ_0 also enters the recursion of the stochastic process $\{f_t(\theta_0)\}_{t\in\mathbb{Z}}$ through the scaled-score function $s_t(f_t(\theta_0), \lambda_0)$ in (3).

As noted by Creal et al. (2011), Harvey (2013), and Blasques et al. (2014), many time series models, including conditional location-scale and volatility models, can be formulated within the score-driven framework. This includes standard ARMA-GARCH models. For example, a location model can be specified by setting

$$g(f_t(\theta_0), \epsilon_t) = f_t(\theta_0) + \epsilon_t, \tag{4}$$

where $f_t(\theta_0)$ represents a stochastic conditional location (or mean), such that $\mathcal{F} = \mathbb{R}$, and the noise term ϵ_t follows a normal, Student's *t*, or skewed distribution. Similarly, a conditional scale (or volatility) model can be obtained by defining

$$g(f_t(\theta_0), \epsilon_t) = \sqrt{f_t(\theta_0)}\epsilon_t, \tag{5}$$

such that $\mathcal{F} = [\underline{\omega}_0, \infty)$, where $\underline{\omega}_0 > 0$ is the lower bound of the conditional variance process, see Section 4.1 for details on the classical GARCH(1,1) model and Section 6 for the Beta-*t*-GARCH(1,1) model. Furthermore, the generality of the score-driven framework enables researchers to construct nonlinear time series models for higher order moments, such as the stochastic tail-shape model proposed by D'Innocenzo et al. (2023) and D'Innocenzo et al. (2024a), which in this case the model becomes

$$g(f_t(\theta_0), \epsilon_t) = e^{f_t(\theta_0)\epsilon_t} - 1, \tag{6}$$

with $\mathcal{F} = [\underline{\omega}_0, \infty)$, where, similarly to the conditional scale model in (5), $\underline{\omega}_0 > 0$ is the lower bound of the conditional tail process, and the noise term ϵ_t has a standard exponential distribution with unit mean. Other examples of popular score-driven models are listed in Table 1.

2.1 | The Filtering Recursions and Invertibility

In practice, the time-varying parameter $\{f_t(\theta_0)\}_{t\in\mathbb{Z}}$ is not observable, and thus, one must approximate its time evolution from real data. Here, this approximation is denoted with $\{\hat{f}_t(\theta_0)\}_{t\in\mathbb{N}}$. Moreover, the real true parameter vector θ_0 is generally unknown, and so for practical purposes, one needs to replace it with some estimator.

To fix the idea, consider a set of observations $\{y_t\}_{t=1}^T$ and a vector of static parameters $\theta \in \Theta$. From the observation-driven nature of the model given in equations (1) to (3) it is clear that for any initial condition $\hat{f}_1(\theta) = f_1$, the sequence $\hat{f}_t(\theta)$ can be propagated iteratively through $\hat{f}_{t+1}(\theta)$ for $t = 1, \dots, T$ using the following recursive equations:

$$\hat{f}_{t+1}(\theta) = \phi\big(\hat{f}_t(\theta), y_t, \theta\big) := \omega + \alpha s_t(\hat{f}_t(\theta), \lambda) + \beta \hat{f}_t(\theta), \tag{7}$$

$$s_t(\hat{f}_t(\theta), \lambda) = S(\hat{f}_t(\theta), y_t, \lambda)\nabla(\hat{f}_t(\theta), y_t, \lambda), \tag{8}$$

with

$$\nabla(\hat{f}_t(\theta), y_t, \lambda) = \frac{\partial \log p_y(y_t | f, \lambda, \mathcal{Y}_{t-1})}{\partial f}\Big|_{f=\hat{f}_t(\theta)},$$

and a univariate scaling function $S(\hat{f}_t(\theta), y_t, \lambda)$.

If for each $f_1 \in \mathcal{F}$ this recursive procedure employed to approximate $f_t(\theta)$ for $t = 1, \dots, T$ converges to a unique stationary and ergodic solution $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ as $t \to \infty$, then the general score-driven model is said to be invertible. More formally, under suitable conditions, the invertibility property ensures that

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

in probability, where $\xrightarrow{e.a.s.}$ stands for *exponentially fast almost sure convergence*, see Straumann and Mikosch (2006) for a detailed explanation of this mode of convergence. Additionally, using equation (1), the invertibility property further allows researchers to approximate the unobserved noise sequence ϵ_t in terms of observed values. That is, one can construct a process $\{\epsilon_t(\theta)\}_{t\in\mathbb{Z}}$ which is adapted to the filtration generated by $\{y_t, y_{t-1}, y_{t-2}, \dots\}$ and depends on $\theta \in \Theta$.

Since the unobserved $f_t(\theta)$ must be approximated using $\hat{f}_t(\theta)$, which depends on the chosen initial condition f_1, one can only

TABLE 1 | Links, inverses, and derivative terms $(\dot{g},\ \nabla f_t/\dot{g})$ that enter the score recursion for six standard score-driven specifications.

Model type	$\boldsymbol{g(f_t(\theta_0), \epsilon_t)}$	$\boldsymbol{g^{-1}(f_t(\theta), y_t)}$	$\boldsymbol{\dot{g}(f_t(\theta), \epsilon_t)}$	$\boldsymbol{\frac{\nabla f_t(\theta)}{\dot{g}(f_t(\theta),\epsilon_t)}}$
Location	$f_t(\theta_0) + \epsilon_t$	$y_t - f_t(\theta)$	-1	$-\nabla f_t(\theta)$
Logistic location	$\frac{1}{1+e^{-f_t(\theta_0)}} + \epsilon_t$	$y_t - \frac{1}{1+e^{-f_t(\theta)}}$	$-\frac{(1+e^{f_t(\theta)})^2}{e^{f_t(\theta)}}$	$-\frac{e^{f_t(\theta)}}{(1+e^{f_t(\theta)})^2}\nabla f_t(\theta)$
Scale	$\sqrt{f_t(\theta_0)}\epsilon_t$	$\frac{y_t}{\sqrt{f_t(\theta)}}$	$-\frac{2}{\epsilon_t}\frac{\sqrt{f_t(\theta)}}{\sqrt{f_t(\theta_0)}}f_t(\theta)$	$-\frac{\epsilon_t}{2}\frac{\sqrt{f_t(\theta_0)}}{\sqrt{f_t(\theta)}}\frac{\nabla f_t(\theta)}{f_t(\theta)}$
Exponential scale	$e^{f_t(\theta_0)}\epsilon_t$	$y_t e^{-f_t(\theta)}$	$-\frac{1}{\epsilon_t}\frac{e^{f_t(\theta)}}{e^{f_t(\theta_0)}}$	$-\epsilon_t\frac{e^{f_t(\theta_0)}}{e^{f_t(\theta)}}\nabla f_t(\theta)$
Duration	$f_t(\theta_0)\epsilon_t$	$\frac{y_t}{f_t(\theta)}$	$-\frac{1}{\epsilon_t}\frac{f_t(\theta)}{f_t(\theta_0)}f_t(\theta)$	$-\epsilon_t\frac{f_t(\theta_0)}{f_t(\theta)}\frac{\nabla f_t(\theta)}{f_t(\theta)}$
Tail	$e^{f_t(\theta_0)\epsilon_t} - 1$	$\frac{\ln(1+y_t)}{f_t(\theta)}$	$-\frac{1}{\epsilon_t}\frac{f_t(\theta)}{f_t(\theta_0)}f_t(\theta)$	$-\epsilon_t\frac{f_t(\theta_0)}{f_t(\theta)}\frac{\nabla f_t(\theta)}{f_t(\theta)}$

obtain an approximate noise process $\{\hat{\epsilon}_t(\theta)\}_{t=1}^{T}$. This process is obtained by substituting $f_t(\theta)$ with $\hat{f}_t(\theta)$ in equation (1):

$$\hat{\epsilon}_t(\theta) = g^{-1}(\hat{f}_t(\theta), y_t), \qquad t = 1, \ldots, T, \tag{9}$$

where $g^{-1}(\cdot,\cdot)$ denotes the inverse function of $g(\cdot,\cdot)$ defined in equation (1) with respect to its second argument. As noted by Granger and Andersen (1978), the invertibility property ensures that, for each $f_1 \in \mathcal{F}$, the unobserved noise $\{\epsilon_t(\theta)\}_{t\in\mathbb{Z}}$ can be consistently estimated, meaning that

$$\sup_{\theta\in\Theta}\left|\hat{\epsilon}_t(\theta) - \epsilon_t(\theta)\right| \xrightarrow{e.a.s.} 0, \quad \text{as} \quad t \to \infty,$$

where

$$\epsilon_t(\theta) = g^{-1}(f_t(\theta), y_t), \qquad t \in \mathbb{Z},$$

is strictly stationary and ergodic and does not depend on the chosen initial condition. This property is crucial for nonlinear dynamic models, as it allows researchers to represent the process $\{\epsilon_t(\theta)\}_{t\in\mathbb{Z}}$ as a function of the data $\{y_t\}_{t\in\mathbb{Z}}$.

Remark 2. It is crucial to remark that in general it holds that $\epsilon_t(\theta) \neq \epsilon_t$ for all t, unless $\theta = \theta_0$ and the postulated score-driven model described is correctly specified, that is, the process $\{y_t\}_{t\in\mathbb{Z}}$ is generated by the model in equations (1) to (3).

Sufficient conditions for the invertibility of the filter in (7) are discussed in Blasques et al. (2018). In particular, Proposition 3.1 of Blasques et al. (2018) provides conditions that ensure the continuous invertibility of the filter $\{\hat{f}_t(\theta)\}_{t\in\mathbb{N}}$, a concept originally introduced by Wintenberger (2013). As in Wintenberger (2013), Blasques et al. (2018) rely on Theorem 3.1 of Bougerol (1993). Specifically, Proposition 3.1 of Blasques et al. (2018) establishes the following sufficient conditions for the invertibility of the filter $\{\hat{f}_t(\theta)\}_{t\in\mathbb{N}}$.

Assumption 1. For the score-driven model given in equations (1) to (3), it holds that:

(*i*) The process $\{y_t\}_{t\in\mathbb{Z}}$ is a stationary and ergodic sequence of random variables.

(*ii*) For any $\theta \in \Theta$, there exists $\overline{f} \in \mathcal{F}$ such that

$$\mathbb{E}\left[\log^{+}\sup_{\theta\in\Theta}\left|\phi\left(\overline{f}, y_t, \theta\right) - \overline{f}\right|\right] < \infty.$$

(*iii*) It holds that $\log\sup_{f\in\mathcal{F}}\left|\dot{\phi}(f, y_0, \theta)\right|$ is a.s. continuous on Θ and that

$$\mathbb{E}\left[\log^{+}\sup_{\theta\in\Theta}\sup_{f\in\mathscr{F}}\left|\dot{\phi}(f, y_t, \theta)\right|\right] < \infty, \quad \text{and}$$

$$\mathbb{E}\left[\log\sup_{f\in\mathscr{F}}\left|\dot{\phi}(f, y_0, \theta)\right|\right] < 0 \quad \text{for any} \quad \theta \in \Theta,$$

where $\dot{\phi}(f, y_t, \theta) := \partial\phi(f, y_t, \theta)/\partial f$ is defined as the stochastic Lipschitz coefficient.

From equations (7) and (8), it is easy to see that the stochastic Lipschitz coefficient defined in point (*iii*) of Assumption 1 can be explicitly expressed as

$$\dot{\phi}(f, y_t, \theta) := \beta + \alpha\frac{\partial s_t(f, \lambda)}{\partial f}.$$

Remark 3. Assumption 1 is directly analogous to the SRE conditions used in the nonlinear GARCH literature. For a model of the form $y_t = \sigma_t(\theta)\epsilon_t$ with variance recursion $\sigma_{t+1}^2(\theta) = \varphi(\sigma_t^2(\theta), y_t, \theta)$ for some (nonlinear) function $\varphi(\cdot,\cdot,\cdot)$, Straumann and Mikosch (2006) define the random Lipschitz coefficient $\Lambda_t(\theta) := \sup_{x_1\neq x_2}\frac{|\varphi(x_1,y_t,\theta)-\varphi(x_2,y_t,\theta)|}{|x_1-x_2|}$, and impose Bougerol-type log-moment and contraction conditions such as $\mathbb{E}\left[\log^{+}|\sigma_0^2|\right] < \infty$ and $\mathbb{E}\left[\log|\Lambda_0(\theta_0)|\right] < 0$ (together with uniform variants) to obtain existence and invertibility of the volatility recursion. In the score-driven recursion $f_{t+1}(\theta) = \phi(f_t(\theta), y_t, \theta)$ considered here, the quantity $\sup_{f\in\mathcal{F}}\left|\dot{\phi}(f, y_t, \theta)\right|$ plays the analogous role of a random Lipschitz coefficient, and point (iii) of Assumption 1 imposes the corresponding log-moment and contraction conditions via $\mathbb{E}\left[\log\sup_{f\in\mathcal{F}}\left|\dot{\phi}(f, y_0, \theta_0)\right|\right] < 0$. Assumption 1 can therefore be viewed as the score-driven analogue of the Lipschitz and log-moments and contraction assumptions for nonlinear GARCH models in Straumann and Mikosch (2006).

Due to the potentially high degree of nonlinearity in the dynamic model under investigation, proving invertibility can be highly cumbersome. Typically, very restrictive sufficient conditions are required, which may not hold in practical applications. This is because the parameter space that ensures invertibility may be extremely small or even degenerate. In the latter case, this means that an invertibility region may not exist at all.

To address these issues, I propose a set of additional sufficient conditions that directly relate to the tail behavior of the filter $\hat{f}_t(\theta)$. These conditions ensure the existence of a non-degenerate invertibility region and further establish a sharp rate of convergence to a unique stationary and ergodic solution $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ for general invertible score-driven filters, using maximal inequalities.

3 | Filtering Approximation Error

3.1 | Orlicz Spaces and Entropy With Bracketing

Let ψ be a convex, nondecreasing, nonzero function on $[0,\infty)$, with $\psi(0)=0$ (an Orlicz function). The Orlicz norm of the random variable X is defined as

$$\|X\|_{\psi} := \inf\left\{K>0 : \mathbb{E}\left[\psi\left(\frac{|X|}{K}\right)\right]\leq 1\right\},$$

if the expectation is finite. For example, when $\psi(x)=x^p$, with $p\geq 1$, the Orlicz norm $\|X\|_{\psi}$ coincides with the standard $L_p(\mathbb{P})$ norm, that is, $\|X\|_p := (\mathbb{E}[|X|^p])^{1/p}$. In what follows, I restrict attention to Orlicz norms corresponding to the exponential-type functions:

$$\psi_p := e^{x^p}-1, \quad p\geq 1. \tag{10}$$

For a given separable stochastic process $\{X_t\}_{t\in T}$ and a fixed Orlicz function ψ_p, I define the pseudo-metric:

$$\rho_X^p(t,s) := \|X_t - X_s\|_{\psi_p}, \quad \forall t,s.$$

The transformation in (10) is particularly useful for score-driven models. As shown in Lemma 8.1 of Kosorok (2008), it provides necessary and sufficient conditions for the ψ_p-Orlicz norm to be finite based on the tail behavior of stochastic processes. Specifically, $\|X_t\|_{\psi_2}<\infty$, corresponds to sub-Gaussian tails, while $\|X_t\|_{\psi_1}<\infty$, corresponds to sub-exponential tails. For further details, I refer to Chapter 2 of Boucheron et al. (2013) or section 2.2 of van der Vaart and Wellner (2023).

Now, consider the function $\xi_{f,\theta}$ defined as

$$\xi_{f,\theta}(y_t) := \dot{\phi}(f,y_t,\theta) = \beta + \alpha\frac{\partial s_t(f,\lambda)}{\partial f}, \tag{11}$$

such that $\xi_{f,\theta}\in\Xi_\theta$, and define

$$X_t(\xi_{f,\theta}) := \xi_{f,\theta}(y_t), \tag{12}$$

where, for all $(f,\theta)\in\mathcal{F}\times\Theta$, the map $\xi_{f,\theta}:\mathbb{R}\mapsto\mathbb{R}$ is deterministic, whereas $\xi_{f,\theta}(y_t)$ is the corresponding real-valued random variable obtained by evaluating $\xi_{f,\theta}$ at y_t. This notation clarifies that I focus on the stochastic process $\{X_t(\xi_{f,\theta})\}_{\xi_{f,\theta}\in\Xi_\theta}$, indexed by the functions $\xi_{f,\theta}\in\Xi_\theta$. I derive a maximal inequality to obtain a sharp upper bound on the expected supremum of this process.

To control this supremum, I measure the size of the index class Ξ_θ via its covering number $N(\varepsilon,\Xi_\theta,\rho_X^p)$ under the pseudo-metric

$$\rho_X^p(f,f') := \left\|X_t(\xi_{f,\theta}) - X_t(\xi_{f',\theta})\right\|_{\psi_p},$$

for every $(f,f')\in\mathcal{F}\times\mathcal{F}$. The covering number $N(\varepsilon,\Xi_\theta,\rho_X^p)$ represents the minimal number of balls $B(f,\varepsilon)$ of radius $\varepsilon>0$ needed to cover Ξ_θ. Its logarithm, that is, $\log N(\varepsilon,\Xi_\theta,\rho_X^p)$, is the (metric) entropy. Similarly, the bracketing number $N_{[]}(\varepsilon,\Xi_\theta,\rho_X^p)$ represents the minimal number of brackets of diameter at most ε required to cover Ξ_θ, and $\log N_{[]}(\varepsilon,\Xi_\theta,\rho_X^p)$ is the bracket entropy (see van der Vaart and Wellner (2023), 147).

3.2 | Maximal and Deviation Inequalities

Consider the score-driven model as defined in equations (1) to (3), and the Orlicz function ψ_p as given in (10) for $p\geq 1$. I begin by deriving a maximal inequality for the stochastic process $\{X_t(\xi_{f,\theta})\}_{\xi_{f,\theta}\in\Xi_\theta}$. For this purpose, I assume the existence of a bracketing function.

Formally, for $\delta>0$ define

$$\begin{aligned} X_t^\star(\xi_{f,\theta}) &:= \sup_{\rho_X^p(f,f')\leq\delta} X_t(\xi_{f',\theta}),\\ X_{t\star}(\xi_{f,\theta}) &:= \inf_{\rho_X^p(f,f')\leq\delta} X_t(\xi_{f',\theta}), \end{aligned}$$

where $X_{t\star}(\xi_{f,\theta}) < X_t^\star(\xi_{f,\theta})$, $\forall (f,f')\in\mathcal{F}\times\mathcal{F}$. Suppose further that, for some $p\geq 1$

$$X_{t\star}(\xi_{f,\theta}) \leq X_t(\xi_{f,\theta}) \leq X_t^\star(\xi_{f,\theta}), \quad \forall t\geq 1.$$

I now introduce the following lemma, which applies to several classes of score-driven models. Examples include the Beta-t-GARCH model of Harvey (2013) and the fat-tailed dynamic location model of Harvey and Luati (2014), both of which are extensively discussed in Blasques et al. (2018).

Lemma 1. *For any $f_0\in\mathcal{F}$ and every $\theta\in\Theta$, it holds that*

$$\left\|\sup_{\xi_{f,\theta}\in\Xi_\theta}\left|X_t(\xi_{f,\theta})\right|\right\|_{\psi_p} \leq L(\xi_{f_0,\theta},\eta), \quad \text{where}$$

$$L(\xi_{f_0,\theta},\eta) := \left\|X_t(\xi_{f_0,\theta})\right\|_{\psi_p} + \int_0^\eta \left(\log\left(1+N_{[]}(\varepsilon,\Xi_\theta,\rho_X^p)\right)\right)^{1/p} \mathrm{d}\varepsilon, \tag{13}$$

with $\eta := diam(\Xi_\theta) = \left\|\sup_{\xi_{f,\theta}\in\Xi_\theta}\left|X_t^\star(\xi_{f,\theta}) - X_{t\star}(\xi_{f,\theta})\right|\right\|_{\psi_p}.$

Remark 4. Lemma 1 shows that a classical chaining argument, applied to the *random* function class Ξ_θ, delivers a finite Orlicz norm for $\sup_{\xi_{f,\theta}\in\Xi_\theta}|X_t(\xi)|$ whenever the entropy integral in (13) is bounded. Consequently, I obtain a non-degenerate invertibility region without imposing the uniform one-step contraction required in Assumption 4.4-(ii) of Blasques et al. (2022). To the best of my knowledge, the standard empirical-process literature—for example, van der Vaart and Wellner (2023); Rio (2018)—treats only deterministic or weakly dependent index sets, whereas here the index $f_t(\theta)$ is itself a random iterated map. The combination of Orlicz-tail control with an entropy bound therefore provides a novel route to invertibility for heavy-tailed score-driven filters.

Using this result, I now state the following theorem, which establishes sharp maximal and deviation inequalities for invertible score-driven filters. For simplicity of notation, I define $L_i(\xi) := L\big(\xi_{f_0,\theta},\eta\big)^i$ with $i \in \mathbb{N}$, where $L_1(\xi) = L\big(\xi_{f_0,\theta},\eta\big)$ is given in (13).

Theorem 1. *Consider the score-driven model as given in equations (1) to (3) and let Assumption 1 holds true. Furthermore, assume that*:

(*i*) *For any $f_1, f_0 \in \mathcal{F}$, it holds that*

$$\left\| \sup_{\theta\in\Theta} \left| f_1 - \phi\big(f_0, y_t, \theta\big) \right| \right\|_{\psi_p} \leq \tau < \infty, \quad p \geq 1.$$

(*ii*) *The filter process $\{\hat{f}_t(\theta)\}_{t\in\mathbb{N}}$ initialized at some non-random $f_1 \in \mathcal{F}$ satisfies*

$$\left\| \sup_{\theta\in\Theta} \left| \phi\big(\hat{f}_t(\theta), y_t, \theta\big) - \phi\big(f_t(\theta), y_t, \theta\big) \right| \right\|_{\psi_p} \leq L_1(\xi) \left\| \sup_{\theta\in\Theta} \left| \hat{f}_t(\theta) - f_t(\theta) \right| \right\|_{\psi_p}, \quad p \geq 1,$$

where for any $f_0 \in \mathcal{F}$ and every $\theta \in \Theta$, it holds that $L_1(\xi) < 1$.

Then

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \left| \hat{f}_{t+1}(\theta) - f_{t+1}(\theta) \right| \right] \leq \tau\big(\log\big(1 + L_t(\xi)\tau\big)\big)^{1/p} \to 0, \quad t \to \infty, \tag{14}$$

and consequently, for any $z \geq 0$

$$\mathbb{P}\left(\sup_{\theta\in\Theta} \left| \hat{f}_{t+1}(\theta) - f_{t+1}(\theta) \right| \geq z \right) \leq 2\exp\left\{ -\frac{z^2}{2L_t(\xi)\tau\big(L_t(\xi)\tau + z\big)} \right\}. \tag{15}$$

The inequality obtained in (15) resembles the classical Bernstein inequality for independent random variables (see van der Vaart and Wellner (2023), 152, 153). A similar result has also been established by Dedecker and Fan (2015) for separately Lipschitz functionals of iterated random functions.

This upper bound is sub-Gaussian for small values of z and of sub-exponential type for large values of z. Therefore, using the approximation function of van der Vaart and Wellner (2023), 153, I derive the following corollary.

Corollary 1.1. *Let the conditions of Theorem 1 hold true.*

Then, for any $u > 0$

$$\mathbb{P}\left(\sup_{\theta\in\Theta} \left| \hat{f}_{t+1}(\theta) - f_{t+1}(\theta) \right| \geq 2L_t(\xi)\tau(\sqrt{u} + u) \right) \leq 2e^{-u}.$$

Therefore, choosing $u = c\log t$ for any constant $c > 1$ gives

$$\sup_{\theta\in\Theta} \left| \hat{f}_{t+1}(\theta) - f_{t+1}(\theta) \right| = O_{a.s.}\Big(L_t(\xi)\tau(\sqrt{\log t} + \log t) \Big).$$

To conclude this section, I provide some comments on conditions (i) and (ii) imposed in Theorem 1.

First, condition (i) requires exponential control of the recurrence filtering equation under analysis, uniformly over Θ. This holds for all score-driven models with bounded scaled-score functions $s_t(f_t(\theta), \lambda)$ that have bounded derivatives. For instance, Harvey (2013) showed that score-driven models with heavy-tailed noise densities p_ϵ in (1) inherit scaled-score functions $s_t(f_t(\theta), \lambda)$ with sub-exponential tails and consequently bounded unconditional moments.

Second, condition (ii) is essentially a one-step contraction condition, which generalizes condition (iii) in Assumption 1 by also ensuring the boundedness of moments for $\{f_t(\theta)\}_{t\in\mathbb{Z}}$. Additionally, it is important to note that this condition is less restrictive than condition (iv) in Proposition 3.2 (or equivalently Assumption 4.4–(ii)) imposed by Blasques et al. (2022), which enforces a uniform one-step contraction condition over $\mathcal{F} \times \mathbb{R} \times \Theta$.

Condition (ii) is automatically met by score-driven models whose scaled score is bounded, and together with condition (i) forms the core of Theorem 1. These two conditions are stated in general terms; each model must be checked against them. In Section 6, I verify both (i) and (ii) explicitly for the Beta-*t*-GARCH(1,1) specification, providing a fully worked example in which Theorem 1, and all subsequent results that rely on it, hold without any additional assumptions.

4 | The Empirical Process of the Approximate Noise

As discussed in the previous section, the time-varying parameter $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ is not observable. However, given a set of observations $\{y_t\}_{t=1}^T$, it can be sufficiently well-approximated by $\{\hat{f}_t(\theta)\}_{t\in\mathbb{N}}$ uniformly over Θ, provided that the invertibility property holds. In practice, however, the parameter vector θ, which lies in the compact set $\Theta \subset \mathbb{R}^{\dim(\theta)}$, is typically unknown and must be estimated. In score-driven models, this estimation is conveniently performed using the maximum likelihood (ML) method, denoted here by $\hat{\theta}_T$. The likelihood function is available in closed form via a prediction error decomposition, making ML estimation straightforward.

As a result, the approximate noise process $\{\hat{\epsilon}_t(\theta)\}_{t=1}^T$ defined in (9) becomes

$$\hat{\epsilon}_t(\hat{\theta}_T) := g^{-1}(\hat{f}_t(\hat{\theta}_T), y_t), \qquad t = 1, \dots, T. \tag{16}$$

In this section, I investigate the asymptotic properties of the approximate empirical distribution function

$$\hat{F}_T(x, \hat{\theta}_T) := \frac{1}{T}\sum_{t=1}^T \mathbb{1}\{\hat{\epsilon}_t(\hat{\theta}_T) \le x\}, \qquad -\infty < x < \infty, \tag{17}$$

and the respective empirical process

$$\hat{\mathcal{W}}_T(x, \hat{\theta}_T) := \sqrt{T}\big(\hat{F}_T(x, \hat{\theta}_T) - F(x)\big), \qquad -\infty < x < \infty, \tag{18}$$

where

$$F(x) := \mathbb{P}(\epsilon_t \le x), \qquad -\infty < x < \infty.$$

Before presenting the Glivenko-Cantelli-type theorem for score-driven filters, I first introduce a preliminary lemma concerning the closeness between the approximate sequence $\{\hat{\epsilon}_t(\hat{\theta}_T)\}_{t=1}^T$ and the unobserved noise process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ as defined in (1). All subsequent results hinge on a careful analysis of the derivative of the link function g in (1) with respect to its first argument:

$$\dot{g}\big(f_t(\theta_0), \epsilon_t\big) := \frac{\partial g\big(f, \epsilon_t\big)}{\partial f}\Big|_{f=f_t(\theta_0)}. \tag{19}$$

Since the approximate filter $\hat{f}_t(\hat{\theta}_T)$ generally differs from the unobserved $f_t(\theta_0)$, the residual approximation error naturally splits into two parts:

$$\hat{\Delta}_t(\hat{\theta}_T) := \hat{\epsilon}_t(\hat{\theta}_T) - \epsilon_t = \big(\hat{\epsilon}_t(\hat{\theta}_T) - \epsilon_t(\hat{\theta}_T)\big) + \big(\epsilon_t(\hat{\theta}_T) - \epsilon_t\big),$$
$$t = 1, \dots, T. \tag{20}$$

The first term $\big(\hat{\epsilon}_t(\hat{\theta}_T) - \epsilon_t(\hat{\theta}_T)\big)$ in (20) is the *pure filter-approximation error*, which persists even if the parameter were known, because $f_t(\theta)$ is replaced by its proxy $\hat{f}_t(\theta)$. The second term $\big(\epsilon_t(\hat{\theta}_T) - \epsilon_t\big)$ in (20) is the *parameter-estimation error*, which vanishes when $\hat{\theta}_T = \theta_0$.

Thus, to control the first term on the rhs of the difference process $\{\hat{\Delta}_t(\hat{\theta}_T)\}_{t=1}^T$ in equation (20), I assume the following:

(B1) The link function g is invertible and continuously differentiable, with derivative denoted by $\dot{g}$. Define

$$M_{1,t} := \sup_{\theta\in\Theta}\left|\frac{1}{\dot{g}\big(f_t(\theta), \epsilon_t\big)}\right|,$$

and assume that $M_{1,t} > 0$ almost surely and $\mathbb{E}\big[\log^+ M_{1,t}\big] < \infty$, where $\log^+ x := \max\{0, \log(x)\}$.

Assumption (B1) stipulates that the link function g is strictly monotone, and therefore invertible, and continuously differentiable. These properties ensure that g^{-1} is single-valued and smooth, while the accompanying log-moment bound controls, on average, the worst-case inverse slope. Consequently, occasional flat regions of g or episodes of heavy-tailed noise do not endanger filter stability or the subsequent limit theory.

Similarly, to control the second term on the rhs of equation (20), I introduce an additional assumption.

(B2) Let

$$M_{2,t} := \sup_{\theta\in V_\delta(\theta_0)}\left\|\frac{\nabla f_t(\theta)}{\dot{g}\big(f_t(\theta), \epsilon_t\big)}\right\|,$$

where $V_\delta(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \le \delta\}$ and $\{\nabla f_t(\theta)\}_{t\in\mathbb{Z}} := \left\{\frac{\partial f_t(\theta)}{\partial\theta}\right\}_{t\in\mathbb{Z}}$ is the strictly stationary and ergodic vector which collects the first derivative process of the stochastic time-varying parameter $f_t(\theta)$.
Assume that either

$$\|M_{2,t}\|_{\psi_p} \le M_2 < \infty \quad \text{for some} \quad p \ge 1, \tag{21}$$

or

$$\|M_{2,t}\|_p \le M_2 < \infty \quad \text{for some} \quad p > 2. \tag{22}$$

In Assumption (B2), $M_{2,t}$ is the largest value that the stochastic factor $\nabla_\theta f_t(\theta)\big/\dot{g}\big(f_t(\theta), \epsilon_t\big)$ can take in a δ-neighborhood of the true parameter. Condition (21) states that this factor has sub-exponential tails, while (22) allows heavier tails but requires a finite pth moment with $p > 2$. Either alternative is strong enough to keep the noise approximation remainder summable and $o(T^{-1/2})$, which is exactly what the subsequent stability and limit arguments rely on. Table 1 lists several widely used link functions in score-driven models that meet Assumptions (B1) and (B2). Rigorous checks for each case are provided in Appendix A.

Strong consistency of the ML estimator for score-driven models can be obtained under five low-level conditions: (i) stationarity and ergodicity with an integrable one-step log-likelihood, (ii) identifiability, (iii) filter invertibility with a contraction and finite log-moments, (iv) a linear envelope that links likelihood and filter errors, and (v) a mild integrability bound that guarantees an upper-semicontinuous limit; see Theorem 4.1 of Blasques et al. (2018). Theorem 4.6 of Blasques et al. (2022) shows that the same set of conditions also suffices under misspecification. Because these requirements are automatically satisfied once Theorem 1, Assumptions (B1) and (B2), and the moment restrictions stated above are in force, I impose strong consistency directly:

(B3) The ML estimator $\hat{\theta}_T$ is strongly consistent for θ_0, that is, $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$ as $T \to \infty$, where $\theta_0 \in \mathrm{int}(\Theta)$, and that there exists a non-stochastic decreasing sequence $\{D(\delta_T)\}_{T\in\mathbb{N}}$ that only depends on δ_T which satisfies $D(\delta_T) \xrightarrow{a.s.} 0$ as $T \to \infty$ with $\lim_{T\to\infty}\delta_T = 0$.

Assumption (B3) also introduces the deterministic sequence $D(\delta_T)$. Its only purpose is to keep track of how the envelope of the derivative factor $M_{2,t}$ in Assumption (B2) behaves on shrinking neighborhoods $V_{\delta_T}(\theta_0)$. The condition $D(\delta_T) \xrightarrow{a.s.} 0$ as $\delta_T \downarrow 0$

ensures that this tail component is asymptotically negligible and does not affect the $\sqrt{T}$-limit in Theorem 3. See also Remark 5 for how $D(\delta_T)$ enters the maximal-deviation bound.

With Assumptions (B1)–(B3) in place, I can now control the gap between the approximate and the true noise sequences. The next lemma provides the key technical bound needed for the Glivenko-Cantelli result later stated in Theorem 2.

Lemma 2. *Let the conditions of Theorem 1, and Assumptions (B1)–(B3) hold true.*

Then, for every $t \geq 1$ and $u > 0$

$$\mathbb{P}\left(\left|\hat{\Delta}_{t+1}(\hat{\theta}_T)\right| \geq 2M_{1,t}L_t(\xi)\tau(\sqrt{u}+u)+t^{\gamma}u^{-1/p}\right) \leq \mathcal{T}_{\gamma}(t,u).$$

where,

$$\left(\gamma, \mathscr{T}_{\gamma}(t,u)\right) = \begin{cases} (0, 2e^{-u}) & \text{if} \quad (\mathbf{B}2)-(21) \quad \text{is satisfied,} \\ \left(1+1/p, 2e^{-u}+CD(\delta_T)^p t^{-(p+1)}u^2\right) & \text{if} \quad (\mathbf{B}2)-(22) \quad \text{is satisfied,} \end{cases}$$

and $C > 0$ does not depend on t or u.

Therefore, choosing $u = c\log t$ for any constant $c > 1$ gives

$$\left|\hat{\Delta}_{t+1}(\hat{\theta}_T)\right| = O_{a.s.}\left(L_t(\xi)\tau(\sqrt{\log t}+\log t)+D(\delta_T)(\log t)^{1/p}\right).$$

Remark 5. Lemma 2 establishes a maximal-deviation inequality that does not rely on the global one-step contraction required by Assumption 4.4(ii) in Blasques et al. (2022). The threshold decomposes into two terms. The filter component, $L_t(\xi)\tau(\sqrt{\log t}+\log t)$, reflects geometric forgetting of the recursion and therefore holds for every parameter value. The tail component adapts to the heaviness of the noise: under the sub-exponential Orlicz condition (21) the deviation probability decays like e^{-u}; under the finite-moment condition (22) with $p > 2$ the deviation event uses the deterministic threshold $t^{\gamma}(\log t)^{-1/p}$ with any $\gamma > 1/p$ (I fix $\gamma = 1 + 1/p$ in the lemma), yielding a summable factor $D(\delta_T)^p(\log t)^2 t^{-(p+1)}$ in the probability bound. As a consequence, the almost sure envelope used downstream can be taken as $L_t(\xi)\tau(\sqrt{\log t}+\log t)+D(\delta_T)(\log t)^{1/p}$, which shrinks because $D(\delta_T) \to 0$ a.s. The bound remains valid for heavy tails as weak as $p > 2$ (e.g., Student's t with $2+\delta$ degrees of freedom or Pareto with index $2+\delta$), and of course covers the Gaussian case. Since the corresponding tail probabilities are summable (for $u = c\log t$), the deviation is $o_{\mathbb{P}}(T^{-1/2})$, so the subsequent limit theorems require no uniform contraction assumption.

I am now ready to introduce the next theorem, where I show that under certain regularity conditions, the approximate empirical distribution function $\hat{F}_T(x,\hat{\theta}_T)$ belongs to the Glivenko-Cantelli class (see section 2.4 of van der Vaart and Wellner (2023)). The key step in the proof consists of establishing that the difference between $\hat{F}_T(x,\hat{\theta}_T)$ and the empirical distribution function of the unobserved noise process $\left\{\epsilon_t\right\}_{t\in\mathbb{Z}}$, given by

$$F_T(x) := \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\left\{\epsilon_t \leq x\right\}, \qquad -\infty < x < \infty, \tag{23}$$

converges to zero uniformly and almost surely. Analogously, I then define the corresponding empirical process as

$$\mathcal{W}_T(x) := \sqrt{T}\left(F_T(x)-F(x)\right), \qquad -\infty < x < \infty. \tag{24}$$

To properly study the asymptotic behavior of the empirical distribution functions in (17) and (23), and the corresponding empirical processes in (18) and (24), respectively, I need the following Lipschitz continuity condition.

(B4) The distribution function F is Lipschitz continuous with K_F Lipschitz constant.

Theorem 2. *Let the conditions of Theorem 1, and Assumptions (B1)–(B4) hold true.*

Then,

$$\sup_{x\in\mathbb{R}}\left|\hat{F}_T(x,\hat{\theta}_T)-F(x)\right| \xrightarrow{a.s.} 0, \qquad T \to \infty.$$

Having established a Glivenko-Cantelli-type theorem for score-driven filters, I now present the final theorem of this Section, in which I derive the asymptotic distribution of the approximate empirical distribution function $\hat{F}_T(x,\hat{\theta}_T)$. To proceed, I impose some additional assumptions. Specifically, I require that the unobserved noise process $\left\{\epsilon_t\right\}_{t\in\mathbb{Z}}$ has a smooth density.

(B5) The noise density p_ϵ is continuous on $\mathbb{R}$ and satisfies $\sup_{x\in\mathbb{R}} p_\epsilon(x,\lambda) = \boldsymbol{B} < \infty, \quad \forall\, \lambda \in \Lambda$.

As noted in Section 2, a key feature of score-driven models is that the density p_ϵ of the unobserved noise process $\left\{\epsilon_t\right\}_{t\in\mathbb{Z}}$ depends directly on the unknown parameter vector $\lambda_0 \in \Lambda$ which also appears in the recursion of the stochastic parameter process $\left\{f(\theta_0)\right\}_{t\in\mathbb{Z}}$ (see Remark 1). To control this parameter variation, I impose the following assumption, which is a p-th order moment dominance condition required for performing Taylor expansions of the distribution function (see the proof of Theorem 3).

(B6) There exist some positive constant $M_3 > 0$ such that

$$\left\|\sup_{\theta\in V_\delta(\theta_0)}\left\|q_\epsilon\left(\epsilon_t(\theta),\lambda\right)\right\|\right\|_p \leq M_3 < \infty,$$

$$q_\epsilon\left(\epsilon_t(\theta),\lambda\right) := p_\epsilon\left(\epsilon_t(\theta),\lambda\right)\frac{\partial}{\partial\theta}\log p_\epsilon\left(\epsilon_t(\theta),\lambda\right), \quad p > 2.$$

It is well-known that ensuring the asymptotic normality of the ML estimator in score-driven models requires additional conditions on both the time-varying parameter $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ and the density p_ϵ. These assumptions further restrict the parameter space Θ to guarantee the existence of at least $2+\delta$ moments (for some $\delta > 0$) and at least $1+\delta$ moments for their first and second derivatives, respectively.

Thus, to establish the asymptotic normality of the ML estimator $\hat{\theta}_T$, the following moment dominance conditions must be satisfied.

(B7) For some positive constants $M_4 > 0$ and $M_4' > 0$, it holds that

$$\left\| \sup_{\theta \in V_\delta(\theta_0)} \left\| \nabla \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right) \right\| \right\|_p \leq M_4 < \infty,$$

$$\left\| \sup_{\theta \in V_\delta(\theta_0)} \left\| \nabla f_t(\theta) \right\| \right\|_p \leq M_4' < \infty, \qquad p > 2,$$

where the vector of first derivatives $\nabla \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right)$ is strictly stationary and ergodic, and is given by

$$\nabla \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right) := \nabla(\epsilon_t(\theta), \lambda)\nabla f_t(\theta) + \frac{\partial}{\partial\theta} \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right).$$

(B8) For some $M_5 > 0$ and $M_5' > 0$, it holds that

$$\left\| \sup_{\theta \in V_\delta(\theta_0)} \left\| \nabla^2 \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right) \right\| \right\|_p \leq M_5 < \infty,$$

$$\left\| \sup_{\theta \in V_\delta(\theta_0)} \left\| \nabla^2 f_t(\theta) \right\| \right\|_p \leq M_5' < \infty, \qquad p > 1,$$

where the matrix of second derivatives $\nabla^2 \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right)$ is strictly stationary and ergodic, non-singular, and is given by

$$\nabla^2 \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right) := \nabla^2(\epsilon_t(\theta), \lambda)\nabla f_t(\theta)\nabla f_t(\theta)^\top + \nabla(\epsilon_t(\theta), \lambda)\nabla^2 f_t(\theta) + \frac{\partial^2}{\partial\theta\partial\theta^\top} \log p_\epsilon\left(\epsilon_t(\theta), \lambda\right),$$

with $\left\{\nabla^2 f_t(\theta)\right\}_{t\in\mathbb{Z}} := \left\{ \frac{\partial^2 f_t(\theta)}{\partial\theta\partial\theta^\top} \right\}_{t\in\mathbb{Z}}$ denoting the strictly stationary and ergodic matrix which collects the second derivative processes of the stochastic time-varying parameter $f_t(\theta)$.

Given these assumptions, I can establish the asymptotic distribution of $\hat{F}_T(x, \hat{\theta}_T)$.

Theorem 3. *Let the conditions of Theorem 1, and Assumptions (B1)–(B8) hold true.*

Then

$$\hat{\mathscr{W}}_T(x, \hat{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(1\left\{\epsilon_t \leq x\right\} - F(x)\right) + \boldsymbol{H}(x)^\top \boldsymbol{\mathcal{J}}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla \log p_\epsilon(\epsilon_t, \lambda_0) + o_{\mathbb{P}}(1), \quad (25)$$

where

$$\boldsymbol{H}(x) := p_\epsilon(x, \lambda_0)\left(\boldsymbol{D}(x) + q_\epsilon(x, \lambda_0)\right),$$
$$\boldsymbol{D}(x) := \mathbb{E}\left[\nabla f_t(\theta_0) / \dot{g}(f_t(\theta_0), x)\right], \quad \text{and}$$
$$\boldsymbol{\mathcal{I}} := \mathbb{E}\left[-\nabla^2 \log p_\epsilon(\epsilon_t, \lambda_0)\right].$$

Furthermore

$$\hat{\mathscr{W}}_T(x, \hat{\theta}_T) \Rightarrow \boldsymbol{\mathcal{N}}\left(0, F(x)(1 - F(x)) + \boldsymbol{H}(x)^\top \boldsymbol{\Sigma} \boldsymbol{H}(x) + 2\boldsymbol{H}(x)^\top \boldsymbol{\mathcal{J}}^{-1} \boldsymbol{\Gamma}(x)\right),$$

with

$$\boldsymbol{\Sigma} := \boldsymbol{\mathcal{I}}^{-1} \boldsymbol{V} \boldsymbol{\mathcal{I}}^{-1},$$

$$\boldsymbol{V} := \mathbb{E}\left[\nabla \log p_\epsilon(\epsilon_t, \lambda_0) \nabla \log p_\epsilon(\epsilon_t, \lambda_0)^\top\right], \quad \text{and}$$

$$\boldsymbol{\Gamma}(x) := \mathbb{E}\left[1\left\{\epsilon_t \leq x\right\} \nabla f_t(\theta_0) \nabla(\epsilon_t, \lambda_0)\right] + \mathbb{E}\left[1\left\{\epsilon_t \leq x\right\} \frac{\partial}{\partial\theta} \log p_\epsilon(\epsilon_t, \lambda_0)\right].$$

The asymptotic variance in Theorem 3 splits into three parts

$$\lim_{T\to\infty} \mathbb{V}\left[\hat{\mathscr{W}}_T(x, \hat{\theta}_T)\right] = \underbrace{F(x)(1 - F(x))}_{\textit{Binomial sample noise}} + \underbrace{\boldsymbol{H}(x)^\top \boldsymbol{\Sigma} \boldsymbol{H}(x)}_{\textit{Estimation term}} + \underbrace{2\boldsymbol{H}(x)^\top \boldsymbol{\mathcal{J}}^{-1} \boldsymbol{\Gamma}(x)}_{\textit{Interaction term}}. \quad (26)$$

The binomial sampling noise $V_{BIN}(x) := F(x)(1 - F(x))$ is the variance of the indicator $\mathbb{1}\left\{\epsilon_t \leq x\right\}$ for *IID* data and it persists even if θ_0 were known. The estimation term $V_{EST}(x) := \boldsymbol{H}(x)^\top \boldsymbol{\Sigma} \boldsymbol{H}(x)$ arises because an estimated parameter vector is substituted in the approximate noise; its size is quadratic in the estimation error through $\boldsymbol{\Sigma} = \boldsymbol{\mathcal{I}}^{-1} \boldsymbol{V} \boldsymbol{\mathcal{I}}^{-1}$, that is, is the asymptotic covariance of $\sqrt{T}(\hat{\theta}_T - \theta_0)$, but is also modulated by the model through the weight $\boldsymbol{H}(x)$, which is model dependent. Finally, the interaction term $V_{INT}(x) := 2\boldsymbol{H}(x)^\top \boldsymbol{\mathcal{I}}^{-1} \boldsymbol{\Gamma}(x)$ captures the cross-effect between estimation error and the filter's nonlinear dynamics. When the noise density is parameter-free, for example, in the Gaussian GARCH(1,1) analyzed in Section 4.1, this cross term simplifies since all θ_0-dependent factors cancel, so the variance coincide (under Gaussian noise) to the AMF and AED structure of Francq and Zakoïan (2022). This simplification is proved formally in Corollary 3.1. For more general score-driven models, such as the heavy-tailed Beta-*t*-GARCH(1,1) treated in Section 6, no such simplification occurs and the interaction component remains fully operative. Hence the last two addends in (26) quantify, respectively, the extra variability injected by parameter uncertainty alone and the additional variability created by its coupling with model dynamics. Their relative magnitudes are documented for the Beta-*t*-GARCH(1,1) example in Section 6.

Remark 6. Having established the main theorem of this Section, I can now compare my assumptions with those of Francq and Zakoïan (2022), continuing the discussion begun in the Introduction. In particular, Francq and Zakoïan (2022) work with two series of conditions, A1–A12 for pure conditional scale models and B1–B8 when a conditional mean is added. The discussion below aligns each of my assumptions with its Francq and Zakoïan (2022) counterpart and highlights the divergences that matter for the limit results:

- Assumption 1 evaluated at $\theta = \theta_0$ plays the same role as A1 and B1 of Francq and Zakoïan (2022). All three guarantee a unique strictly stationary, ergodic solution, but the contraction here is applied to the score recursion instead of the conditional location-scale equations.

- My requirement (B1), strict monotonicity and a positive lower bound for the time-varying parameter $f_t(\theta)$, echoes A2 for scale models and B2 for location-scale models in Francq and Zakoïan (2022), which bound the conditional scale process away from zero and impose differentiability. Assumption (B1), however, still allows link functions that can approach, or even attain, zero (for instance the logistic or exponential link), provided that the expected value of $1/\dot{g}\left(f_t(\theta), \epsilon_t\right)$, with $\dot{g}\left(f_t(\theta), \epsilon_t\right)$ as in (19), remains finite.
- Assumptions A3 (scale) and B3 (location-scale) in Francq and Zakoïan (2022) impose a geometric-decay inequality on the conditional location-scale recursion, yielding invertibility. The same role is played in this article by Theorem 1, where I obtain invertibility for the score-driven filter via a global Lipschitz contraction combined with Orlicz-norm tail bounds.
- In (B2) I ask for either an Orlicz bound (21) or a finite p-moment with $p > 2$ (22) for the stochastic factor $\nabla f_t(\theta) \Big/ \dot{g}(f_t(\theta), \epsilon_t)$. This mirrors A4/B4 of Francq and Zakoïan (2022): those assumptions are written with $r > 0$, but the proofs that yield asymptotic normality actually set $r > 2$ (and sometimes $r \geq 4$). My $p > 2$ branch therefore matches their effective requirement, while the optional Orlicz branch, though stronger, delivers the sub-Gaussian maximal and deviation bounds used in Lemma 1 and Theorem 1.
- Assumption (B3) states strong consistency explicitly, in line with Proposition 3.2 of Blasques et al. (2018). Francq and Zakoïan (2022) obtain the same property via their Assumptions A1–A3 for pure-scale models, and impose it directly as B2 in the conditional location-scale case. Thus the three sets of premises are equivalent.
- The Lipschitz requirement on the noise distribution function F in (B4) coincides with Assumption A5 of Francq and Zakoïan (2022). That assumption remains in force when they move to the conditional location-scale setting, so there is no separate B-condition for it. Likewise, the smoothness condition on the noise density p_ϵ in (B5), matches their Assumption A11, which is also retained unchanged in the conditional location-scale framework. However, note that since score-driven models are fully parametric, and the noise density p_ϵ may depends directly on the unknown parameter vector λ_0, (B5) must hold for every λ in the compact parameter set Λ that contains λ_0, a condition that is unnecessary in the semi-parametric framework of Francq and Zakoïan (2022).
- (B6) is specific to the score-driven framework. In fact, since Francq and Zakoïan (2022) treat the noise law as parameter-free, no counterpart to (B6) appears in their assumption lists. Note that when the score-driven specification degenerates to the Gaussian GARCH(1,1) case, the situation covered by Corollary 3.1, the density is no longer parameter-dependent and (B6) becomes superfluous.
- Finally, Assumptions (B7) and (B8) bounds the $p > 2$-moments of the first derivatives, and $p > 1$-moments of the second derivatives, respectively, uniformly over a small neighborhood of θ_0. The closest items in Francq and Zakoïan (2022) are A9 (pure conditional scale) and B7 (location-scale), which control the derivatives of the conditional scale and the conditional locations. In the broader score-driven setting treated here no universal coefficient inequalities exist, so verification of (B7) and (B8) must be carried out model-by-model once the specific recursion and noise distribution are fixed.

Remark 7. Francq and Zakoïan (2022) offer two reference cases. First, for pure conditional scale models, their Theorem 2.2 yields an AMF and AED variance that is entirely filter-free under their Assumptions A1–A12. Second, for conditional location-scale models, Theorem 3.2 adds a generally AMD term. This term disappears, and the variance becomes AMF, only if both (a) the extension of the scaling assumption (denoted A12* in Francq and Zakoïan (2022)) and (b) the moment-orthogonality condition introduced in their Remark 3.2 hold. That conjunction is satisfied, for example, in centered ARMA-GARCH models without intercept, leading to Corollary 3.2 of Francq and Zakoïan (2022). In contrast, in the score-driven setting studied here the AMD component persists for all parameter values because the same vector θ_0 appears in both the recursion and the noise density; it vanishes only in special cases such as the Gaussian-GARCH model covered by Corollary 3.1.

Remark 8. Since the finite-dimensional parameter $\lambda_0 \subset \theta_0$ also enters the noise density $p_\epsilon(\cdot, \lambda_0)$, the limiting variance of the approximate empirical distribution function is never fully AMF in score-driven models. Therefore, the residual-kernel estimators that are valid only under the AMF scenario identified by Francq and Zakoïan (2022) (Theorem 2.2, Remark 2.2, Remark 3.3) are not applicable here. Consistent inference instead requires joint parametric estimation of $(\omega_0, \alpha_0, \beta_0, \lambda_0)$, rather than the plug-in kernel methods used for ARMA (Hart and Vieu 1990; Wu and Mielniczuk 2002) and GARCH (Kulperger and Yu 2005; Horváth and Zitikis 2006) models.

4.1 | The Special Case of the GARCH(1,1) Model

As shown by Creal et al. (2011), when (1) is a volatility model with Gaussian noise, $\epsilon_t \sim \mathcal{N}(0,1)$ and $y_t = \sqrt{f_t(\theta_0)}\epsilon_t$, the score–driven specification (1) to (3) specializes to the classical GARCH(1,1) model of Engle (1982); Bollerslev (1986). In this Gaussian benchmark, Theorem 3 yields the limiting law and variance derived below for GARCH(1,1); these coincide with the standard formulas. This agrees with the Gaussian specialization of Francq and Zakoïan (2022), noting that their results are semi-parametric while my analysis is fully parametric; the two frameworks are complementary, and under Gaussian noise their limits coincide.

4.1.1 | Relationship to Francq and Zakoïan (2022)

Francq and Zakoïan (2022) analyze the empirical distribution of residuals in conditional location-scale models $y_t = m_t(\theta_0) + \sigma_t(\theta_0)\eta_t$, where m_t, σ_t depend on past observations and $\{\eta_t\}_{t\in\mathbb{Z}}$ is *IID* with density free of θ_0. Under suitable conditions they show that the empirical process is AED and AMF: its limiting variance involves only the noise density, not the particular choice of m_t or σ_t.

In the score-driven framework studied here, the static parameter θ_0 enters both the noise density $p_\epsilon(\cdot,\lambda_0)$ and the recursion $f_{t+1}(\theta_0) = \phi(f_t(\theta_0), y_t, \theta_0)$ (see (1) to (3)). As a result, the limit variance in Theorem 3 contains an AMD component in addition to the AED term (see Remarks 7 and 8); the AMF property of Francq and Zakoïan (2022) therefore need not hold once parameters influence both the noise law and the dynamics.

The two sets of results nevertheless intersect. When the link function is fixed at $y_t = \sqrt{f_t(\theta_0)}\epsilon_t$ with Gaussian noise, the score recursion becomes the GARCH(1,1) update; the AMD term simplifies, and Corollary 3.1 exactly reproduces the variance expression of Theorem 2.2 in Francq and Zakoïan (2022). Thus, under Gaussian noise, the conditional scale result of Francq and Zakoïan (2022) is recovered within the score-driven framework.

One key structural difference remains: Francq and Zakoïan (2022) work in a semi-parametric setting since the density of η_t is unspecified, whereas score-driven models are fully parametric and the noise density $p_\epsilon(\cdot,\lambda_0)$ is indexed by the unknown vector λ_0, thereby introducing AMD. Accordingly, Francq and Zakoïan (2022) provide an AMF and AED benchmark for conditional location-scale models, while this article complements their results by pinpointing when and why the limit law becomes inherently AMD in general score-driven dynamics.

This equivalence is formally stated in the following corollary.

Corollary 3.1. *Suppose that the score-driven model in equations (1) to (3) satisfies*

$$y_t = \sqrt{f_t(\theta_0)}\epsilon_t, \quad \epsilon_t \overset{IID}{\sim} \mathcal{N}(0,1),$$

$$f_{t+1}(\theta_0) = \omega_0 + \beta_0 f_t(\theta_0) + \alpha_0 y_t^2,$$

and let conditions A.1–A.6 in chapter 7 of Francq and Zakoian (2019) *hold true.*

Then, Theorem 3 implies that

$$\hat{\mathscr{W}}_T(x,\hat{\theta}_T) \Rightarrow \mathcal{N}\left(0,\ F(x)(1-F(x)) + \frac{(xp_\epsilon(x))^2}{4}\left(\mathbb{E}\left[|\epsilon_t|^4\right]-1\right) + xp_\epsilon(x)\zeta_\epsilon(x)\right),$$

with

$$\zeta_\epsilon(x) := \mathbb{E}\left[\mathbb{1}\left\{\epsilon_t \le x\right\}\epsilon_t^2\right] - F(x).$$

Remark 9. In this Gaussian GARCH(1,1) limit the variance in Corollary 3.1 is AMF but AED (see Francq and Zakoïan (2022), Remark 1); thus a residual-kernel estimator for p_ϵ can be used here, in contrast with the generic score-driven case highlighted in Remarks 7 and 8.

5 | A Residual-Based Kolmogorov–Smirnov Test

Goodness-of-fit (GoF) checks are indispensable once a score-driven (observation-driven) model has been estimated: we need evidence that its conditional density actually describes the data. A natural candidate is the Kolmogorov–Smirnov (KS) test, that can be formulated in terms of the probability integral transform (PIT) of the noise process $\{\epsilon_t\}_{t\in\mathbb{Z}}$. The PIT of a variable is given by its cumulative distribution function $PIT(x) = F(x)$, which, under correct specification, is uniformly distributed on the interval $[0,1]$ (see Pearson (1938) or Theorem 2.1.10 in Casella and Berger (2002)). Recall that the noise process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ in equation (1) is assumed *IID* with distribution function $F(x) = \mathbb{P}(\epsilon_t \le x)$.

Let $U_t := F(\epsilon_t)$, where $F(\epsilon_t)$ is the PIT of the noise process $\{\epsilon_t\}_{t\in\mathbb{Z}}$. Under correct specification, the PIT sequence $\{U_t\}_{t=1}^T$ satisfies $U_t \overset{\text{IID}}{\sim} \mathcal{U}(0,1)$, where $\mathcal{U}(0,1)$ denotes the uniform distribution in $[0,1]$. A formal goodness-of-fit therefore tests

$$H_0 : \{U_t\}_{t=1}^T \quad IID \quad \mathcal{U}(0,1). \tag{27}$$

Since the noise process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is not observable in practice, it must be replaced with its approximation $\{\hat{\epsilon}_t(\hat{\theta}_T)\}_{t=1}^T$, thus yielding $\hat{U}_t = F\left(\hat{\epsilon}_t(\hat{\theta}_T)\right)$, that is, the PIT of the approximate noise process $\{\hat{\epsilon}_t(\hat{\theta}_T)\}_{t=1}^T$. While the classical KS statistic has been employed in the score-driven framework by Harvey (2013), 54–55, the AMD property discussed here implies that its use is not theoretically valid, as the null distribution is no longer distribution-free and the usual Brownian bridge critical values are incorrect. To address this issue, I adopt the modified KS test of Bai (2003), which relies on the martingale transformation of Khmaladze (1981) to obtain an asymptotically distribution-free procedure for general nonlinear dynamic models.

To properly adapt this test for score-driven models, I then introduce a new empirical process

$$\mathcal{K}_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\mathbb{1}\left\{U_t \le r\right\} - r\right),$$

for $r \in [0,1]$. Consequently, I define the approximate empirical process

$$\hat{\mathcal{K}}_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\mathbb{1}\left\{\hat{U}_t \le r\right\} - r\right). \tag{28}$$

A closer examination of the proof of Theorem 3 reveals that the asymptotic representation in (25) follows from the asymptotic expansion:

$$\hat{\mathscr{W}}_T(x,\hat{\theta}_T) = \mathscr{W}_T(x) + p_\epsilon(x,\lambda_0)\left(\frac{1}{T}\sum_{t=1}^T \frac{\nabla f_t(\theta_0)}{\dot{g}\left(f_t(\theta_0),x\right)} + q_\epsilon\left(x,\lambda_0\right)\right)^\top \times \sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_{\mathbb{P}}(1), \tag{29}$$

see equations (S.12), (S.13) and Lemma S4 in Appendix B. The asymptotic expansion in (29) is the key for addressing the GoF problem using the martingale transform of Khmaladze (1981).

By noting that $\hat{\mathcal{W}}_T(x, \hat{\theta}_T) = \hat{\mathcal{K}}_T(F(x))$, and letting $F(x) = r$, it follows that

$$\hat{\mathscr{K}}_T(r) = \mathscr{K}_T(r) + p_\epsilon(Q(r), \lambda_0)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla f_t(\theta_0)}{\dot{g}\left(f_t(\theta_0), Q(r)\right)} + q_\epsilon\left(Q(r), \lambda_0\right)\right)^\top \times \sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_{\mathbb{P}}(1), \tag{30}$$

where $Q(r) = F^{-1}(r) = x$ is the inverse function of $F(x) = r$. The asymptotic representation in (30) has the same form as eq. (2) of Bai (2003). In fact, under the conditions of Theorem 3, there exist a vector $\boldsymbol{u}(r)$ that satisfies

$$\lim_{T\to\infty}\mathbb{P}\left(\left\|p_\epsilon(Q(r), \lambda_0)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla f_t(\theta_0)}{\dot{g}\left(f_t(\theta_0), Q(r)\right)} + q_\epsilon\left(Q(r), \lambda_0\right)\right) - \boldsymbol{u}(r)\right\| > z\right) = 0,$$

for all $z > 0$ so that, equation (30) can be rewritten as follows:

$$\hat{\mathcal{K}}_T(r) = \mathcal{K}_T(r) + \boldsymbol{u}(r)^\top\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_{\mathbb{P}}(1),$$

with

$$\boldsymbol{u}(r) = \operatorname*{plim}_{T\to\infty}\, p_\epsilon(Q(r), \lambda_0)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla f_t(\theta_0)}{\dot{g}\left(f_t(\theta_0), Q(r)\right)} + q_\epsilon\left(Q(r), \lambda_0\right)\right).$$

The second term on the rhs of equation (30) shows how estimation turns this bridge into a drift term that must be removed. The martingale transformation of Khmaladze (1981) then takes the form:

$$\hat{\mathcal{M}}_T(r) = \hat{\mathcal{K}}_T(r) - \int_0^r\left(\dot{\boldsymbol{u}}(s)^\top\boldsymbol{C}^{-1}(s)\int_s^1\dot{\boldsymbol{u}}(\iota)\mathrm{d}\hat{\mathcal{K}}_T(\iota)\right)\mathrm{d}s, \tag{31}$$

where $\dot{\boldsymbol{u}}(s) = \partial\boldsymbol{u}(s)/\partial s$, and $\boldsymbol{C}(s) = \int_s^1\dot{\boldsymbol{u}}(\iota)\dot{\boldsymbol{u}}(\iota)^\top\mathrm{d}\iota$. Then, the test statistics of interest is given by:

$$\hat{\mathcal{B}}_T = \sup_{0\le r\le 1}\left|\hat{\mathcal{M}}_T(r)\right|, \tag{32}$$

that is the sup-distance between the approximate empirical distribution function of $\hat{\epsilon}_t(\hat{\theta}_T)$ and the true model distribution function of ϵ_t.

Under the conditions of Theorem 3, the martingale transformation in (31) is well defined. Assumptions (B1), (B2), and (B6) ensure that $\boldsymbol{u}(s)$ is continuously differentiable on $s \in [0, 1]$, and its second moment is uniformly bounded. Assumptions (B7) and (B8) guarantee that the weighting matrix $\boldsymbol{C}(s)$ is well defined and nonsingular for every $s \in [0, 1)$.

Therefore, according to Khmaladze (1981) and Corollary 1 of Bai (2003) the process $\hat{\mathcal{M}}_T(r)$ in equation (31) converges weakly to a standard Brownian motion $\mathcal{M}(r)$ so that, by the continuous mapping theorem, it follows that:

$$\hat{\mathcal{B}}_T \Rightarrow \mathcal{B} = \sup_{0\le r\le 1}|\mathcal{M}(r)|, \qquad T \to \infty. \tag{33}$$

In practice, the distribution of the supremum of the absolute value of standard Brownian motion $\mathcal{B}$ given in equation (33) can be easily obtained via simulations. However, its distribution is already known in closed-form, and it is given by:

$$F_{\mathscr{B}}(z) := \mathbb{P}(\mathscr{B} \le z) = \frac{4}{\pi}\sum_{T=0}^{\infty}\frac{(-1)^T}{2T+1}\exp\left\{-\frac{(2T+1)^2\pi^2}{8z^2}\right\}, \quad z > 0. \tag{34}$$

TABLE 2 | Critical values for $\mathcal{B} = \sup_{0\le r\le 1}|\mathcal{M}(r)|$ Quantiles are obtained by numerically solving $F_{\mathcal{B}}(z) = 1 - \alpha$ for z, where $F_{\mathcal{B}}(z)$ is given in (34).

Tail probability α	Confidence level	c_α
0.10	90%	1.960
0.05	95%	2.241
0.025	97.5%	2.498
0.01	99%	2.807
0.005	99.5%	3.023
0.001	99.9%	3.481

Note: Ten terms of the series yield accuracy better than 10^{-8}; the values are rounded at the third decimal.

This limiting distribution was already proved by Erdös and Kac (1946) (see also Feller (1971), 343 for a detailed derivation). The upper-tail quantile c_α for any significance level α can be easily obtained by inverting the upper-tail probability $\mathbb{P}(\mathcal{B} > z) = 1 - F_{\mathcal{B}}(z)$ using the closed-form series formula in (34) for $z > 0$, and are listed in Table 2.

The martingale transform in (31) is implemented via a discrete approximation on a fine grid, leading to the statistic $\tilde{\mathcal{B}}_T$ in (S.36). An explicit discretization scheme, including the construction of the grids, the numerical evaluation of $\dot{\boldsymbol{u}}(s)$ and $\boldsymbol{C}(s)$, and a stabilized pseudo-inverse for $\boldsymbol{C}(s)$, is provided in Appendix D. This scheme refines the algorithmic approach of Bai (2003) by tailoring the Khmaladze transform to score-driven models and by treating the numerical integration and matrix inversion steps in a more systematic way.

In the next section, devoted to the Beta-t-GARCH(1,1) model, I employ this implementation to carry out simulation experiments evaluating the size and power of the test, directly comparing the results with those obtained from the classical KS approach proposed by Harvey (2013), 54–55, and to present an empirical illustration based on EUR/USD returns.

6 | Application to the Beta-*t*-GARCH(1,1)

This section illustrates the practical applicability of the theoretical results derived above by verifying explicitly the assumptions for the Beta-t-GARCH(1,1) model. The Beta-t-GARCH(1,1) model was introduced by Harvey (2013). In this model, Equations (1) and (2) take the form:

$$y_t = g\left(f_t(\theta_0), \epsilon_t\right) = \sqrt{f_t(\theta_0)}\epsilon_t, \quad \epsilon_t \overset{IID}{\sim} p_\epsilon(\epsilon_t, \lambda_0), \tag{35}$$

where $\left\{\epsilon_t\right\}_{t\in\mathbb{Z}}$ is an *IID* sequence of standard Student's t random variables with $\lambda_0 > 2$ degrees of freedom, having density:

$$p_\epsilon(\epsilon_t, \lambda_0) = \frac{\Gamma\left(\frac{\lambda_0+1}{2}\right)}{\sqrt{\pi\left(\lambda_0 - 2\right)}\Gamma\left(\frac{\lambda_0}{2}\right)}\left(1 + \frac{\epsilon_t^2}{\lambda_0 - 2}\right)^{-\frac{\lambda_0+1}{2}}, \tag{36}$$

and the stochastic variance process follows

$$f_{t+1}(\theta_0) = \omega_0 + \beta_0 f_t(\theta_0) + \alpha_0 \frac{\left(\lambda_0 + 1\right) y_t^2}{\lambda_0 - 2 + y_t^2 / f_t(\theta_0)}, \tag{37}$$

where $\omega_0 > 0$, $\alpha_0 \geq 0$ and $\beta_0 \geq 0$, so that $\theta_0 = \left(\omega_0, \alpha_0, \beta_0, \lambda_0\right)^\top \in \Theta \subset \mathbb{R}^4$. From equations (36) and (37), it is clear that the parameter λ_0 appears in both the noise density and the stochastic variance process $\left\{f_t(\theta_0)\right\}_{t\in\mathbb{Z}}$. As $\lambda_0 \to \infty$, the Beta-*t*-GARCH(1,1) model reduces to the standard GARCH(1,1) model. This occurs because, in this limit, the Student's *t* distribution converges to the Gaussian distribution, and the scaled-score function

$$s_t\left(f_t(\theta_0), \lambda_0\right) = \frac{\left(\lambda_0 + 1\right) y_t^2}{\lambda_0 - 2 + y_t^2 / f_t(\theta_0)},$$

satisfies $\lim_{\lambda_0 \to \infty} s_t\left(f_t(\theta_0), \lambda_0\right) = y_t^2$. Substituting this into the recursion equation (37) yields

$$f_{t+1}(\theta_0) = \omega_0 + \beta_0 f_t(\theta_0) + \alpha_0 y_t^2,$$

which is exactly the conditional variance equation of the standard GARCH(1,1) model (see Section 4.1).

It is straightforward to rewrite equation (37) in terms of the noise process $\left\{\epsilon_t\right\}_{t\in\mathbb{Z}}$:

$$f_{t+1}(\theta_0) = \omega_0 + \beta_0 f_t(\theta_0) + \alpha_0 f_t(\theta_0) \frac{\left(\lambda_0 + 1\right)\epsilon_t^2}{\lambda_0 - 2 + \epsilon_t^2}.$$

Applying the Cauchy root test, a strictly stationary and the ergodic solution exists if

$$\mathbb{E}\left[\log\left|\beta_0 + \alpha_0 \frac{\left(\lambda_0 + 1\right)\epsilon_t^2}{\lambda_0 - 2 + \epsilon_t^2}\right|\right] < 0, \tag{38}$$

see Harvey (2013), 125, or Blasques et al. (2018). Also, note that equations (35) to (37) define a nonlinear GARCH(1,1) model of the general type studied in Straumann and Mikosch (2006). In their SRE framework, the random Lipschitz coefficient of the variance update is precisely $\beta_0 + \alpha_0 \frac{(\lambda_0+1)\epsilon_t^2}{\lambda_0 - 2 + \epsilon_t^2}$, so the contraction condition in (38) is of the same log-contraction type as their stationarity and ergodicity assumptions for nonlinear GARCH processes.

Several statistical properties of this model have been extensively studied by Creal et al. (2011), Harvey (2013) and Blasques et al. (2018). The invertibility of this model was first analyzed in Blasques et al. (2018), whereas a full derivation of the asymptotic theory for the MLE can be found in Blasques et al. (2022). Further statistical properties, such as the local asymptotic normality, has been recently established by Francq and Zakoian (2023).

6.1 | Maximal and Deviation Inequalities

As discussed above, it is evident that, when evaluated at the true parameter θ_0, the process in (37) is linear in $f_t(\theta_0)$. However, for any $\theta \in \Theta$ such that $\theta \neq \theta_0$, the stochastic recurrence equation in (7) satisfies:

$$f_{t+1}(\theta) = \omega + \beta f_t(\theta) + \alpha \frac{(\lambda + 1) y_t^2}{\lambda - 2 + y_t^2 / f_t(\theta)}. \tag{39}$$

It is clear that this recursion is highly nonlinear in $f_t(\theta)$, as the driving force is given by $s_t\left(f_t(\theta), \lambda\right) = \frac{(\lambda+1)y_t^2}{\lambda - 2 + y_t^2/f_t(\theta)}$. Nevertheless, in what follows, I demonstrates that the results obtained in Sections 3 and 4 remain valid despite this nonlinearity.

I start by specializing the maximal inequality obtained in Lemma 1. By straightforward calculation, for all $f \in \mathcal{F}$ and when $\lambda < \infty$, the stochastic process defined in (12) is given by

$$X_t\left(\xi_{f,\theta}\right) = \beta + \alpha \frac{(\lambda + 1) y_t^4}{\left((\lambda - 2) f + y_t^2\right)^2}.$$

This process, is uniformly bounded over $\mathcal{F} \times \mathbb{R} \times \Theta$ and for all $t \in \mathbb{Z}$. In fact, by compactness of the parameter space, one can define Θ as a subset of $\mathbb{R}^4$ of the following type

$$\begin{aligned} \Theta = \{\theta : 0 < \underline{\omega} \leq \omega \leq \overline{\omega} < \infty,\ 0 < \underline{\alpha} \leq \alpha \leq \overline{\alpha} < \infty, \\ 0 < \underline{\beta} \leq \beta \leq \overline{\beta} < 1,\ 2 < \underline{\lambda} \leq \lambda \leq \overline{\lambda} < \infty\}, \end{aligned} \tag{40}$$

in which a finite Orlicz norm for $\sup_{\xi_{f,\theta}\in\Xi_\theta} |X_t(\xi)|$ always exists since the entropy integral in (13) is bounded, and condition (*i*) and (*ii*) of Theorem 1 are also satisfied, see equations (S.28–S.31) in Appendix C for a detailed verification. It follows that Lemma 1, Theorem 1 and Corollary 1.1 take the following simplified form.

Corollary 3.2. *For the Beta-t-GARCH(1,1) model in (35) to (37) if (38) and (40) are satisfied, then* $\exists \varrho$ *which satisfies* $0 < \varrho < 1$ *such that*

$$\begin{aligned} L_1(\xi) &= d + \int_0^{\alpha(\lambda+1)} \sqrt{\log(1 + 1/\varepsilon)}\mathrm{d}\varepsilon \leq \varrho, \\ d &:= \left\| \beta + \alpha \frac{(\lambda + 1) y_t^4}{\left((\lambda - 2) f_0 + y_t^2\right)^2} \right\|_{\psi_2}. \end{aligned} \tag{41}$$

Consequently, it holds that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\hat{f}_{t+1}(\theta) - f_{t+1}(\theta)\right|\right] \leq \tau\sqrt{\log(1 + \varrho^t\tau)} \to 0, \quad t \to \infty,$$

and therefore, for any $u > 0$

$$\mathbb{P}\left(\sup_{\theta\in\Theta}\left|\hat{f}_{t+1}(\theta) - f_{t+1}(\theta)\right| \geq 2\varrho^t\tau\sqrt{u}\right) \leq 2e^{-u}.$$

Thus, choosing $u = c\log t$ *for any constant* $c > 1$ *gives*

$$\sup_{\theta\in\Theta}\left|\hat{f}_{t+1}(\theta) - f_{t+1}(\theta)\right| = O_{a.s.}\left(\varrho^t\tau\sqrt{\log t}\right).$$

6.2 | Empirical Process and Its Asymptotic Distribution

As shown in the previous subsection, the Beta-*t*-GARCH(1,1) model satisfies the conditions of Theorem 1. Therefore, to establish that its approximate empirical distribution function belongs to the Glivenko–Cantelli class, it suffices to verify the additional condition required for obtaining the probability bound in Lemma 2. This, in turn, ensures that Theorem 2 also holds.

By inverting the relationship in equation (35), I obtain

$$\epsilon_t = \frac{y_t}{\sqrt{f_t(\theta_0)}}.$$

Applying the inverse function rule, for any given $\theta \in \Theta$, I get

$$\epsilon_t(\theta) = \frac{y_t}{\sqrt{f_t(\theta)}}, \quad \dot{g}\left(f_t(\theta), \epsilon_t\right) = -\frac{2}{\epsilon_t} \cdot \sqrt{\frac{f_t(\theta)}{f_t(\theta_0)}} \cdot f_t(\theta). \tag{42}$$

The compactness of Θ defined in (40) ensures that $f_t(\theta)$ lies in the set $[\underline{\omega}/(1-\overline{\beta}), \infty)$, as $\underline{\omega}/(1-\overline{\beta}) = \inf_{\theta \in \Theta} \omega/(1-\beta)$ with $\underline{\omega} > 0$ and $\overline{\beta} < 1$. Hence, taking the supremum of the inverse of equation (42) yields

$$M_{1,t} = \frac{|\epsilon_t|}{2} \cdot \frac{\sqrt{f_t(\theta_0)}}{(\underline{\omega}/(1-\overline{\beta}))^{3/2}},$$

where $M_{1,t} > 0$ almost surely and $\mathbb{E}\left[\log^+ M_{1,t}\right] < \infty$, which shows that Assumption (B1) is satisfied.

Now, the verification of Assumption (B2) requires the calculation of the vector of first derivative processes $\left\{\nabla f_t(\theta)\right\}_{t\in\mathbb{Z}}$. For the Beta-*t*-GARCH(1,1) model, I have

$$\frac{\nabla f_t(\theta)}{\dot{g}\left(f_t(\theta), \epsilon_t\right)} = -\frac{\epsilon_t}{2} \cdot \sqrt{\frac{f_t(\theta_0)}{f_t(\theta)}} \cdot \frac{\nabla f_t(\theta)}{f_t(\theta)}, \tag{43}$$

where $\nabla f_t(\theta)$ is given by

$$\nabla f_{t+1}(\theta) := \begin{pmatrix} \frac{\partial f_{t+1}(\theta)}{\partial \omega} \\ \frac{\partial f_{t+1}(\theta)}{\partial \alpha} \\ \frac{\partial f_{t+1}(\theta)}{\partial \beta} \\ \frac{\partial f_{t+1}(\theta)}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{(\lambda+1)y_t^2}{\lambda-2+y_t^2/f_t(\theta)} \\ f_t(\theta) \\ \alpha \frac{f_t(\theta)y_t^2\left(y_t^2-3f_t(\theta)\right)}{\left((\lambda-2)f_t(\theta)+y_t^2)\right)^2} \end{pmatrix} + \left(\beta + \alpha \frac{(\lambda+1)y_t^4}{\left((\lambda-2)f_t(\theta)+y_t^2\right)^2}\right)\nabla f_t(\theta). \tag{44}$$

Since the noise process $\left\{\epsilon_t\right\}_{t\in\mathbb{Z}}$ is Student's *t* distributed with $\lambda_0 > 2$ degrees of freedom, the stochastic factor in (43) is heavy-tailed and Assumption (B2)–(21) fails. However, under Assumption 1, since

$$M_{2,t} = \frac{|\epsilon_t|}{2} \cdot \sup_{\theta \in V_\delta(\theta_0)} \left\| \sqrt{\frac{f_t(\theta_0)}{f_t(\theta)}} \cdot \frac{\nabla f_t(\theta)}{f_t(\theta)} \right\|,$$

I can use Lemma G.1 of Francq and Zakoian (2023) to prove that each term of the ratio vector process $\left\{\nabla f_t(\theta)/f_t(\theta)\right\}_{t\in\mathbb{Z}}$ can be bounded uniformly over $V_\delta\left(\theta_0\right)$. Moreover, analogously to the usual GARCH(1,1) model, the ratio process $\left\{\sqrt{f_t(\theta_0)/f_t(\theta)}\right\}_{t\in\mathbb{Z}}$ is naturally bounded uniformly over $V_\delta\left(\theta_0\right)$. Hence, $\left\|M_{2,t}\right\|_p \leq M_2 < \infty$ for any $2 < p < \lambda_0$, thus verifying Assumption (B2)–(22).

Therefore, by specifying Lemma 2 for the Beta-*t*-GARCH(1,1), I obtain the following corollary which gives the tails bound for the difference process $\left\{\hat{\Delta}_{t+1}(\hat{\theta}_T)\right\}_{t=1}^T$ in (20).

Corollary 3.3. *For the Beta-t-GARCH(1,1) model in (35) to (37) if (38), (40) and (41) are satisfied, I have*

$$\mathbb{P}\left(\left|\hat{\Delta}_{t+1}(\hat{\theta}_T)\right| \geq \varrho^t \tau \left|\epsilon_t\right| \cdot \frac{\sqrt{f_t(\theta_0)}}{(\underline{\omega}/(1-\overline{\beta}))^{3/2}} \sqrt{u} + t^{1+1/p} u^{-1/p}\right)$$
$$\leq 2e^{-u} + C D(\delta_T)^p t^{-(p+1)} u^2,$$

so that the choice $u = c \log t$ for any constant $c > 1$ implies that

$$\left|\hat{\Delta}_{t+1}(\hat{\theta}_T)\right| = O_{a.s.}\left(\varrho^t \tau \sqrt{\log t} + D(\delta_T)(\log t)^{1/p}\right),$$

for any $2 < p < \lambda_0$.

The MLE for the Beta-*t*-GARCH(1,1) model is strongly consistent under the low-level regularity conditions already imposed: (*i*) Stationarity and ergodicity ensured by the contraction condition in (38); (*ii*) Compactness of the parameter set specified in (40); Invertibility guaranteed by (41). These three ingredients satisfy the requirements of the general consistency results for score-driven models in Theorem 6.3 of Blasques et al. (2018) and Theorem 2.2 of Blasques et al. (2022). Consequently, Assumption (B3) is automatically fulfilled.

I conclude this section by discussing the conditions from Assumption (B4) through Assumption (B8) required for applying Theorem 3. Clearly, the Student's *t* distribution and density functions satisfy the Lipschitz continuity and smoothness conditions in Assumptions (B4) and (B5), respectively. For the remaining condition, a number of derivatives need to be calculated.

In particular, it can be shown that since all the involved derivative functions are uniformly bounded, Assumptions (B6)–(B8) are clearly satisfied, from which the results established in Theorems 2 and 3 follow without additional restrictions. Full derivations can be found in equations (S.32–S.35) in Appendix C.

Corollary 3.4. *The Glivenko-Cantelli Theorem 2 and the asymptotic results derived in Theorem 3 are valid for the Beta-t-GARCH(1,1) model in (35) to (37) if (38), (40) and (41) are satisfied.*

Given the explicit verification above, the Beta-*t*-GARCH(1,1) model meets all the required theoretical conditions of Sections 3 and 4. Therefore, the maximal and deviation inequalities, consistency results, and empirical-process limits derived previously are directly applicable, validating the theoretical framework in practice.

6.3 | Numerical Illustrations

For illustration, I numerically evaluate the theoretical results derived in this article based on the data-generating process defined in equations (35) to (37). Using these recursive equations, I generate a time series dataset of size $T = 25,000$ for a wide range of values for the static parameters ω_0, α_0, β_0 and λ_0.

6.3.1 | Invertibility Region

With these generated datasets, I first present the invertibility region that satisfies the condition $L_1(\xi) < 1$, which is necessary

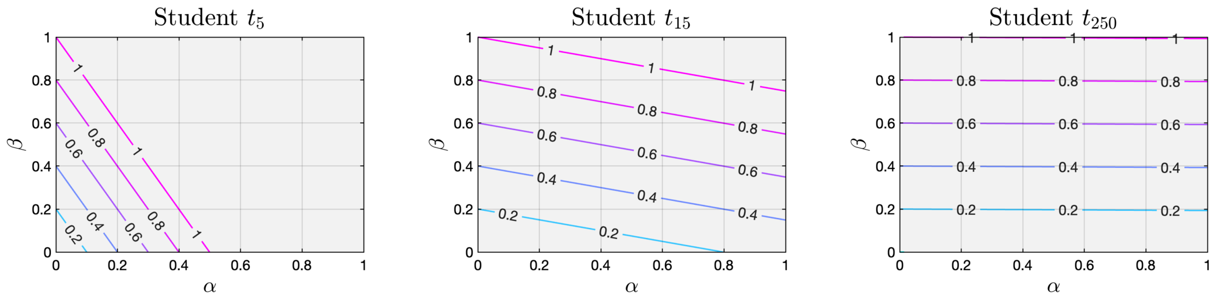

FIGURE 1 | Simulated invertibility region based on $L_1(\xi)$ in equation (13) for the Beta-*t*-GARCH(1,1) model, where the noise follows a Student's *t* distribution with $\lambda_0 = \{5, 15, 250\}$.

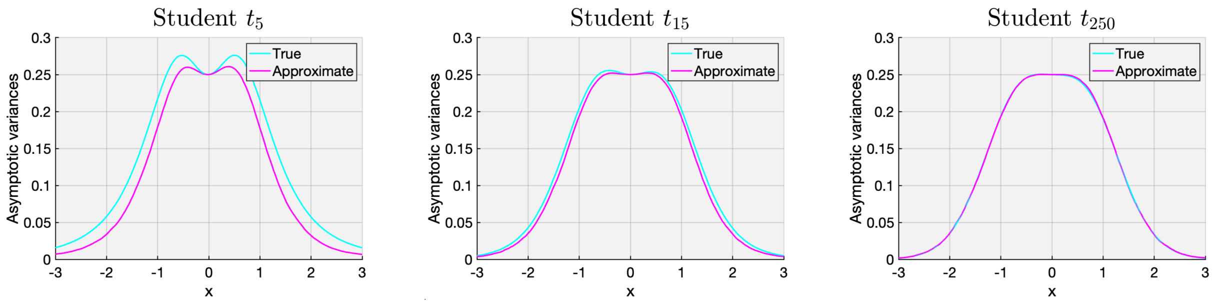

FIGURE 2 | Asymptotic variances of the empirical process for the approximated (pink) and true (cyan) noise in the Beta-*t*-GARCH(1, 1) model. The noise follows a Student's *t* distribution with degrees of freedom $\lambda_0 \in \{5, 15, 250\}$. See Appendix C for the explicit formula used to produce this figure.

to establish the maximal and deviation inequalities in Theorem 1 and all subsequent theoretical results. Here, $L_1(\xi)$ is defined in equation (13) of Lemma 1. This invertibility region is depicted in Figure 1.

When $\lambda_0 = 5$, the contours in Figure 1 are steep and diagonal, indicating a tight inverse relation between α_0 and β_0 and a relatively small invertibility region, in line with the strong nonlinearity induced by heavy tails. For $\lambda_0 = 15$, the contours flatten and the admissible region widens, so a broader set of (α_0, β_0) pairs satisfies $L_1(\xi) < 1$. When $\lambda_0 = 250$ (the near-Gaussian case), the contours are almost horizontal and invertibility depends essentially only on β_0, which aligns perfectly with the fact that the Beta-*t*-GARCH(1,1) model converges to the classical GARCH(1,1) model, which remains invertible as long as $\beta < 1$.

6.3.2 | Asymptotic Variance of the Approximate Empirical Process

Next, guided by the invertibility regions in Figure 1, I examine the effect of estimation on the asymptotic distribution of the empirical process of the approximate noise $\hat{\mathcal{W}}_T(x, \hat{\theta}_T)$. To this end, I conduct a second numerical study with empirically relevant parameter values that satisfy the Lipschitz-type condition in Lemma 1. Specifically, I set:

$$\omega_0 = 0.01, \quad \alpha_0 = 0.015, \quad \beta_0 = 0.8, \quad \lambda_0 \in \{5, 15, 250\}.$$

These values span cases ranging from the heavy-tailed Beta-*t*-GARCH(1,1) model to its Gaussian GARCH(1,1) limiting case. Similar parameter values have been used in empirical illustrations by Blasques et al. (2018). The asymptotic variances of the empirical process in (26), specialized to the Beta-*t*-GARCH(1, 1) model, for both the approximated noise and the true (unobserved) noise are shown in Figure 2. The full closed-form expression, including $\boldsymbol{H}(x)$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}(x)$, is provided in Appendix C.

Figure 2 compares the asymptotic variance of the empirical process based on the approximated noise with that for the true (unobserved) noise, for $\lambda_0 \in \{5, 15, 250\}$. The discrepancy is largest under heavy tails ($\lambda_0 = 5$) and shrinks steadily as λ_0 increases, with an almost perfect match in the near-Gaussian case $\lambda_0 = 250$. This reflects the dependence of $H(x)$ and $\Gamma(x)$ on the score and Fisher information of the Student's *t* density: when λ_0 is small, heavy tails amplify the interaction term, whereas as $\lambda_0 \to \infty$ the *t* distribution converges to the normal and the approximation becomes essentially exact.

Finally, to determine how large are the two estimation-driven pieces in the asymptotic variance in (26) in practice, recall the three variance contributions:

$$\begin{aligned} V_{BIN}(x) &= F(x)(1 - F(x)), \\ V_{EST}(x) &= \boldsymbol{H}(x)^\top \boldsymbol{\Sigma} \boldsymbol{H}(x), \\ \text{and} \quad V_{INT}(x) &= 2\boldsymbol{H}(x)^\top \boldsymbol{\mathcal{I}}^{-1} \boldsymbol{\Gamma}(x). \end{aligned} \tag{45}$$

The three panels of Figure 3 display the three terms in (45) for the Beta-*t*-GARCH(1, 1) model where, as before, the noise follow Student's *t* distributions with $\lambda_0 = \{5, 15, 250\}$. All other parameters and the sample size coincide with those used to produce Figure 2 above, so the graphs are strictly comparable. For clarity, the binomial benchmark $V_{BIN}(x)$ and the interaction piece $V_{INT}(x)$ are plotted on the left axis (purple scale), whereas the

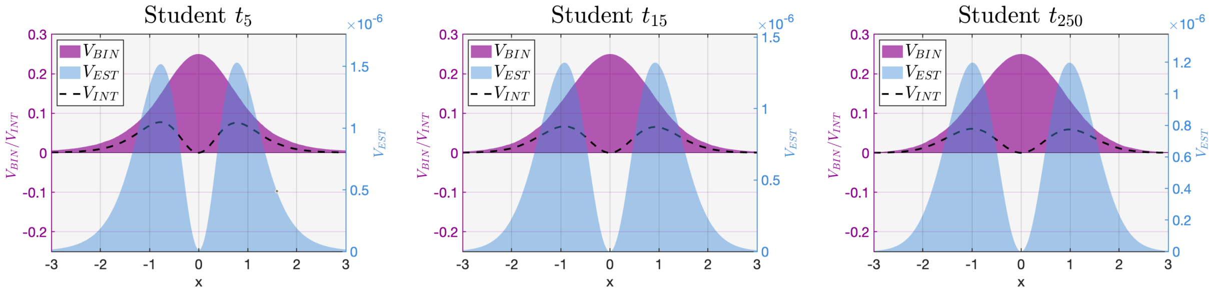

FIGURE 3 | Asymptotic variance terms of the empirical process for the approximated noise processes in the Beta-t-GARCH(1,1) model, where the noise follows a Student's t distribution with $\lambda_0 = \{5, 15, 250\}$. In the three panels, the shaded areas represent the non-negative terms $V_{BIN}(x)$ (purple) and $V_{EST}(x)$ (light blue), while the black dashed line is the signed interaction term $V_{INT}(x)$. The left ordinate (magenta and black) measures $V_{BIN}(x)$ and $V_{INT}(x)$; the right ordinate (light blue) measures $V_{EST}(x)$. All panels share the same horizontal scale $|x| \leq 3$.

estimation layer $V_{EST}(x)$ uses the right axis (blue scale, 10^{-6} factor).

With heavy tails ($\lambda_0 = 5$) the interaction term is negligible at the origin, then becomes positive in the shoulders. $V_{INT}(x)$ peaks around 0.09 at $|x| \cong 1$, implying $V_{INT}(x) \approx 0.09$ and inflating the variance in that region from 0.25 to roughly 0.26. $V_{INT}(x)$ returns to zero for $|x| \gtrapprox 2.2$, so the binomial curve governs the outer flanks. The estimation layer never exceeds 1.6×10^{-6}, six orders of magnitude smaller than $V_{BIN}(0)$. For $\lambda_0 = 15$ the maximum of $V_{INT}(x)$ falls to about 0.03, giving $V_{INT}(x) \lessapprox 0.06$ and a maximum total variance near 0.25. $V_{INT}(x)$ is already below 0.01 for $|x| \gtrapprox 1.7$, so the departure from the binomial envelope is modest and confined to a narrow band around the center, while $V_{EST}(x)$ remains at the 10^{-6} scale. When $\lambda_0 = 250$ (almost Gaussian) the dashed profile stays below 0.01 for all $|x| \leq 2$, so $V_{INT}(x) < 0.01$ and the binomial envelope is a good approximation throughout. $V_{EST}(x)$ dips slightly further, never exceeding 1.2×10^{-6}.

Because $\boldsymbol{\Sigma} = O(T^{-1})$ with $T = 25,000$ and the score-driven filter keeps $\|\boldsymbol{H}(x)\|$ bounded, the estimation layer is negligible in every scenario, whereas the interaction component supplies the only material deviation from binomial sampling variance, and its magnitude declines monotonically as the noise distribution approaches Gaussianity. This behavior exemplifies the discussion following (26) and Remarks 7 and 8.

6.3.3 | Empirical Size and Power of the KS-Type Tests

This subsection evaluates, via Monte Carlo experiments, the finite-sample performance of the classical KS statistic advocated in Harvey (2013), 54–55, and of the Khmaladze transformed KS statistic developed in Section 5. Both tests are applied to the PITs $\{\hat{U}_t\}_{t=1}^{T}$ of the fitted Beta-t-GARCH(1,1) model.

In each replication, the classical KS statistic is $\hat{\mathcal{B}}_T^{\star} := \sup_{0 \leq r \leq 1} \left| \hat{\mathcal{K}}_T(r) \right|$ with $\hat{\mathcal{K}}_T(r)$ as defined in (28). It is compared with the Kolmogorov (Brownian bridge) critical value at level α (see Simard and L'Ecuyer (2011)). The Khmaladze statistic is $\hat{\mathcal{B}}_T := \sup_{0 \leq r \leq 1} \left| \hat{\mathcal{M}}_T(r) \right|$ defined in (32) with $\hat{\mathcal{M}}_T(r)$ given in (31). It is computed using the discrete approximation described in Appendix D. By its Brownian motion limit, $\hat{\mathcal{B}}_T$ is compared with the Brownian motion critical value at level α reported in Table 2.

6.3.3.1 | Empirical Size.

The first experiment assesses the empirical size of the two procedures under the correctly specified Beta-t-GARCH$(1,1)$ null. I simulate 10,000 independent time series samples of size $T \in \{500, 1000, 5000\}$ from the DGP described in Section 6, with the same parameters as those used in Section 6.3.2, fit the null model by ML, and compute the corresponding test statistics. The empirical rejection frequencies, expressed in percentage, are reported in Table 3.

KS$_K$ tracks the nominal levels closely across all sample sizes and tail scenarios, with small deviations that shrink as T increases. By contrast, the classical KS exhibits noticeable size distortions that depend on tail thickness and sample size: under heavy tails it tends to under-reject, whereas for moderate-to-light tails it can over-reject, especially at the 1% level. These patterns are consistent with estimation-induced drift in the PIT empirical process, which is left uncorrected by the classical KS but removed by the Khmaladze projection, so that KS$_K$ attains near-nominal size even in relatively small samples.

6.3.3.2 | Empirical Power.

The second experiment evaluates power against a stochastic volatility (SV) alternative with generalized error (GED) noise process. This alternative differs from the null both in the dynamics and in the noise distribution family, while remaining symmetric. I generate data from an SV-GED model and, as in the size experiment, estimate the null Beta-t-GARCH$(1,1)$ on each replication.

Formally, the DGP is

$$\begin{aligned} y_t &= \sqrt{f_t}(\theta_0) z_t, \\ \log f_{t+1}(\theta_0) &= \mu_0 + \phi_0 \big(\log f_t(\theta_0) - \mu_0\big) + \sigma_{\eta,0}\eta_{t+1}, \\ \eta_{t+1} &\overset{IID}{\sim} \mathcal{N}(0,1), \end{aligned}$$

where $\{z_t\}_{t \in \mathbb{Z}}$ are *IID* standardized GED(δ_0) noise process with mean 0 and variance 1, independent of $\{\eta_t\}_{t \in \mathbb{Z}}$. A convenient parametric form is

$$p(z_t, \delta_0) = \frac{\delta_0}{2 b_{\delta_0} \Gamma\left(\frac{1}{\delta_0}\right)} e^{-\left|z_t / b_{\delta_0}\right|^{\delta_0}}, \quad b_{\delta_0} = \sqrt{\frac{\Gamma\left(\frac{1}{\delta_0}\right)}{\Gamma\left(\frac{3}{\delta_0}\right)}},$$

TABLE 3 | Empirical rejection frequencies (%) for the classical KS test Harvey (2013) and the Khmaladze KS test (KS_K) proposed in Section 5 under the Beta-t-GARCH(1, 1) null.

	$\alpha = 10\%$		$\alpha = 5\%$		$\alpha = 1\%$	
T	**KS**	$\mathbf{KS}_K$	**KS**	$\mathbf{KS}_K$	**KS**	$\mathbf{KS}_K$
Scenario A:$\lambda_0 = 5$						
500	6.44	9.28	2.97	5.15	0.25	1.45
1000	8.75	9.62	3.14	4.98	0.79	1.07
5000	9.73	9.89	4.21	5.01	0.87	1.01
Scenario B:$\lambda_0 = 15$						
500	7.90	8.37	3.04	4.30	3.10	1.27
1000	8.02	8.98	4.00	4.70	1.35	1.07
5000	8.55	9.53	4.60	5.00	0.97	1.05
Scenario C:$\lambda_0 = 250$						
500	8.28	8.97	4.00	4.65	2.45	1.58
1000	8.80	9.06	4.28	4.80	2.01	1.04
5000	8.86	9.03	4.57	4.97	1.89	1.03

Note: Three degrees-of-freedom scenarios for λ_0. Results from 10,000 Monte Carlo replications.

TABLE 4 | Empirical rejection frequencies (%) for the classical KS test Harvey (2013) and the Khmaladze KS test (KS_K) proposed in Section 5.

	$\alpha = 10\%$		$\alpha = 5\%$		$\alpha = 1\%$	
T	**KS**	$\mathbf{KS}_K$	**KS**	$\mathbf{KS}_K$	**KS**	$\mathbf{KS}_K$
Scenario A:$\delta_0 = 1$						
500	80.8	88.0	70.5	86.0	45.4	68.2
1000	91.5	96.4	79.3	94.1	63.2	76.4
5000	96.0	99.8	92.1	99.8	88.9	97.7
Scenario B:$\delta_0 = 1.5$						
500	22.0	47.5	12.2	28.0	7.40	10.0
1000	38.4	56.1	35.7	47.1	12.3	15.8
5000	55.3	75.4	51.4	63.2	20.1	29.3
Scenario C:$\delta_0 = 3$						
500	25.0	26.4	15.0	15.1	10.0	10.0
1000	48.3	65.3	40.7	59.4	17.0	17.1
5000	74.9	78.0	52.3	70.1	23.2	23.4

Note: DGP: stochastic volatility with generalized error (GED) noise (SV-GED), standardized to mean 0 and variance 1. Fitted model under the null: Beta-t-GARCH(1, 1). Three tail-shape scenarios for the GED parameter $\delta_0 \in \{1, 1.5, 3\}$; the unconditional variance is matched to the null. Results from 10,000 Monte Carlo replications.

which yields $\mathbb{V}[z_t] = 1$ for any $\delta_0 > 0$ (the cases $\delta_0 = 2$ and $\delta_0 = 1$ correspond to the normal and Laplace laws, respectively, with $\delta_0 < 2$ heavier-than-Gaussian and $\delta_0 > 2$ lighter-than-Gaussian).

I initialize $\log f_1 = \mu_0$ (so $f_1 = e^{\mu_0}$). Since $\mathbb{V}[z_t] = 1$, the unconditional variance of y_t equals

$$\mathbb{V}[y_t] = \mathbb{E}[f_t(\theta_0)] = e^{\mu_0 + \frac{\sigma^2_{\eta,0}}{2(1-\phi_0^2)}},$$

and I choose μ_0 so that $\mathbb{E}[f_t]$ matches the unconditional variance under the null. In the experiments, I consider several tail shapes $\delta_0 \in \{1, 1.5, 3\}$ to span heavy, near-normal, and light tails. The empirical rejection frequencies are reported in Table 4.

Across all tail scenarios, the Khmaladze-based statistic KS_K dominates or matches the classical KS in power. The gains are most pronounced for the strongest departure ($\delta_0 = 1$), where KS_K delivers substantially higher rejection frequencies at all significance levels and sample sizes; the advantage remains clear for $\delta_0 = 1.5$ and is still present, though smaller, for $\delta_0 = 3$. Within each (δ_0, α) configuration, power increases monotonically with T, as expected. Overall, these experiments indicate that, relative to the classical KS, the Khmaladze correction yields tests that are

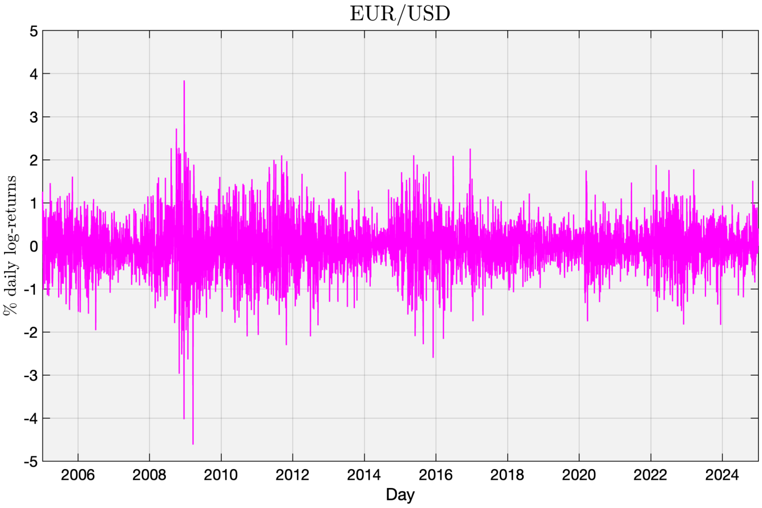

FIGURE 4 | EUR/USD daily percentage log-returns over the sample 2005–2025.

TABLE 5 | Key summary statistics for EUR/USD daily returns ($y_t = 100(\log P_t - \log P_{t-1})$) and standardized approximate noise from the fitted Beta-t-GARCH(1, 1).

	Returns (%/day)	**Approx. noise (decimal)**
Mean	0.005	0.012
Std. dev.	0.554	0.995
Skewness	−0.131	0.019
Excess kurtosis	3.594	1.243
ACF(1)	0.013	0.016
LBQ p (lag 10)	0.093	0.758
JB p	0.001	0.001

Note: Sample size $T = 5,218$.

both better sized and uniformly more powerful against this representative dynamic alternative.

6.3.4 | Empirical Application

I use daily EUR/USD spot rates sampled on trading days from 1 January 2005 to 1 January 2025. Prices P_t are converted to percentage log-returns, $y_t = 100(\log P_t - \log P_{t-1})$, without winsorization or prefiltering. The resulting sample contains $T = 5,218$ observations. Major FX rates are a standard benchmark setting for conditional volatility models: they are highly liquid, free of corporate actions, and typically display (i) near-zero unconditional mean, (ii) weak linear autocorrelation in returns but marked volatility clustering, and (iii) symmetric yet heavy-tailed noise. These features make EUR/USD a natural environment for the Beta-t-GARCH(1,1) benchmark and for assessing goodness-of-fit tests for the one-step predictive distribution. The series and returns are shown in Figure 4; key summary statistics appear in Table 5, and the ML estimates $\hat{\theta}_T$ are reported in Table 6.

TABLE 6 | ML estimates for the Beta-t-GARCH(1, 1) on EUR/USD daily returns.

$\hat{\omega}_T$	$\hat{\alpha}_T$	$\hat{\beta}_T$	$\hat{\lambda}_T$
0.0010	0.0452	0.9557	7.8784
(0.0001)	(0.0125)	(0.0374)	(0.1158)

Note: Standard errors are in parenthesis.

Basic summary statistics for y_t (percent per day) and the standardized approximate noise $\hat{\epsilon}_t(\hat{\theta}_T)$ in Table 5 align with the usual stylized facts: mean 0.0052, standard deviation 0.554, skewness −0.131, and excess kurtosis 3.594 ($T = 5218$). A Ljung–Box test at lag 10 yields a p-value of 0.093 on returns (mild serial dependence), while the Jarque–Bera test rejects Gaussianity (heavy tails). After standardization, linear dependence weakens markedly (LBQ $p = 0.758$) though non-Gaussianity persists (excess kurtosis ≈ 1.243), providing an informative backdrop for the KS and KS_K goodness-of-fit analysis that follows.

Given the MLE $\hat{\theta}_T = \left(\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T, \hat{\lambda}_T\right)^\top$ of the Beta-t-GARCH(1,1) model in Table 6, define the approximate noise $\hat{\epsilon}_t(\hat{\theta}_T) = y_t / \sqrt{\hat{f}_t(\hat{\theta}_T)}$ and the PITs

$$\hat{U}_t = F_{t_{\hat{\lambda}_T}}\left(\frac{\hat{\epsilon}_t(\hat{\theta}_T)}{c(\hat{\lambda}_T)}\right), \quad c(\hat{\lambda}_T) = \sqrt{\frac{\hat{\lambda}_T - 2}{\hat{\lambda}_T}},$$

where $F_{t_{\hat{\lambda}_T}}$ denotes the distribution function of a Student's t with $\hat{\lambda}_T$ degrees of freedom. For these empirical estimates, the asymptotic conditions in Section 6 are still satisfied. A numerical evaluation of the log-contraction condition (38) at $\hat{\theta}_T$ yields a small negative value (about -5×10^{-4}), so the corresponding Beta-t-GARCH(1,1) recursion is strictly stationary but highly persistent, in line with $\hat{\alpha}_T + \hat{\beta}_T \approx 1.0009$. In addition, the point $(\hat{\alpha}_T, \hat{\beta}_T)$ lies in the admissible region suggested by Figure 1, and the corresponding value of $L_1(\xi)$ at $(\hat{\alpha}_T, \hat{\beta}_T, \hat{\lambda}_T)$ is about 0.9945, so the invertibility condition $L_1(\xi) < 1$ in (41) is also compatible with the estimated model. For the EUR/USD daily returns considered here, the theoretical results in Corollaries 3.2 to 3.4 therefore apply.

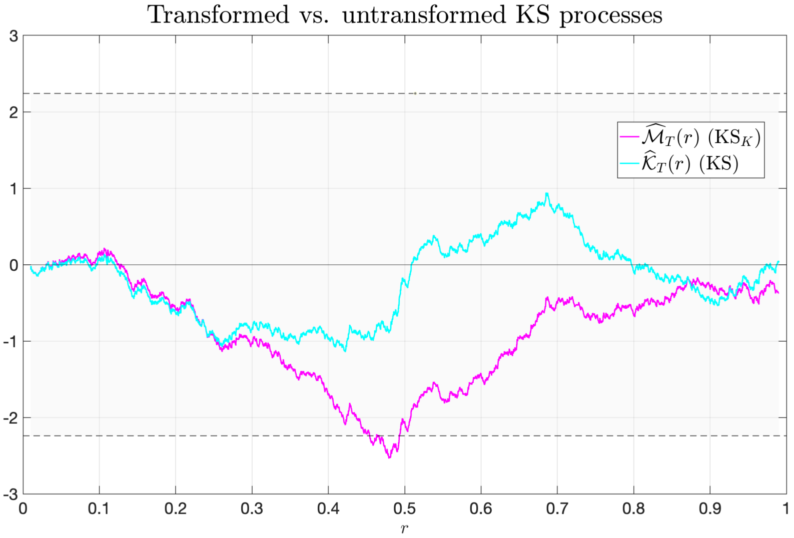

FIGURE 5 | Empirical KS process [$\hat{\mathcal{K}}_T(r)$] and Khmaladze transformed process [$\hat{\mathcal{M}}_T(r)$] for EUR/USD daily returns (2005–2025). Dashed lines denote the 95% Brownian motion band on $[0,1]$. The transformed path [$\hat{\mathcal{M}}_T(r)$] crosses the band, leading to rejection at the 5% level.

On this basis, I compute (i) the classical KS statistic for the empirical distribution of $\{\hat{U}_t\}$, whose limit is a Brownian bridge, and (ii) the Khmaladze KS$_K$ statistic, that is, the sup-norm of the martingale-transformed process, whose limit is standard Brownian motion (cf. Section 5). Alongside decisions based on the usual asymptotic critical values, I also report parametric bootstrap p-values obtained by simulating from the fitted model, re-estimating on each replication (R), and recomputing the test statistics (here $R = 10,000$).

Both the untransformed process $\hat{\mathcal{K}}_T(r)$ and the transformed process $\hat{\mathcal{M}}_T(r)$ are plotted in Figure 5. The two horizontal lines show the 95% simultaneous band for a standard Brownian motion on $[0,1]$, namely ± 2.241 (appropriate for KS$_K$). Since $\hat{\mathcal{M}}_T(r)$ crosses this band, the Khmaladze KS test rejects at the 5% level. Numerically, $\hat{\mathcal{B}}_T = 2.699 > 2.241$. Hence the null that the PITs $\left\{\hat{U}_t\right\}_{t=1}^{T}$ are *IID* $\boldsymbol{\mathcal{U}}(0,1)$ under the fitted Beta-t-GARCH(1, 1) is rejected. (The classical KS decision uses Brownian bridge critical values.)

Complementing Figure 5, a size check under the fitted Beta-t-GARCH(1, 1) (parametric bootstrap with re-estimation, $R = 10,000$) shows that KS$_K$ attains the nominal 5% level (empirical size 0.050), whereas the classical KS is conservative (empirical size 0.020). On the EUR/USD sample (2005–2025), the parametric bootstrap p-values are 0.010 for KS$_K$ and 0.110 for KS. Thus KS$_K$ rejects the null that the PITs $\left\{\hat{U}_t\right\}_{t=1}^{T}$ are *IID* $\boldsymbol{\mathcal{U}}(0,1)$ under the fitted model, while the classical KS does not, in line with the theory and the simulation evidence.

7 | Conclusion and Future Work

In this article, I examined the filtering approximation-theoretic properties of score-driven models and analyzed the asymptotic behavior of both the empirical distribution function and the empirical process of the approximated noise. Theoretical results were derived under Lipschitz-type and tail conditions, leading to sharp upper bounds on the filtering approximation error. These bounds are fundamental for understanding the asymptotic properties of the empirical distribution function as well as the limiting behavior of the empirical process of the approximated noise. The assumptions imposedon the score-driven model are relatively weak and remain compatible with specifications featuring bounded updates, provided that the driving force $s_t(f_t(\theta_0), \lambda_0)$ and its derivatives have bounded Orlicz norm.

In Francq and Zakoïan (2022), the empirical process of the approximated noise is AMF and AED under their baseline assumptions for pure conditional scale models, and an AMD term appears only when their stability-by-scaling condition is relaxed or when a conditional mean is introduced. In the fully parametric score-driven setting analyzed here, the same parameter vector enters both the noise density and the recursion, so the empirical process is generically AMD and AED. The two theories coincide only in the Gaussian GARCH(1,1) special case. This model-dependence complicates goodness-of-fit assessment, but by exploiting the filter approximation error and the empirical distribution of the approximated noise I outline a feasible testing approach.

Beyond the univariate case, an important future direction is the extension of this theoretical framework to multivariate score-driven models, where an N-dimensional vector process is accompanied by an N-dimensional vector of time-varying parameters. In such models, the problem of filter invertibility becomes significantly more complex, as the invertibility region determined by Bougerol's SRE theory and Assumption 1 may be too small or impractical for real-world applications. A promising strategy to address this issue is the use of k-fold contractions, where k iterations of the recurrence equation for the parameter process $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ are considered. This technique has already proved effective in establishing strict stationarity, ergodicity, and moment existence, as demonstrated by D'Innocenzo et al. (2023) for a robust score-driven multivariate model and by D'Innocenzo et al. (2024b) for a time-varying tail-shape bivariate process.

A natural next step is to generalize the Lipschitz and tail conditions presented in this article by incorporating k-fold contractions instead of single-step contractions. This approach could facilitate the development of non-degenerate invertibility regions and lead to sharper upper bounds on filtering and residual approximation errors in multivariate score-driven models.

Acknowledgments

Open access publishing facilitated by Universita degli Studi di Bologna, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

Data and codes are available from the corresponding author upon request. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] For an updated overview of research on score-driven models, see the articles section on www.gasmodel.com.

References

Bai, J. 2003. "Testing Parametric Conditional Distributions of Dynamic Models." *Review of Economics and Statistics* 85, no. 3: 531–549.

Berkes, I., S. Hörmann, and L. Horváth. 2008. "The Functional Central Limit Theorem for a Family of Garch Observations With Applications." *Statistics & Probability Letters* 78, no. 16: 2725–2730.

Berkes, I., S. Hörmann, and J. Schauer. 2009. "Asymptotic Results for the Empirical Process of Stationary Sequences." *Stochastic Processes and Their Applications* 119, no. 4: 1298–1324.

Berkes, I., S. Hörmann, and J. Schauer. 2011. "Split Invariance Principles for Stationary Processes." *Annals of Probability* 39, no. 6: 2441–2473.

Berkes, I., and L. Horváth. 2003. "Limit Results for the Empirical Process of Squared Residuals in Garch Models." *Stochastic Processes and Their Applications* 105, no. 2: 271–298.

Beutner, E. A., Y. Lin, and A. Lucas. 2023. *Consistency, Distributional Convergence, and Optimality of Score-Driven Filters, Technical Report*. Tinbergen Institute Discussion Paper.

Blasques, F., P. Gorgi, and S. J. Koopman. 2019. "Accelerating Score-Driven Time Series Models." *Journal of Econometrics* 212, no. 2: 359–376.

Blasques, F., P. Gorgi, S. J. Koopman, and O. Wintenberger. 2018. "Feasible Invertibility Conditions and Maximum Likelihood Estimation for Observation-Driven Models." *Electronic Journal of Statistics* 12: 1019–1052.

Blasques, F., S. J. Koopman, and A. Lucas. 2014. "Stationarity and Ergodicity of Univariate Generalized Autoregressive Score Processes." *Electronic Journal of Statistics* 8, no. 1: 1088–1112.

Blasques, F., S. J. Koopman, and A. Lucas. 2015. "Information-Theoretic Optimality of Observation-Driven Time Series Models for Continuous Responses." *Biometrika* 102, no. 2: 325–343.

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas. 2022. "Maximum Likelihood Estimation for Score-Driven Models." *Journal of Econometrics* 227, no. 2: 325–346.

Boldin, M. V. 1983. "Estimation of the Distribution of Noise in an Autoregression Scheme." *Theory of Probability and Its Applications* 27, no. 4: 866–871.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31, no. 3: 307–327.

Boucheron, S., G. Lugosi, and P. Massart. 2013. "Concentration Inequalities: A Nonasymptotic Theory of Independence." In *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Bougerol, P. 1993. "Kalman Filtering With Random Coefficients and Contractions." *SIAM Journal on Control and Optimization* 31, no. 4: 942–959.

Casella, G., and R. Berger. 2002. *Statistical Inference*. Vol. 2. Duxbury Resource Center.

Creal, D., S. J. Koopman, and A. Lucas. 2011. "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations." *Journal of Business & Economic Statistics* 29, no. 4: 552–563.

Creal, D., S. J. Koopman, and A. Lucas. 2013. "Generalized Autoregressive Score Models With Applications." *Journal of Applied Econometrics* 28, no. 5: 777–795.

Dedecker, J., and X. Fan. 2015. "Deviation Inequalities for Separately Lipschitz Functionals of Iterated Random Functions." *Stochastic Processes and Their Applications* 125, no. 1: 60–90.

Dehling, H., T. Mikosch, and M. Sörensen. 2002. *Empirical Process Techniques for Dependent Data*. Springer.

D'Innocenzo, E., A. Luati, and M. Mazzocchi. 2023. "A Robust Score-Driven Filter for Multivariate Time Series." *Econometric Reviews* 42, no. 5: 441–470.

D'Innocenzo, E., A. Lucas, B. Schwaab, and X. Zhang. 2024a. *Joint Extreme Value-At-Risk and Expected Shortfall Dynamics With a Single Integrated Tail Shape Parameter. Technical Report*. Tinbergen Institute Discussion Paper.

D'Innocenzo, E., A. Lucas, B. Schwaab, and X. Zhang. 2024b. "Modeling Extreme Events: Time-Varying Extreme Tail Shape." *Journal of Business & Economic Statistics* 42, no. 3: 903–917.

Doukhan, P., and O. Wintenberger. 2008. "Weakly Dependent Chains With Infinite Memory." *Stochastic Processes and Their Applications* 118, no. 11: 1997–2013.

Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation." *Econometrica: Journal of the Econometric Society* 50: 987–1007.

Erdös, P., and M. Kac. 1946. "On Certain Limit Theorems of the Theory of Probability." *Bulletin of the American Mathematical Society* 52, no. 12: 292–302.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications. Vol. II*. John Wiley & Sons Inc.

Francq, C., and J. Zakoian. 2019. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Francq, C., and J.-M. Zakoian. 2023. "Local Asymptotic Normality of General Conditionally Heteroskedastic and Score-Driven Time-Series Models." *Econometric Theory* 39, no. 5: 1067–1092.

Francq, C., and J.-M. Zakoïan. 2022. "Adaptiveness of the Empirical Distribution of Residuals in Semi-Parametric Conditional Location Scale Models." *Bernoulli* 28, no. 1: 548–578.

Gorgi, P., C. Lauria, and A. Luati. 2024. "On the Optimality of Score-Driven Models." *Biometrika* 111, no. 3: 865–880.

Granger, C. W., and A. Andersen. 1978. "On the Invertibility of Time Series Models." *Stochastic Processes and Their Applications* 8, no. 1: 87–92.

Hart, J. D., and P. Vieu. 1990. “Data-Driven Bandwidth Choice for Density Estimation Based on Dependent Data.” *Annals of Statistics* 18, no. 2: 873–890.

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

Harvey, A. C., and A. Luati. 2014. “Filtering With Heavy Tails.” *Journal of the American Statistical Association* 109, no. 507: 1112–1122.

Hörmann, S. 2008. “Augmented Garch Sequences: Dependence Structure and Asymptotics.” *Bernoulli* 14, no. 2: 543–561.

Horváth, L., and R. Zitikis. 2006. “Testing Goodness of Fit Based on Densities of GARCH Innovations.” *Econometric Theory* 22, no. 3: 457–482.

Khmaladze, E. V. 1981. “Martingale Approach in the Theory of Goodness-Of-Fit Tests.” *Theory of Probability and Its Applications* 26, no. 2: 240–257.

Kosorok, M. 2008. *Introduction to Empirical Processes and Semiparametric Inference*, Springer Series in Statistics. Springer.

Koul, H. L. 1992. *Weighted Empiricals and Linear Models*, Lecture Notes-Monograph Series. Vol. 21, i–264. IMS.

Koul, H. L., and S. Ling. 2006. “Fitting an Error Distribution in Some Heteroscedastic Time Series Models.” *Annals of Statistics* 34, no. 2: 994–1012.

Kulperger, R., and H. Yu. 2005. “High Moment Partial Sum Processes of Residuals in GARCH Models and Their Applications.” *Annals of Statistics* 33, no. 5: 2395–2422.

Pearson, E. S. 1938. “The Probability Integral Transformation for Testing Goodness of Fit and Combining Independent Tests of Significance.” *Biometrika* 30, no. 1/2: 134–148.

Rio, E. 2018. *Asymptotic Theory of Weakly Dependent Random Processes*, Probability Theory and Stochastic Modelling. Springer.

Simard, R., and P. L’Ecuyer. 2011. “Computing the Two-Sided Kolmogorov-Smirnov Distribution.” *Journal of Statistical Software* 39, no. 11: 1–18.

Straumann, D., and T. Mikosch. 2006. “Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach.” *Annals of Statistics* 34, no. 5: 2449–2495.

van der Vaart, A., and J. Wellner. 2023. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Wintenberger, O. 2013. “Continuous Invertibility and Stable Qml Estimation of the EGARCH(1,1) Model.” *Scandinavian Journal of Statistics* 40, no. 4: 846–867.

Wu, W. B., and J. Mielniczuk. 2002. “Kernel Density Estimation for Linear Processes.” *Annals of Statistics* 30, no. 5: 1441–1459.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1.** Supporting Information.