Realized Dynamic Score Exponential GARCH

Tibor Pál Giuseppe Storti Kok Haur Ng

October 12, 2026

Abstract

We introduce the realized dynamic score exponential GARCH model, which jointly models returns and a realized range-based volatility measure using heavy-tailed and skewed conditional distributions. Beyond embedding a score-driven framework within the volatility dynamics, the model provides an enhanced treatment of leverage effects by specifying leverage functions in both quadratic and score-driven forms and incorporating them into both the measurement and volatility equations. The score-driven dynamics linking returns, realized ranges, and future volatility, together with flexible conditional distributions, yield substantial improvements in empirical fit relative to competing specifications. In addition, we introduce a two-component volatility structure to capture the highly persistent, long-memory behavior of return volatility. Empirical analysis based on the Standard & Poor's (S&P) 500 and three individual asset return series shows that the proposed model delivers superior in-sample fit and out-of-sample VaR and ES forecasting performance.

Keywords: realized EGARCH, score-driven models, component volatility models, Value-at-Risk, Expected Shortfall

1 Introduction

Modeling volatility is central to financial econometrics, as understanding volatility dynamics is fundamental to asset pricing, risk management, and forecasting. Volatility clustering, the key stylized fact that motivated the development of the canonical autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and generalized autoregressive conditional heteroskedasticity (GARCH) models (Bollerslev, 1986), is naturally captured within the baseline formulations of this modeling framework. However, these models often provide a limited account of other empirical features of financial returns, such as heavy tails, skewness, leverage effects, and long memory in volatility, particularly under non-standard market conditions. Moreover, rapid shifts in volatility can cause standard GARCH-type models to adjust slowly to new volatility regimes.

This study addresses these challenges jointly. To this end, we introduce a variant of the realized GARCH model that uses score-driven updating dynamics to link

the return process and realized range volatility jointly, with volatility dynamics driven by scores from both the return and measurement equations. We call this specification the realized dynamic score EGARCH (RDS-EGARCH) model.

The proposed framework makes several contributions. First, it jointly exploits information from returns and realized range-based volatility within a unified score-driven updating mechanism. Second, it incorporates a two-component volatility structure to more effectively capture persistence. Third, it extends leverage-effect modeling by including both quadratic and score-driven terms in the return and measurement equations. Finally, it combines these features with flexible heavy-tailed and skewed conditional distributions, enhancing robustness to extreme observations.

Early contributions in the literature propose a broad range of model extensions that address specific empirical features of volatility dynamics, but typically in isolation. Bollerslev (1987) incorporated heavy tails into GARCH models, and Hansen (1994) later generalized them to allow asymmetric tail behavior. Recognizing the asymmetric impact of returns on conditional volatility, leverage effects were introduced through the EGARCH model of Nelson (1991) and the Glosten–Jagannathan–Runkle (GJR) GARCH model of Glosten et al. (1993), among others. Persistence in volatility was further addressed by introducing dynamic component structures, most notably in the seminal work of Ding and Granger (1996), which showed how component-based models can approximate long-memory behavior. Relatedly, Baillie et al. (1996) formalized long-memory dynamics explicitly within the fractionally integrated GARCH (FIGARCH) framework. However, component models, such as that of Ding and Granger (1996), provide a more tractable tool for modeling persistent volatility dynamics and overcome the statistical difficulties implied by fractionally integrated dynamic models.

Although these frameworks provide essential building blocks for researchers and practitioners, they adapt slowly to volatility changes during periods of market stress. This limitation primarily stems from their reliance on a reduced set of information. The development of the score-driven framework (Creal et al., 2008) marked an important milestone in this respect. Specifically, when dynamics depend on the conditional score rather than specific conditional moments, and the score is derived from a heavy-tailed density, the resulting volatility update reacts more strongly and quickly to large but admissible shocks under that density. Moreover, Harvey and Sucarrat (2014) showed that combining score-driven dynamics with a skew-t distribution yields additional robustness by downweighting observations that would be considered outliers under Gaussianity.

Despite the substantial advances offered by the score-driven framework, its performance remains limited when the updating equation relies solely on an individual signal, as a single return observation conveys little information about prevailing volatility. Consequently, as Andersen et al. (2003) note, return-based filters often fail to capture abrupt shifts in the true volatility level. With the emergence of realized volatility measures based on high-frequency data, Hansen et al. (2012) introduced the realized GARCH framework for modeling returns and volatility jointly. By incorporating a measurement equation that links the realized measure to the conditional variance, the framework expands the information set used to update volatility dynamics.

Building on evidence that realized measures are often distorted by jumps (Opschoor et al., 2018), Harvey and Palumbo (2023) extended the score-driven methodology to model realized volatility. Using members of the exponential generalized

beta of the second kind (EGB2) distribution family, the model accommodates these distortions and achieves robust updating through the conditional score. Although this approach provides a flexible, practical framework for modeling realized volatility on its own, it does not use information embedded in the return process. In addition, a model for realized volatility alone cannot generate tail risk forecasts in a fully coherent manner unless embedded in a joint return–volatility specification, which limits its applicability in risk-management settings.

Against this background, we propose the realized dynamic score EGARCH (RDS-EGARCH) model to model asset returns and realized range volatility jointly. The score-driven framework provides a systematic approach to updating volatility dynamics based on the conditional density score, enabling a closed-form likelihood and efficient estimation.

Within this framework, we incorporate leverage effects, persistence, and distributional flexibility through augmented score-driven terms, a two-component volatility structure, and heavy-tailed, skewed conditional distributions.

Finally, we demonstrate the empirical relevance of the proposed model through applications to the Standard & Poor's (S&P) 500 index and three individual stocks, AAPL, IBM, and BAC. The results indicate that our specification improves in-sample fit and delivers superior out-of-sample performance in forecasting value-at-risk (VaR) and expected shortfall (ES) relative to benchmark beta-skew-t-EGARCH variants (Harvey and Sucarrat, 2014), Realized GARCH (Hansen et al., 2012), and the Realized EGARCH specification with a single realized measure (Hansen and Huang, 2016). These findings highlight the benefits of incorporating realized measures and flexible error distributions within a score-driven EGARCH framework for financial volatility modeling.

The remainder of the paper is organized as follows. Section 2 presents the statistical framework of the proposed model, focusing on volatility dynamics with leverage effects, score-driven updating, and estimation. Section 3 reports the empirical analysis and model comparison, with particular emphasis on forecast evaluation. Section 4 concludes the paper.

2 The statistical framework

The proposed model builds on the realized GARCH framework of Hansen et al. (2012). This framework jointly models returns and realized volatility, linking the realized measure to the latent conditional volatility through a measurement equation. As a result, it explicitly accounts for measurement errors in realized measures and allows conditional volatility to react more promptly and smoothly to information extracted from high-frequency data. For ease of presentation, we will start with a simple baseline version of the RDS-EGARCH model and then gradually deepen our analysis of its components.

The structure of the baseline RDS-EGARCH model takes the form

$$r_t = \alpha_t z_t = \exp(\lambda_t) z_t \tag{1}$$

$$y_t = \omega_x \alpha_t^{\rho} \varepsilon_t = \omega_x \exp(\rho \lambda_t) \varepsilon_t, \tag{2}$$

where r_t is the log-return of an asset at time t, that is, $r_t = \ln(P_t) - \ln(P_{t-1})$, with P_t being the price at time t; y_t is a realized volatility measure; and $\alpha_t = \exp(\lambda_t)$ is a time-varying scale parameter that is assumed to be a function of the information set up to time $t-1$, $\mathcal{I}_{t-1}$.

In the baseline specification of the RDS-EGARCH model, the dynamics are driven by the two mutually independent error terms, $z_t \sim$ i.i.d. $F(\mu_z)$, and $\varepsilon_t \sim$ i.i.d. $G(\mu_\varepsilon)$, where $F(\mu)$ denotes a distribution with mean μ and G is a distribution with positive support. For ease of exposition, we impose $\mathbb{E}[r_t \mid \mathcal{I}_{t-1}] = 0$, consistent with treating z_t as a zero mean innovation.

The parameters ω_x and ρ are scaling parameters that align the level of the realized measure with the latent volatility. Namely, they correct for the systematic scaling bias (e.g., Christensen and Podolskij, 2007) of the realized measure and allow us to isolate the dynamics in λ_t, ensuring that ε_t reflects only pure innovations. Although we do not explicitly implement bias-corrected estimators as in Christensen et al. (2009), the scaling parameters, together with the score-driven updating mechanism, can partially account for persistent distortions arising from market microstructure noise in the level of the realized measure. This does not replace explicit bias correction, and incorporating adjusted realized measures is a natural extension for future research.

The following subsections provide more insight into the structure of the model, focusing on the distribution of the error components (Sec. 2.1), the maximum likelihood (ML) estimation procedure (Sec. 2.2), the dynamic volatility model (Sec. 2.3), the treatment of the leverage effect (Sec. 2.4) and the characterization of the score-driven dynamics (Sec. 2.5).

2.1 Heavy tails and skewness

This section provides insight into the distributional framework underlying the RDS-EGARCH model presented in Eqs. (1)–(2). Our settings allow us to account for heavy tails and skewness in the conditional distribution of returns, a well-known stylized fact. To this aim, the return innovations z_t are assumed to follow a skewed Student's t distribution, $z_t \sim$ i.i.d. Skew-$t(\nu, \gamma)$, where ν ($\nu > 2$) is the degrees of freedom parameter and γ is an additional parameter controlling skewness. More specifically, the skewed t distribution is obtained through contamination of a $t(\nu)$ distribution following the methodology proposed by Fernández and Steel (1998): given a variable z whose density $f(z)$ is continuous, unimodal, and symmetric around zero, skewness is achieved by introducing inverse scale factors in the positive and the negative orthants (see Fernández and Steel, 1998, for more details). The resulting distribution can account for both heavy tails and skewness, and its density is given by

$$f(z_t \mid \gamma) = \frac{2}{\gamma + \gamma^{-1}} f\left(\frac{z_t}{\gamma^{\text{sgn}(z_t)}}\right), \tag{3}$$

where z_t becomes uncentered, with $\gamma \in (0, \infty)$ governing the degree of skewness. Symmetry occurs at $\gamma = 1$, while values of $\gamma < 1$ and $\gamma > 1$ indicate negative and positive skewness, respectively.

As observed by Giot and Laurent (2003), the conditional expectation of the demeaned return

$$\mathbb{E}_{t-1}[r_t] = \mu_z \, \exp\big(\lambda_{t|t-1}\big), \tag{4}$$

where μ_z is the uncentered mean, is not constant. Consequently, r_t cannot be a martingale difference (MD). To restore the MD property, following Harvey (2013, Sec. 4.1, p. 146), the model is reformulated as

$$r_t = -\mu_z \exp(\lambda_{t|t-1}) + \exp(\lambda_{t|t-1}) z_t, \tag{5}$$

where, given the skew t distribution assumed in the return process, μ_z is defined as

$$\mu_z = \sqrt{\nu}\left(\gamma - \frac{1}{\gamma}\right)\frac{\mathrm{B}\left(\frac{\nu-1}{2}, \frac{1}{2}\right)}{\pi} = \sqrt{\nu}\left(\gamma - \frac{1}{\gamma}\right)\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}}. \tag{6}$$

In the measurement equation, we use the Burr (or Singh-Maddala) distribution to match the positive support of the observations. The Burr distribution is a special case of the generalized beta of the second kind (GB2) density, which allows for fat-tailed behavior through its two shape parameters. Using a fat-tailed distribution is important to accommodate the outliers commonly observed in high-frequency data. Our choice of the Burr distribution is also supported by Harvey and Palumbo (2023), which identifies it as one of the best-performing specifications for modeling realized volatility in score-driven settings. The standard form of GB2, as given in Kleiber and Kotz (2003), is

$$f_{t-1}(y_t; \alpha_t, a, p, q) = \frac{a(y_t/\alpha_t)^{ap-1}}{\lambda_t \,\mathrm{B}(p,q)\,\left[(r_t/\alpha_t)^a + 1\right]^{p+q}}, \qquad 0 < y_t < \infty, \quad \alpha_t, a, p, q > 0,$$

where $\alpha_t = \exp(\lambda_t)$ is the scale parameter, a, p, and q are the shape parameters, and $\mathrm{B}(p,q)$ denotes the beta function; the notation f_{t-1} indicates conditioning on information available at time $t-1$. We use the Burr specification obtained by setting $p = 1$ in the GB2 density, which results in

$$f_{t-1}(y_t; \alpha_t, a, q) = \frac{aq(y_t/\alpha_t)^{a-1}}{\alpha_t\left[(r_t/\alpha_t)^a + 1\right]^{1+q}}, \qquad 0 < y_t < \infty, \quad \alpha_t, a, q > 0.$$

To enhance the model's flexibility, we use the exponential Singh–Maddala (ESM) distribution, obtained by applying a logarithmic transformation to the measurement equation. Consequently, Eq. 2 is now expressed as

$$x_t = \log\omega_x + \rho\lambda_t + \log\varepsilon_t, \tag{7}$$

where $x_t = \log(y_t)$, with conditional location $\log\omega_x + \rho\lambda_t$. Finally, the density function of the ESM distribution takes the form

$$f_{t-1}(x_t; \lambda_t, a, q, \omega_x, \rho) = \frac{aq\,\exp\{a(x_t - \log\omega_x - \rho\lambda_t)\}}{\left[1 + \exp\{a(x_t - \log\omega_x - \rho\lambda_t)\}\right]^{1+q}}, \qquad a, q, \rho > 0.$$

2.2 ML Estimation

This section focuses on maximum likelihood (ML) estimation of the RDS-EGARCH model in Eqs. (19)- (23). Relying on standard probability arguments and on the assumed independence of the returns and measurement errors, z_t and ϵ_t, the joint conditional distribution of (r_t, x_t) given $\mathcal{I}_{t-1}$ can be factorized as follows

$$f_{rx,t-1}(r_t, x_t) = f_{r,t-1}(r_t) f_{x|r,t-1}(x_t|r_t). \tag{8}$$

After taking logs, Eq. (8) leads to the following additive decomposition for the conditional log-likelihood

$$\ln f_{rx,t-1}(r_t, x_t) = \ell_{rx,t-1}(r_t, x_t; \theta) = \ell_{r,t-1}(r_t; \theta) + \ell_{x|r,t-1}(x_t; \theta) \tag{9}$$

where θ is the vector of static model parameters to be estimated, $\ell_{r,t-1}(r_t;\theta) = \ln f_{r,t-1}(r_t;\theta)$ and $\ell_{x,t-1}(x_t;\theta) = \ln f_{x,t-1}(x_t;\theta)$. Under distributional assumption in Sec. 2.1, the returns log-likelihood component $\ell_{r,t-1}(r_t;\theta)$ is given by

$$\begin{aligned}\ell_{r_t} = \ln(2) - \ln\left(\gamma + \frac{1}{\gamma}\right) + \ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) \\ - \frac{\ln(\pi\nu)}{2} - \lambda_t - \frac{\nu+1}{2}\ln\left(1 + \frac{(r_t + \mu_z \exp(\lambda_t))^2}{\gamma^{2\,\mathrm{sign}(r_t+\mu_z\exp(\lambda_t))}\exp(2\lambda_t)\nu}\right),\end{aligned} \tag{10}$$

while the log-likelihood component related to the measurement equation is

$$\ell_{x|r,t-1}(x_t;\theta) = \ln(a) + a\varepsilon_t + \ln(q) - (1+q)\ln(\exp(a\varepsilon_t) + 1)\,. \tag{11}$$

The ML estimator of θ is then defined as the minimizer of the log-likelihood in Eq. (9)

$$\hat{\theta} = \arg\max_{\theta}\ \ell_{rx,t-1}(r_t, x_t;\theta).$$

2.3 Scale dynamics

Unlike the standard realized GARCH framework, since z_t is not constrained to have unit variance, the scale parameter α_t can be interpreted as the conditional variance of returns only up to a proportionality factor. Relying on an exponential link function, the dynamic model is therefore specified in terms of the log-transformed scale, λ_t. A distinctive feature of the proposed modeling framework, relative to the standard realized GARCH specification, is the incorporation of a score-driven updating mechanism. The key components of the scale dynamics are the conditional scores of the likelihood functions, through which information from the most recent observations enters directly into updating the latent scale process. In addition, as shown below, the score-driven framework naturally accommodates a leverage channel, with the conditional score of the return innovation density generating asymmetric effects on volatility.

More precisely, three key departures from the general form of the realized GARCH model of Hansen et al. (2012) are as follows. First, the dynamic specification of λ_t adopts a two-component structure, similar to Harvey and Sucarrat (2014). This component structure allows reproducing long-memory behavior, in line with Ding and Granger (1996), who showed that a component GARCH structure can generate patterns observationally equivalent to long memory. It is worth noting that, to guarantee identifiability of the components and avoid the so-called exchangeability problem, it is necessary to impose constraints on their persistence. These constraints allow us to refer to the two components as the long- and short-memory components, respectively.

The volatility level is therefore expressed as a linear combination of the two components, each governed by its own autoregressive (AR) dynamics. This structure lets the model capture both highly persistent movements and short-term, rapid responses to new information. Modeling long- and short-term volatility separately has clear advantages, including a clearer characterization of the leverage effect, as discussed later.

Second, instead of including the realized measure directly in the GARCH equation, the model uses the conditional score of the measurement error density to drive the volatility updates.

Third, the dynamic model for λ_t can be further generalized to include an additional leverage term, which can generate dynamic asymmetric effects on volatility. This can be achieved by adding to the short-memory component specification a term that depends on the sign of past returns and on the score of the conditional return log-likelihood.

The resulting (baseline) specification for λ_t is

$$\lambda_{t|t-1} = \omega_\lambda + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \tag{12}$$

$$\lambda_{1,t+1|t} = \phi_1 \lambda_{1,t|t-1} + \kappa_1 (u_{r,t} + u_{x,t}), \tag{13}$$

$$\lambda_{2,t+1|t} = \phi_2 \lambda_{2,t|t-1} + \kappa_2 (u_{r,t} + u_{x,t}), \tag{14}$$

where $\lambda_{1,t|t-1}$ and $\lambda_{2,t|t-1}$ characterize the long-term and short-term components, respectively; $u_{r,t}$ and $u_{x,t}$ are the scores of the log-likelihood components implied by the returns and measurement equations, respectively:

$$u_{r,t} = \frac{\partial \ell_{r,t-1}}{\partial \lambda_t}, \tag{15}$$

$$u_{x,t} = \frac{\partial \ell_{x,t-1}}{\partial \lambda_t}, \tag{16}$$

where θ is a vector collecting the static parameters of the model; for ease of reference, we refer to $u_{r,t}$ and $u_{x,t}$ as the returns and measurement scores, respectively. It is trivial to show that the overall score $u_{rx,t}$ is given by the sum of $u_{r,t}$ and $u_{x,t}$

$$u_{rx,t} = \frac{\partial \ell_{rx,t-1}}{\partial \lambda_t} = \frac{\partial \ell_{r,t-1}}{\partial \lambda_t} + \frac{\partial \ell_{x,t-1}}{\partial \lambda_t} = u_{r,t} + u_{x,t}.$$

The long-term component, denoted by $\lambda_{1,t|t-1}$, typically has an AR coefficient ϕ_1 close to one, or is even set to unity. In contrast, the short-term component, denoted by $\lambda_{2,t|t-1}$, usually places relatively greater weight on the score innovation κ_2 and is combined with a smaller AR coefficient. The model is not identifiable when $\phi_2 = \phi_1$. Therefore, following Harvey and Sucarrat (2014), identifiability and stationarity are ensured by restricting the AR coefficients to satisfy $0 < |\phi_2| < |\phi_1| < 1$. This constraint also prevents the exchangeability of the long- and short-term components.

The two conditional scores, $u_{r,t}$ and $u_{x,t}$, enter both components with common weight parameters, κ_1 and κ_2, summarizing the information contained in the densities of the return and measurement equations.

2.4 Leverage effect

Although introducing skewness can induce a slight leverage effect, Harvey (2013) finds that this effect is modest in magnitude. The model therefore incorporates two distinct mechanisms through which the leverage effect can operate directly.

First, following Harvey (2013), we introduce a leverage term in the volatility dynamics. Consistent with Engle and Lee (1999) and subsequent findings, we restrict it to operate only through the short-term dynamics. Thus, the short-term component defined in Eq. 14 becomes

$$\lambda_{2,t+1|t} = \phi_2 \lambda_{2,t|t-1} + \kappa_2 (u_{r,t} + u_{x,t}) + \kappa^* \operatorname{sgn}(-r_t)\,(u_{r,t} + 1), \tag{17}$$

where κ^* determines the contribution of the leverage term.

We restrict the leverage function to $u_{r,t}$ to ensure that volatility asymmetry is driven solely by signed return innovations. Since the realized measure is defined on a positive support, its score $u_{x,t}$ reflects magnitude rather than directional shocks. Including $u_{x,t}$ in the leverage term would therefore introduce sign-independent variation and weaken identification of the weight parameter κ^*.

Second, as in Hansen et al. (2012), an additional leverage function enters the measurement equation in a simple quadratic form of $\tau(z_t) = \tau_1 z_t + \tau_2(z_t^2 - 1)$. Consequently, the measurement equation defined in Eq. 7 is modified to

$$x_t = \log \omega_x + \lambda_t + \log \varepsilon_t + \tau(z_t). \tag{18}$$

As shown below, allowing the leverage effect to operate through multiple channels makes the model more flexible and further improves its ability to capture the dynamic relationship between returns and volatility.

2.5 Score-driven dynamics

Within the proposed modeling framework, an updating mechanism driven by the conditional score of the joint likelihood governs volatility dynamics. Specifically, the returns and measurement conditional scores, which extract information from the likelihood components associated with the returns and measurement equations, respectively, enter both the long- and short-term components of volatility dynamics.

In this section, we provide analytical derivations of the returns and measurement conditional scores under the frameworks introduced in Sections 2.1, 2.3, and 2.4. More specifically, we assume a skew t distribution for the returns innovation z_t, a Burr distribution for the measurement error ε_t; the long- and short- term volatility components are specified as in Eq. 13 and 17, respectively, while the measurement equation is defined as in Eq. 23. Analytically, the general RDS-EGARCH takes the form

$$\begin{aligned}
r_t &= \exp(\lambda_t) z_t = \sigma_t z_t, && (19)\\
\lambda_t &= \omega_\lambda + \lambda_{1,t} + \lambda_{2,t}, && (20)\\
\lambda_{1,t+1} &= \phi_1 \lambda_{1,t} + \kappa_1(u_{r,t} + u_{x,t}), && (21)\\
\lambda_{2,t+1} &= \phi_2 \lambda_{2,t} + \kappa_2(u_{r,t} + u_{x,t}) + \kappa^* \operatorname{sgn}(-r_t)\,(u_{r,t} + 1), && (22)\\
x_t &= \log \omega_x + \lambda_t + \log \varepsilon_t + \tau(z_t). && (23)
\end{aligned}$$

Formally, the returns score is computed as

$$u_{r,t} = \frac{(\nu + 1)\,[r_t^2 + r_t \mu_z \exp(\lambda_t)]}{\nu \exp(2\lambda_t)\gamma^{2\operatorname{sgn}(c_t)} + c_t^2} - 1, \tag{24}$$

where the term c_t is given by

$$c_t = r_t + \mu_z \exp(\lambda_t).$$

The measurement score is derived as

$$u_{x,t} = -a \exp(-2\,\lambda_t) \left(2\,\tau_2\, r_t^2 + \tau_1 \exp(\lambda_t)\, r_t - \rho \exp(2\,\lambda_t)\right) b_t, \tag{25}$$

where b_t is given by

$$b_t = \frac{q d_t - 1}{d_t + 1},$$

with the term d_t specified as

$$d_t = \exp\left(-\tau_2 \exp(-2\,\lambda_t)\, r_t^2 - \tau_1 \exp(-\lambda_t)\, r_t - \rho\,\lambda_t - \omega_x + \tau_2 + x_t\right)^a.$$

The analytical expressions for $u_{r,t}$ and $u_{x,t}$ derived in Eq. (24) and Eq. (25), respectively, make clear that the dynamics of the two components, $\lambda_{1,t}$ and $\lambda_{2,t}$, non-linearly depend, through both $u_{r,t}$ and $u_{x,t}$, on the past squared returns as well as on past values of the realized measure, through x_t. Also, it is worth noting that the score components depend on the return r_t, thus contributing to the estimation of the leverage effect through a third channel, in addition to those highlighted in Section 2.4. Therefore, past values of the score contribute to the dynamics of the scale parameter λ_t in a complex way, leveraging both low- and high-frequency information.

The different innovation distributions imply heterogeneous score dynamics and different expected curvature magnitudes for the two log-likelihood components in Eq. (9). This makes the scaling of the forcing variables in the dynamic models for $\lambda_{1,t}$ and $\lambda_{2,t}$, i.e., the scores, of fundamental importance. In the score-driven literature, it is standard practice to scale the score by the inverse of the expected Fisher information (Harvey and Palumbo, 2023). The overall expected Fisher information with respect to λ_t is given by

$$\begin{aligned}\mathbb{E}(\mathcal{I}_{rx,t}) &= -E\left(\frac{\partial^2 \ell_{rx,t-1}}{\partial\lambda_t \partial\lambda_t}\right) \\ &= -E\left(\frac{\partial u_{rx,t}}{\partial\lambda_t \partial\lambda_t}\right) = -E\left(\frac{\partial u_{r,t}}{\partial\lambda_t}\right) - E\left(\frac{\partial u_{x,t}}{\partial\lambda_t}\right) = \bar{\mathcal{I}}_r + \bar{\mathcal{I}}_x. \end{aligned} \tag{26}$$

Since the score has zero mean and the information identity holds,

$$\mathrm{Var}(u_{i,t}) = \bar{\mathcal{I}}_i, \qquad i = r, x, rx.$$

Under the chosen distributional assumptions, following the score specification in Harvey (2013), the expected information associated with the return score $\bar{\mathcal{I}}_r$ can be written as

$$\bar{\mathcal{I}}_r = \frac{2\nu}{\nu+3} s, \tag{27}$$

where s denotes the adjustment induced by skewness and the mean correction, and is given by

$$s = 1 + \frac{\nu+1}{2\nu}\mu_z^2, \tag{28}$$

while the measurement expected information is given by

$$\bar{\mathcal{I}}_x = \frac{a^2 q}{2+q}. \tag{29}$$

The scaled score is then obtained as

$$s_{rx,t} = \bar{\mathcal{I}}_{rx}^{-1/2}(u_{r,t} + u_{x,t}). \tag{30}$$

Compared to the unscaled case, scaling the aggregated score by the overall inverse Fisher information would only affect the scale of the estimated score coefficients κ_1 and κ_2. Because the score components have different variances, the measurement score contribution would be dominated by the returns score, which has higher

variability. An alternative solution is then to separately standardize the score components by their variances, leading to

$$s_{rx,t} = \bar{\mathcal{I}}_r^{-1/2} u_{r,t} + \bar{\mathcal{I}}_x^{-1/2} u_{x,t}. \tag{31}$$

Individual standardization of the score components better separates their contributions while keeping the number of model parameters unchanged. Furthermore, compared to the approach presented in Eq.(30), the model configuration based on the scaling procedure in Eq.(31) has been found to yield a remarkably better fit. These considerations motivate our preference for this model setting.

3 Empirical analysis

3.1 Data description and estimated models

The in-sample analysis uses high-frequency data for the S&P 500 index from January 5, 1998, to July 2, 2025, covering $T = 6916$ trading days. The forecasting exercise additionally considers three individual stocks: Apple Inc. (AAPL) and Bank of America Corporation (BAC), for which data are available from January 5, 1998, to October 25, 2024, and International Business Machines Corporation (IBM), for which the sample extends to October 28, 2024.

The realized volatility measure is constructed from intraday one-minute prices observed within the standard trading window from 09:30 to 16:00 ET. Specifically, we use a volatility estimator based on the realized range (RR), defined as $y_t = \sqrt{RR_t}$, which has been shown to exhibit lower mean-squared error than realized variance, supported by simulation and empirical evidence (Martens and van Dijk, 2007). The estimator aggregates intraday high–low price ranges and applies a scaling factor originally proposed by Parkinson (1980) to obtain a range-based estimate of daily variance.

Formally,

$$RR_t = \frac{1}{4 \ln 2} \sum_{i=1}^{I_t} \left(\ln H_{t,i} - \ln L_{t,i} \right)^2, \tag{32}$$

where

$$H_{t,i} = \max_{1 \le s \le n_{t,i}} \{P_{t,i,s}\}, \qquad L_{t,i} = \min_{1 \le s \le n_{t,i}} \{P_{t,i,s}\}.$$

Here, $t = 1, 2, \ldots, T$ indexes trading days, $i = 1, 2, \ldots, I_t$ indexes intraday intervals within day t, and $s = 1, 2, \ldots, n_{t,i}$ indexes price observations within interval i. Thus, $H_{t,i}$ and $L_{t,i}$ denote the highest and lowest price levels observed within the corresponding intraday interval.

Table 1 reports descriptive statistics for S&P 500 index returns and the associated realized range. The realized range has a mean of 1.16 and a standard deviation of $\bar{\sigma} = 4.26$, indicating substantial variability. Its unconditional distribution is strongly right-skewed ($\bar{\zeta} = 27.36$) and has extremely high kurtosis ($\bar{\nu} = 1052.70$), reflecting extreme right-tail observations concentrated mainly around the 2008 global financial crisis (GFC) and the COVID-19 market turmoil. The Jarque-Bera test strongly rejects the null hypothesis of normality, consistent with the pronounced positive skewness and extreme kurtosis of the realized range.

Returns display a mean close to zero ($\bar{\mu} = 0.014$) and a standard deviation of 1.02. The distribution is approximately symmetric, with skewness close to zero ($\bar{\zeta} = -0.03$), and is strongly leptokurtic ($\bar{\nu} = 10.42$), indicating heavy tails relative to the

normal distribution. The Jarque–Bera (JB) test again rejects the null hypothesis of normality, consistent with the well-documented empirical properties of financial return series.

Table 1: Summary statistics of returns and realized range

Variable	$\bar{\mu}$	$\bar{\sigma}$	$\bar{\zeta}$	$\bar{\nu}$	**JB test**
realized range	1.160	4.264	27.358	1052.702	3.18×10^8 (0.001)
Returns	0.014	1.015	-0.025	10.422	15874.73 (0.001)

Note: p-values in parentheses.

These empirical characteristics therefore motivate flexible volatility models that jointly account for heavy tails, asymmetry, and extreme observations in realized range-based volatility measures.

We evaluate and compare twelve specifications. The first two are beta-skew-t-EGARCH models: the original formulation of Harvey and Sucarrat (2014), used here as a benchmark and denoted btE, and an extension that incorporates the realized range into the short-term volatility dynamics, denoted btE-rr. Specifically, we add a new term to the short-term dynamic volatility component of Eq. (14), including the realized measure x_t weighted by parameter δ. We use these specifications to assess the information content of the unmodeled realized measure and its impact on forecast performance relative to the score-driven volatility update.

The next two models are realized volatility models that explicitly incorporate the realized range through a measurement equation. The realized GARCH model of Hansen et al. (2012), denoted rgarch, jointly models returns and the realized measure by linking the latent conditional variance to the observed realized range. The realized EGARCH model of Hansen and Huang (2016), denoted rEgarch, provides an alternative realized-volatility specification that allows multiple realized measures to enter through a system of measurement equations. In addition, it adopts an exponential specification for latent volatility, allowing a more flexible representation of the dependence between returns and volatility. In the present analysis, we consider the single-measure specification that includes only the realized range. These two models therefore provide established realized-measure benchmarks with distinct volatility dynamics against which we can evaluate the score-driven specifications.

The remaining models correspond to alternative specifications of the RDS-EGARCH framework described in Section 2. The model denoted rds-1c1l is the baseline RDS-EGARCH specification with one component and a single leverage channel. The rds-1c2l and rds-2c1l specifications extend the baseline model by introducing, respectively, a second leverage channel and a second volatility component. We then consider the fully specified RDS-EGARCH model, rds-2c2l, defined in Eqs. (19)–(23), which combines a two-component structure with two leverage channels. The remaining four specifications augment each RDS-EGARCH variant with the realized range using the same realized-measure extension introduced for the alternative beta-skew-t-EGARCH specification. This nested modeling scheme, progressing from the most parsimonious RDS-EGARCH specification to increasingly general variants, facilitates assessing the marginal contribution of each modeling component. Table 2 summarizes the models and their corresponding short-

hands, providing a compact overview of the specification hierarchy.

Table 2: Overview of model specifications

Shorthand	**Model Specification**
btE	beta-skew-t-EGARCH (Harvey and Sucarrat, 2014)
btE-rr	beta-skew-t-EGARCH with RR in the volatility process
rgarch	realized GARCH (Hansen et al., 2012)
rEgarch	realized EGARCH (Hansen and Huang, 2016)
rds-1c1l	RDS-EGARCH with one component and one leverage channel
rds-1c2l	RDS-EGARCH with one component and two leverage channels
rds-2c1l	RDS-EGARCH with two components and one leverage channel
rds-2c2l	RDS-EGARCH with two components and two leverage channels
rds-1c1lrr	RDS-EGARCH with one component, one leverage channel, and RR
rds-1c2lrr	RDS-EGARCH with one component, two leverage channels, and RR
rds-2c1lrr	RDS-EGARCH with two components, one leverage channel, and RR
rds-2c2lrr	RDS-EGARCH with two components, two leverage channels, and RR

3.2 In-sample analysis

For the in-sample analysis, we consider eight specifications estimated over the full sample of demeaned S&P 500 returns. Table 3 reports the corresponding parameter estimates, model-fit measures, and diagnostic statistics.

In all specifications, where applicable, inertia in the long-term component is strong, with ϕ_1 close to unity. Such large AR coefficients are consistent with the long-memory characteristics typically observed in volatility dynamics. Together with their statistical significance, these estimates suggest that the model appropriately identifies the long-term component under the two-component specifications.

The estimated score weights further indicate that the long-term component responds more gradually to new information, with κ_1 generally smaller than the corresponding short-term score weight κ_2.

The lower values of δ in the augmented beta-t-EGARCH and RDS-EGARCH models suggest that, once information is mediated through the score-driven mechanism, direct inclusion of the realized range becomes less important for the volatility update.

Turning to the shape parameters, the estimated skewness parameter γ ranges from 0.84 to 0.96, indicating a higher risk of large negative returns compared to large positive returns. Similarly, the exponential Burr shape parameters in the RDS-EGARCH model indicate pronounced asymmetry in the conditional distribution or the realized range measure.

Lastly, leverage is pronounced in all the models considered. The estimated parameter value κ^* associated with the score-driven leverage term is consistent with the standard interpretation that volatility increases after substantial negative returns. Similarly, the shape of the additional leverage function $\tau(z_t)$ in the RDS-EGARCH models follows the expected patterns, as indicated by the signs and significance of the associated coefficients. This suggests that incorporating the leverage function in both the measurement equation and the volatility dynamics is reasonable.

Table 3: Parameter estimates, model fit, and diagnostic measures

Param.	**beta-t**	**beta-t-rr**	**rgarch**	**regarch**	**rds-1c1l**	**rds-2c1l**	**rds-2c2l**	**rds-2c2lrr**
ω_λ	-0.319	-0.808	0.044	-0.017	-0.009	-0.192	-0.496	-0.307
	[0.131]	[0.124]	[0.012]	[0.003]	[0.002]	[0.062]	[0.084]	[0.049]
ω_x			-0.132	-0.115	0.912	0.906	0.950	0.936
			[0.023]	[0.023]	[0.016]	[0.015]	[0.016]	[0.016]
ρ			1.092	1.122	1.128	1.116	1.097	1.047
			[0.035]	[0.037]	[0.021]	[0.020]	[0.020]	[0.019]
Long term components								
ϕ_1	0.997	0.998				0.989	0.993	0.995
	[0.002]	[0.002]				[0.003]	[0.002]	[0.001]
κ_1	0.024	0.011				0.035	0.025	0.009
	[0.005]	[0.002]				[0.006]	[0.004]	[0.002]
Short term components								
ϕ_2	0.937	0.840	0.471	0.967	0.965	0.773	0.889	0.709
	[0.012]	[0.025]	[0.022]	[0.003]	[0.004]	[0.056]	[0.016]	[0.027]
κ_2	0.031	0.024			0.070	0.039	0.038	0.033
	[0.007]	[0.006]			[0.003]	[0.005]	[0.004]	[0.003]
δ		0.089	0.454	0.340				0.170
		[0.016]	[0.023]	[0.019]				[0.020]
Shape								
ν	7.261	7.734			10.995	12.074	9.686	9.342
	[0.603]	[0.678]			[1.051]	[1.280]	[0.927]	[0.889]
γ	0.841	0.836			0.949	0.957	0.844	0.854
	[0.013]	[0.013]			[0.013]	[0.014]	[0.014]	[0.014]
a					10.724	10.885	11.554	12.227
					[0.255]	[0.264]	[0.264]	[0.290]
q					0.547	0.543	0.526	0.482
					[0.024]	[0.024]	[0.023]	[0.020]
Leverage								
κ^*	0.057	0.062					0.046	0.039
	[0.004]	[0.005]					[0.002]	[0.002]
τ_1			-0.100	-0.108	-0.027	-0.029	-0.036	-0.039
			[0.007]	[0.006]	[0.003]	[0.003]	[0.003]	[0.003]
τ_2			0.084	0.052	0.040	0.040	0.029	0.027
			[0.007]	[0.005]	[0.002]	[0.002]	[0.002]	[0.002]
Model fit								
LogL:r_t	-8110.2	-8047.4	-8175.1	-8161.0	-8105.2	-8110.2	-8029.2	-8006.4
LogL:x_t			-4446.1	-4248.7	785.2	869.2	1181.3	1237.4
TotLogL	-8110.2	-8047.4	-12621.1	-12409.7	-7319.9	-7241.1	-6848.0	-6769.0
AIC	16236.4	16112.9	25258.2	24839.5	14661.9	14508.1	13723.9	13567.9
BIC	16291.2	16174.5	25313.0	24907.9	14737.1	14597.0	13819.7	13670.5
K-S:r_t	0.012	0.013	0.034	0.035	0.023	0.025	0.012	0.008
	(0.290)	(0.175)	(0.000)	(0.000)	(0.001)	(0.000)	(0.284)	(0.818)
K-S:x_t			0.059	0.057	0.010	0.010	0.016	0.013
			(0.000)	(0.000)	(0.548)	(0.436)	(0.070)	(0.224)

Note: Standard errors of parameter estimates are reported in square brackets; LogL denotes the log-likelihood. Q-st denotes the Ljung–Box statistic. K–S denotes the Kolmogorov–Smirnov statistic. Corresponding p-values are reported in parentheses.

Regarding model fit, the Kolmogorov–Smirnov (K–S) test provides mixed evi-

dence across specifications. For the return distribution, we do not reject the null at the 5% significance level for the beta-t-EGARCH and fully specified RDS-EGARCH models. For the realized range distribution, we do not reject the null for the RDS-EGARCH specifications, but we reject it for the applicable benchmark models. These results indicate a noticeable improvement in distributional fit for the proposed RDS-EGARCH models, particularly in their fully specified variants..

Based on the partial log-likelihood for returns, the fully specified RDS-EGARCH models provide the best fit among the specifications considered. Allowing the realized range to enter the volatility dynamics directly further improves the return fit, with the fully specified RDS-EGARCH model incorporating the realized range yielding the highest partial log-likelihood. Together with the statistical significance of the score-weight parameters, these results indicate that the realized measure contains relevant information for return dynamics and that incorporating it through the score-driven mechanism materially improves model fit. The additional improvement from directly including the realized range suggests it provides complementary information beyond what the score already captures.

To further investigate model performance, the next section presents results from an application to tail risk forecasting, which provides the main empirical motivation for the proposed modeling framework.

3.3 Forecast evaluation

We use the models described in Table 2 to forecast ES and VaR for daily returns of the S&P 500 index and the three individual assets considered in the analysis at four risk levels, $\alpha \in 0.10, 0.05, 0.025, 0.01$, allowing us to assess performance at both the market and individual-asset levels.

The out-of-sample interval, beginning on July 1, 2007, covers several major episodes of market stress. The evaluation includes the 2007–2009 GFC and the subsequent European sovereign debt crisis, which generated substantial spillovers to global and US financial markets. It also covers the COVID-19 shocks in February 2020, which produced exceptional uncertainty and extreme market turbulence. The period also includes the onset of the Russia–Ukraine war in early 2022, a sudden geopolitical shock that drove sharp moves in equity and energy markets. The period also includes the global inflation surge from the second half of 2021 through 2022, followed by subsequent monetary tightening and elevated macroeconomic uncertainty. The evaluation window also includes the US regional banking stress in March 2023, which triggered a brief but intense spike in risk premia. Finally, the evaluation period extends through the election-related uncertainty and renewed geopolitical tensions observed from mid-2024 to October 2024. Including these episodes in the forecasting horizon allows a robust comparison of model performance during volatile market conditions, where tail-risk forecasting is most relevant.

We begin the forecast evaluation by backtesting the one-period-ahead VaR forecasts over the out-of-sample period. Table 4 summarizes the results by reporting rejection frequencies aggregated across all four assets and all four α values. The Appendix reports the corresponding asset-specific backtesting results.

The RDS-EGARCH specifications exhibit substantially lower rejection frequencies than the realized GARCH and realized EGARCH benchmarks, particularly for the POF test of Kupiec, the CC and CCI tests of Christoffersen (1998), and the DQ test of Engle and Manganelli (2004). Among the proposed models, RDS-1c2l-

Table 4: Overall VaR backtest rejection frequencies

Model	TUFF (%)	POF (%)	CC (%)	CCI (%)	DQ (%)
beta-t	**19**	38	44	19	50
beta-t-rr	**19**	19	12	6	31
RGARCH	12	56	81	38	81
REGARCH	12	56	81	50	81
RDS-1c1l	**19**	19	38	19	44
RDS-1c2l	12	**12**	38	31	31
RDS-2c1l	**19**	25	31	**12**	56
RDS-2c2l	**19**	**12**	**25**	31	31
RDS-1c1l-rr	12	19	31	19	31
RDS-1c2l-rr	12	6	31	38	**38**
RDS-2c1l-rr	**19**	25	38	19	**38**
RDS-2c2l-rr	**19**	19	44	38	44

Note: Entries report rejection frequencies at the 5 percent significance level aggregated across all four assets and all four VaR tail probabilities. Each entry is based on 16 asset–risk-level combinations. TUFF denotes the time-until-first-failure test, POF the proportion-of-failures test, CC and CCI the conditional coverage and conditional coverage independence tests, respectively, and DQ the dynamic quantile test. Lower rejection frequencies indicate more consistent VaR performance across assets and risk levels. The lowest rejection frequency for each test is boxed, while the second-lowest distinct rejection frequency is shown in bold.

rr has the lowest POF rejection frequency, while RDS-2c2l performs particularly well in conditional coverage and dynamic quantile adequacy. Rejection frequencies for the TUFF test of Christoffersen and Pelletier (2004) are comparatively similar across specifications.

The beta-t-rr specification also performs strongly, attaining the lowest rejection frequencies for the CC and CCI tests and tying for the lowest DQ rejection frequency. Thus, while the proposed RDS-EGARCH specifications generally improve upon the realized-volatility GARCH benchmarks and remain competitive across the considered backtests, the results do not indicate uniform dominance over all benchmark specifications. These rejection frequencies nevertheless provide strong evidence of the robustness of the proposed models across assets and VaR tail probabilities.

To complement the diagnostic assessment from the VaR backtests and gain additional insight into forecasting performance, we report in Table 5 the average losses computed over the out-of-sample period. We use two loss functions to obtain these values. First, we employ the standard quantile check loss, which depends solely on the VaR forecasts. Second, since ES forecasts are also available, we compute joint VaR–ES losses using the FZ0 loss function defined in Patton et al. (2019), which is strictly consistent for the pair (VaR, ES). The observed losses consistently indicate that the RDS-EGARCH model, in which realized measures directly enter into the volatility dynamics, is preferred across all risk levels. The advantage over its counterpart is marginal under the standard quantile loss, but becomes more pronounced when the ES enters the loss function.

Table 5: Overall out-of-sample average losses

Model	Average quantile loss	Average FZ0 loss
beta-t	0.13362	1.07883
beta-t-rr	0.13770	1.04989
RGARCH	0.13207	1.09135
REGARCH	0.13147	1.08258
RDS-1c1l	0.13074	1.03623
RDS-1c2l	0.13000	1.02916
RDS-2c1l	0.13104	1.04295
RDS-2c2l	0.13066	1.03558
RDS-1c1l-rr	0.12870	1.01540
RDS-1c2l-rr	**0.12874**	**1.01637**
RDS-2c1l-rr	0.12919	1.02094
RDS-2c2l-rr	0.12941	1.02297

Note: Entries report average out-of-sample losses aggregated across all four assets and all four VaR tail probabilities. Each entry is based on 16 asset–risk-level combinations. Lower values indicate better forecast performance. The lowest average loss for each loss function is boxed, while the second-lowest distinct value is shown in bold.

Average losses provide a preliminary view of out-of-sample forecast performance but do not indicate whether the observed differences are statistically meaningful. To formally assess whether these differences reflect significant variation in predictive ability, we apply the model confidence set (MCS) procedure of Hansen et al. (2011). The MCS allows for the simultaneous comparison of multiple forecasting models while accounting for multiple testing. Following Hansen et al. (2011), we consider two confidence levels, 75% and 90%, implying that models with p-values below 0.25 and 0.10, respectively, are excluded from the preferred set. Table 6 reports the corresponding p-values and identifies the specifications that enter the set of superior predictive models at the 75% and 90% confidence levels, denoted by $\hat{\mathcal{M}}^*_{75}$ and $\hat{\mathcal{M}}^*_{90}$, respectively.

For quantile loss, the results show a clear advantage for RDS specifications augmented with the realized range. The RDS-1c1l-rr model is included in the superior set in all cases, achieving an overall inclusion frequency of 100% at both confidence levels. RDS-1c2lrr and RDS-2c1lrr also perform strongly, with aggregate inclusion frequencies of 87.5% and 75% at the 90% and 75% confidence levels, respectively. In contrast, the beta-t-EGARCH specifications are considerably less robust, with btE attaining overall inclusion rates of 25% and 6.25%, while btE-rr reaches 18.75% and 0%, respectively. These findings indicate that incorporating the realized range within the RDS framework substantially improves VaR forecast performance across the four assets.

The evidence is even stronger for joint VaR–ES losses based on the FZ0 function. RDS-1c1lrr again achieves full inclusion across all assets and probability levels at both confidence levels. RDS-2c1lrr also performs particularly strongly, with overall rates of inclusion of 100% and 87.5%, while RDS-1c2lrr reaches 87.5% and 81.25%. By comparison, btE is almost entirely excluded from the superior set, whereas btE-rr performs better but remains substantially less consistent than the leading RDS specifications. The inclusion frequencies therefore indicate that the

advantage of the realized-range RDS models becomes particularly pronounced once ES information is incorporated.

The beta-t-EGARCH model augmented with the realized range performs better than the baseline beta-t specification in some of the MCS results, particularly under the FZ0 loss. However, its overall rates of inclusion remain below those of the strongest RDS specifications.

Standard quantile-based losses depend on the frequency and magnitude of violations but are relatively insensitive to conservatism when no violation occurs. As a result, models that systematically produce more conservative VaR forecasts may avoid violation penalties while not being sufficiently penalized for excessive capital allocation. To account for this limitation, we also consider the alternative loss functions proposed by Caporin (2008), which evaluate the distance between realized returns and VaR forecasts and therefore incorporate the opportunity costs associated with overly conservative risk estimates:

$$f_1(\varepsilon_{t+1}, VaR_{m,t+1}) = \left| 1 - \left| \frac{\varepsilon_{t+1}}{VaR_{m,t+1}} \right| \right| \tag{33}$$

$$f_2(\varepsilon_{t+1}, VaR_{m,t+1}) = \frac{\left(|\varepsilon_{t+1}| - |VaR_{m,t+1}|\right)^2}{|VaR_{m,t+1}|} \tag{34}$$

$$f_3(\varepsilon_{t+1}, VaR_{m,t+1}) = |\varepsilon_{t+1} - VaR_{m,t+1}| \tag{35}$$

These alternative loss functions complement VaR forecast assessment by penalizing deviations between realized returns and forecasted risk thresholds, even without violations. They therefore help distinguish accurate VaR forecasts from forecasts that achieve adequate coverage primarily by being excessively conservative.

This aspect aligns with the argument in Caporin (2008), which emphasizes that loss functions focused on violations may fail to capture the opportunity costs of overly conservative risk measures. To better understand the source of the similar forecasting performance, we therefore employ the alternative loss specifications in (Eqs. (33)–(35)), which provide an additional perspective on model performance by accounting for these opportunity costs and reducing the tendency to favor excessively conservative VaR forecasts.

Under the alternative loss functions, the ranking is less uniform than for the quantile and FZ0 losses. RDS-2c1l-rr attains the highest overall inclusion frequency, reaching 50% at both confidence levels, while REGARCH is the second-best specification with inclusion frequencies of 45.83% and 41.67%. RDS-2c1l also performs comparatively well, particularly for BAC, whereas the remaining RDS and beta-t specifications are included less frequently. These results therefore provide more mixed evidence, although they continue to show strong performance for the RDS-2c1lrr specification once the opportunity costs associated with conservative VaR forecasts are taken into account.

In summary, the MCS results are broadly consistent with the earlier forecast evaluation and average loss rankings. The evidence is particularly strong for the realized-range RDS specifications under the quantile and FZ0 loss functions, with RDS-1c1lrr exhibiting the most stable performance. Under the alternative losses, the evidence is more heterogeneous, but RDS-2c1lrr achieves the highest overall inclusion rate. The MCS results therefore provide additional support for the RDS-EGARCH framework, particularly when the realized range is incorporated directly into the volatility dynamics..

Table 6: Model confidence set: SPY, AAPL, IBM and BAC

VaR

	SPY		AAPL		IBM		BAC		Overall (%)	
Model	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}
btE	0	0	0	0	0	0	4	1	25.00	6.25
btE-rr	0	0	0	0	0	0	**3**	0	18.75	0.00
rG	0	0	1	0	0	0	4	0	31.25	0.00
rEG	2	0	1	0	0	0	4	2	43.75	12.50
rds-1c1l	2	1	1	**1**	**1**	0	4	**3**	50.00	31.25
rds-1c2l	**3**	2	1	0	0	0	4	**3**	50.00	31.25
rds-2c1l	0	0	1	**1**	0	0	4	2	31.25	18.75
rds-2c2l	2	1	1	0	0	0	4	1	43.75	12.50
rds-1c1lrr	4	4	4	4	4	4	4	4	100.00	100.00
rds-1c2lrr	4	**3**	**2**	**1**	4	4	4	4	**87.50**	**75.00**
rds-2c1lrr	4	**3**	**2**	**1**	4	4	4	4	**87.50**	**75.00**
rds-2c2lrr	**3**	1	1	**1**	4	4	4	4	75.00	62.50

FZ0

	SPY		AAPL		IBM		BAC		Overall (%)	
Model	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}
btE	0	0	0	0	**1**	0	0	0	6.25	0.00
btE-rr	0	0	0	0	4	**2**	**3**	**1**	43.75	18.75
rG	0	0	1	0	0	0	0	0	6.25	0.00
rEG	0	0	1	0	0	0	0	0	6.25	0.00
rds-1c1l	1	0	**2**	1	0	0	**3**	0	37.50	6.25
rds-1c2l	**3**	**2**	**2**	1	0	0	**3**	**1**	50.00	25.00
rds-2c1l	0	0	**2**	1	0	0	2	0	25.00	6.25
rds-2c2l	1	0	**2**	1	0	0	1	0	25.00	6.25
rds-1c1lrr	4	4	4	4	4	4	4	4	100.00	100.00
rds-1c2lrr	4	4	**2**	1	4	4	4	4	**87.50**	81.25
rds-2c1lrr	4	4	4	**2**	4	4	4	4	100.00	**87.50**
rds-2c2lrr	1	0	**2**	1	4	4	4	4	68.75	56.25

Caporin (2008)

	SPY		AAPL		IBM		BAC		Overall (%)	
Model	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}	M_{90}	M_{75}
btE	0	0	0	0	1	0	1	1	4.17	2.08
btE-rr	0	0	0	0	0	0	1	1	2.08	2.08
rG	**4**	3	2	0	1	0	0	0	14.58	6.25
rEG	8	7	**4**	**3**	**6**	**6**	**4**	**4**	**45.83**	**41.67**
rds-1c1l	0	0	0	0	2	0	0	0	4.17	0.00
rds-1c2l	0	0	0	0	0	0	0	0	0.00	0.00
rds-2c1l	**4**	**4**	0	0	2	0	9	8	31.25	25.00
rds-2c2l	0	0	0	0	0	0	0	0	0.00	0.00
rds-1c1lrr	3	2	1	1	5	2	2	1	22.92	12.50
rds-1c2lrr	2	2	1	1	2	1	1	1	12.50	10.42
rds-2c1lrr	**4**	**4**	9	9	8	8	3	3	50.00	50.00
rds-2c2lrr	2	2	1	1	3	2	2	2	16.67	14.58

Note: The table reports how often each model is included in the model confidence set across the four probability levels $\alpha \in \{0.10, 0.05, 0.025, 0.01\}$ for each asset. M_{90} and M_{75} denote inclusion at the 90% and 75% confidence levels, respectively. The "Overall" columns report the corresponding overall rates of inclusion across all assets and risk levels. For the alternative loss functions of Caporin (2008), the overall frequencies additionally aggregate across the three versions of the loss function in Eqs. (33)–(35). The highest inclusion frequency in each column is boxed, while the second-highest distinct inclusion frequency is shown in bold.

4 Conclusions

Traditional GARCH models have several well-known limitations, particularly in accommodating heavy-tailed and skewed return distributions, leverage effects, and persistent volatility dynamics. These limitations reduce the flexibility of conventional models to capture the complex behavior of financial volatility and, consequently, affect the accuracy of volatility and tail-risk forecasts.

To address these challenges, we introduce the realized dynamic score EGARCH (RDS-EGARCH) model. The proposed framework jointly models asset returns and a range-based realized volatility measure within a score-driven setting. It accommodates skewness and heavy tails in the conditional return distribution and incorporates a two-component volatility structure together with extended leverage channels. By exploiting the information contained in the conditional densities through their scores, while also allowing realized volatility to contribute to the volatility dynamics, the framework provides a flexible and adaptive representation of financial market volatility.

The empirical results strongly support the proposed framework. The in-sample analysis for the S&P 500 indicates that the RDS-EGARCH specifications deliver improved distributional fit and return likelihood relative to the benchmark models considered. The out-of-sample analysis, based on VaR and ES forecasts for the S&P 500 and three individual assets, provides further evidence of superior forecasting performance. The RDS-EGARCH specifications consistently outperform the realized GARCH and realized EGARCH benchmarks across VaR backtests and achieve the lowest average quantile and FZ0 losses, with the strongest results obtained when the realized range enters directly into the volatility dynamics. The model confidence set analysis provides particularly strong support for these specifications under both quantile and FZ0 losses, with the leading RDS-EGARCH models exhibiting very high and, in some cases, complete inclusion frequencies across assets and risk levels. Although the alternative loss functions produce a more heterogeneous ranking, an RDS-EGARCH specification still attains the highest overall inclusion frequency. In summary, the empirical evidence indicates a clear forecasting advantage for the proposed framework, particularly when realized range information enters the volatility dynamics directly.

A potential direction for future research is to further develop the proposed framework by allowing multiple realized volatility measures to enter the model jointly. In the spirit of the realized EGARCH framework of Hansen and Huang (2016), incorporating several realized measures could allow the RDS-EGARCH model to exploit complementary information from alternative high-frequency volatility proxies and provide a richer characterization of latent volatility dynamics.

Overall, the results indicate that combining score-driven dynamics with realized range information provides an effective framework for modeling volatility and forecasting tail risk. In particular, the RDS-EGARCH specifications show strong and consistent forecasting performance across assets, tail probabilities, and alternative evaluation criteria, with the best results when realized range enters the volatility dynamics directly. These findings support using realized measures in flexible, score-driven volatility models and highlight the empirical advantages of the proposed framework.

Appendix

Table A.1: VaR backtesting across risk levels (SPY)

	$\alpha = 0.1$					$\alpha = 0.05$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-*t*	0.505	0.965	**0.040**	**0.011**	0.210	0.237	**0.042**	**0.008**	**0.020**	**0.050**
beta-*t*-rr	0.505	0.839	0.180	*0.066*	0.148	0.237	0.102	*0.086*	0.136	0.137
RGARCH	0.505	0.379	0.384	0.285	0.136	0.237	**0.001**	**0.004**	0.428	**0.002**
REGARCH	0.505	0.494	0.531	0.371	0.803	0.237	**0.004**	**0.010**	0.323	**0.028**
RDS-1c1l	0.505	0.558	0.217	*0.100*	*0.090*	0.237	**0.003**	**0.009**	0.375	**0.025**
RDS-1c2l	0.505	0.965	**0.040**	**0.011**	*0.091*	0.237	0.241	0.315	0.334	0.779
RDS-2c1l	0.505	0.353	0.202	0.126	**0.022**	0.237	**0.001**	**0.001**	0.107	**0.003**
RDS-2c2l	0.505	0.626	*0.066*	**0.023**	0.159	0.237	*0.058*	*0.071*	0.196	0.392
RDS-1c1l-rr	0.505	0.329	0.183	0.118	0.249	0.237	**0.042**	*0.096*	0.453	0.171
RDS-1c2l-rr	0.505	0.558	*0.054*	**0.019**	0.199	0.237	*0.077*	*0.098*	0.217	0.299
RDS-2c1l-rr	0.505	0.172	0.164	0.186	0.289	0.237	**0.016**	**0.043**	0.520	0.250
RDS-2c2l-rr	0.505	0.261	**0.023**	**0.012**	0.119	0.237	**0.009**	**0.009**	0.106	*0.071*
	$\alpha = 0.025$					$\alpha = 0.01$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-*t*	0.108	**0.001**	**0.002**	0.129	**0.010**	**0.038**	**0.028**	*0.088*	0.843	**0.004**
beta-*t*-rr	0.108	**0.003**	**0.004**	0.143	**0.003**	**0.038**	**0.028**	*0.088*	0.843	*0.062*
RGARCH	0.108	**0.000**	**0.000**	*0.090*	**0.000**	**0.038**	**0.000**	**0.000**	0.153	**0.000**
REGARCH	0.108	**0.000**	**0.000**	0.190	**0.000**	**0.038**	**0.000**	**0.000**	0.125	**0.000**
RDS-1c1l	0.108	**0.000**	**0.000**	0.438	**0.001**	**0.038**	**0.000**	**0.000**	0.735	**0.000**
RDS-1c2l	0.108	**0.016**	**0.025**	0.202	0.223	0.132	0.131	0.160	0.237	**0.026**
RDS-2c1l	0.108	**0.000**	**0.000**	0.328	**0.000**	**0.038**	**0.000**	**0.000**	0.671	**0.000**
RDS-2c2l	0.108	**0.003**	**0.007**	0.375	*0.052*	**0.038**	**0.009**	**0.034**	0.917	**0.007**
RDS-1c1l-rr	0.108	**0.001**	**0.003**	0.331	**0.017**	0.132	**0.009**	**0.034**	0.917	**0.004**
RDS-1c2l-rr	0.108	**0.006**	**0.003**	**0.039**	**0.031**	0.132	*0.074*	*0.096*	0.221	**0.017**
RDS-2c1l-rr	0.108	**0.001**	**0.001**	0.116	**0.003**	**0.038**	**0.003**	**0.011**	0.989	**0.001**
RDS-2c2l-rr	0.108	**0.003**	**0.005**	0.150	**0.016**	**0.038**	**0.003**	**0.011**	0.989	**0.000**

Note: TUFF denotes the time-until-first-failure test, POF the proportion-of-failures test, CC and CCI the conditional coverage and conditional coverage independence tests, respectively, and DQ the dynamic quantile test. Values for which the null hypothesis is rejected at the 5 percent significance level are shown in bold, while values rejected at the 10 percent but not at the 5 percent significance level are shown in italics.

Table A.2: VaR backtesting across risk levels (AAPL)

	$\alpha = 0.1$					$\alpha = 0.05$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-*t*	**0.032**	0.968	0.497	0.237	0.815	0.352	0.787	0.722	0.447	0.658
beta-*t*-rr	**0.032**	0.642	0.527	0.302	0.196	0.352	0.586	0.250	0.116	0.317
RGARCH	0.721	**0.004**	**0.007**	0.216	**0.045**	0.352	0.572	0.821	0.785	0.378
REGARCH	0.721	**0.001**	**0.003**	0.354	**0.022**	0.352	0.323	0.608	0.891	0.879
RDS-1c1l	**0.032**	0.282	0.278	0.236	0.119	0.352	0.884	0.648	0.358	0.429
RDS-1c2l	**0.032**	0.152	**0.016**	**0.012**	**0.029**	0.352	0.723	0.138	*0.050*	0.125
RDS-2c1l	**0.032**	0.716	0.324	0.145	*0.091*	0.352	0.634	0.621	0.394	0.224
RDS-2c2l	**0.032**	0.504	**0.005**	**0.002**	**0.015**	0.352	0.950	0.392	0.172	0.130
RDS-1c1l-rr	**0.032**	0.642	0.136	*0.052*	0.116	0.352	0.684	0.469	0.246	0.217
RDS-1c2l-rr	**0.032**	0.384	**0.034**	**0.014**	**0.036**	0.352	0.895	0.374	0.163	0.271
RDS-2c1l-rr	**0.032**	0.888	0.275	0.110	0.163	0.352	0.413	0.432	0.315	0.109
RDS-2c2l-rr	**0.032**	0.888	**0.043**	**0.012**	*0.072*	0.352	0.340	*0.068*	**0.035**	**0.044**
	$\alpha = 0.025$					$\alpha = 0.01$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-*t*	0.164	0.889	0.441	0.203	0.694	*0.058*	0.743	0.582	0.323	0.333
beta-*t*-rr	0.164	0.522	0.467	0.291	0.874	*0.058*	0.556	0.582	0.391	0.754
RGARCH	0.164	0.419	0.650	0.648	0.763	0.445	**0.014**	**0.048**	0.872	**0.035**
REGARCH	0.164	0.601	0.711	0.522	0.859	0.445	**0.006**	**0.014**	0.313	**0.029**
RDS-1c1l	0.164	0.348	0.632	0.845	0.703	0.445	0.556	0.582	0.391	0.865
RDS-1c2l	0.164	**0.040**	*0.097*	0.504	0.352	0.445	0.287	0.414	0.427	0.782
RDS-2c1l	0.164	0.813	0.835	0.581	0.740	0.445	0.743	0.582	0.323	0.337
RDS-2c2l	0.164	0.522	0.467	0.291	0.770	0.445	0.633	0.536	0.313	0.319
RDS-1c1l-rr	0.164	0.348	0.401	0.330	0.884	0.445	0.366	0.477	0.415	0.826
RDS-1c2l-rr	0.164	0.255	0.343	0.358	0.867	0.445	0.221	0.350	0.440	0.801
RDS-2c1l-rr	0.164	0.669	0.366	0.176	0.756	0.445	0.901	0.648	0.356	0.823
RDS-2c2l-rr	0.164	0.958	0.475	0.223	0.638	0.445	0.858	0.617	0.334	0.906

Note: See the notes to Table A.1 for definitions of the VaR backtests and the interpretation of the reported formatting.

Table A.3: VaR backtesting across risk levels (IBM)

	$\alpha = 0.1$					$\alpha = 0.05$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-t	0.390	0.165	0.348	0.666	0.485	0.180	0.720	0.928	0.883	0.673
beta-t-rr	0.390	0.199	0.413	0.732	0.428	0.180	0.953	0.799	0.504	0.550
RGARCH	0.390	**0.010**	**0.008**	*0.081*	**0.003**	0.180	0.396	**0.007**	**0.002**	**0.000**
REGARCH	0.390	**0.013**	**0.003**	**0.018**	**0.004**	0.180	0.322	**0.002**	**0.001**	**0.001**
RDS-1c1l	0.390	0.100	0.191	0.438	0.132	0.180	0.357	0.255	0.170	*0.100*
RDS-1c2l	0.390	*0.051*	0.122	0.531	*0.099*	0.180	*0.088*	0.125	0.262	0.146
RDS-2c1l	0.390	0.258	0.525	0.929	0.145	0.180	0.479	0.487	0.333	**0.042**
RDS-2c2l	0.390	*0.090*	0.229	0.796	0.269	0.180	0.178	0.331	0.528	0.153
RDS-1c1l-rr	0.390	0.328	0.586	0.737	0.516	0.180	0.936	0.300	0.121	*0.070*
RDS-1c2l-rr	0.390	0.165	0.348	0.666	0.553	0.180	0.479	0.347	0.204	0.145
RDS-2c1l-rr	0.390	0.534	0.748	0.659	0.539	0.180	0.936	0.187	*0.067*	0.105
RDS-2c2l-rr	0.390	0.639	0.845	0.734	0.518	0.180	0.773	0.542	0.285	0.393
	$\alpha = 0.025$					$\alpha = 0.01$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-t	*0.082*	0.967	0.772	0.472	0.199	**0.029**	0.136	0.306	0.699	**0.023**
beta-t-rr	*0.082*	0.803	0.693	0.413	0.599	**0.029**	0.634	0.536	0.313	0.887
RGARCH	*0.082*	0.368	**0.046**	**0.020**	**0.006**	**0.029**	**0.000**	**0.001**	0.123	**0.000**
REGARCH	*0.082*	0.203	**0.005**	**0.003**	**0.001**	**0.029**	**0.000**	**0.000**	**0.034**	**0.000**
RDS-1c1l	*0.082*	0.458	0.467	0.323	0.544	**0.029**	0.360	0.574	0.603	0.897
RDS-1c2l	*0.082*	*0.080*	0.217	0.960	0.703	**0.029**	0.860	0.774	0.488	0.876
RDS-2c1l	*0.082*	0.803	0.389	0.176	0.292	**0.029**	0.228	0.437	0.650	0.433
RDS-2c2l	*0.082*	0.347	0.368	0.291	0.441	**0.029**	0.860	0.617	0.334	0.913
RDS-1c1l-rr	*0.082*	0.815	0.241	*0.095*	0.494	**0.029**	0.780	0.697	0.423	0.896
RDS-1c2l-rr	*0.082*	0.656	0.865	0.761	0.913	**0.029**	0.555	0.581	0.391	0.931
RDS-2c1l-rr	*0.082*	0.956	0.439	0.200	0.217	**0.029**	0.634	0.735	0.533	0.891
RDS-2c2l-rr	*0.082*	0.728	0.655	0.394	0.381	**0.029**	0.744	0.582	0.323	0.901

Note: See the notes to Table A.1 for definitions of the VaR backtests and the interpretation of the reported formatting.

Table A.4: VaR backtesting across risk levels (BAC)

	$\alpha = 0.1$					$\alpha = 0.05$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-*t*	0.505	**0.001**	**0.003**	0.368	**0.001**	0.237	**0.036**	**0.022**	*0.072*	**0.003**
beta-*t*-rr	0.505	**0.023**	*0.065*	0.579	**0.004**	0.237	0.153	0.198	0.272	**0.005**
RGARCH	0.505	0.306	**0.004**	**0.001**	**0.003**	0.729	*0.058*	**0.000**	**0.000**	**0.000**
REGARCH	0.505	0.642	**0.018**	**0.005**	**0.030**	0.729	0.153	**0.000**	**0.000**	**0.000**
RDS-1c1l	0.505	0.588	0.234	0.106	**0.001**	0.729	0.994	**0.032**	**0.009**	**0.010**
RDS-1c2l	0.505	0.793	0.462	0.225	*0.074*	0.729	0.438	**0.034**	**0.013**	**0.030**
RDS-2c1l	0.505	0.255	0.400	0.466	**0.008**	0.729	0.174	*0.062*	*0.054*	**0.006**
RDS-2c2l	0.505	0.554	0.535	0.342	**0.008**	0.729	0.586	*0.099*	**0.038**	*0.053*
RDS-1c1l-rr	0.505	0.400	0.322	0.211	*0.055*	0.729	0.540	**0.007**	**0.002**	**0.019**
RDS-1c2l-rr	0.505	0.623	0.443	0.239	**0.037**	0.729	0.841	**0.041**	**0.012**	*0.078*
RDS-2c1l-rr	0.505	0.181	0.196	0.226	**0.002**	0.729	0.307	**0.037**	**0.018**	**0.023**
RDS-2c2l-rr	0.505	0.151	0.219	0.323	**0.004**	0.729	0.307	*0.066*	**0.036**	**0.049**
	$\alpha = 0.025$					$\alpha = 0.01$				
	TUFF	POF	CC	CCI	DQ	TUFF	POF	CC	CCI	DQ
beta-*t*	0.362	**0.035**	**0.019**	*0.063*	**0.027**	0.368	0.359	**0.045**	**0.021**	**0.006**
beta-*t*-rr	0.362	0.276	0.104	*0.068*	**0.007**	0.368	0.440	**0.006**	**0.002**	**0.000**
RGARCH	0.362	**0.022**	**0.000**	**0.000**	**0.000**	0.751	**0.000**	**0.000**	**0.000**	**0.000**
REGARCH	0.362	**0.005**	**0.000**	**0.000**	**0.000**	0.751	**0.000**	**0.000**	**0.000**	**0.000**
RDS-1c1l	0.124	0.601	**0.000**	**0.000**	**0.000**	0.646	0.136	**0.000**	**0.000**	**0.000**
RDS-1c2l	0.124	0.123	**0.005**	**0.004**	**0.000**	0.121	0.901	**0.003**	**0.001**	**0.000**
RDS-2c1l	*0.082*	*0.082*	**0.004**	**0.005**	**0.000**	0.646	**0.001**	**0.000**	**0.001**	**0.000**
RDS-2c2l	0.124	0.536	0.110	**0.045**	**0.002**	0.121	0.228	**0.006**	**0.003**	**0.000**
RDS-1c1l-rr	*0.082*	0.889	**0.002**	**0.000**	**0.000**	0.121	0.858	**0.005**	**0.001**	**0.000**
RDS-1c2l-rr	0.124	0.730	**0.005**	**0.001**	**0.000**	0.121	0.781	**0.026**	**0.007**	**0.005**
RDS-2c1l-rr	*0.082*	0.367	**0.004**	**0.002**	**0.000**	0.121	**0.040**	**0.001**	**0.001**	**0.000**
RDS-2c2l-rr	0.362	0.319	**0.043**	**0.022**	**0.004**	0.121	0.103	**0.001**	**0.001**	**0.000**

Note: See the notes to Table A.1 for definitions of the VaR backtests and the interpretation of the reported formatting.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics 69*(3), 542–547.

Caporin, M. (2008, 12). Evaluating value-at-risk measures in presence of long memory conditional volatility. *Journal of Risk 10*.

Christensen, K. and M. Podolskij (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics 141*(2), 323–349.

Christensen, K., M. Podolskij, and M. Vetter (2009, April). Bias-correcting the realized range-based variance in the presence of market microstructure noise. *Finance and Stochastics 13*(2), 239–268.

Christoffersen, P. and D. Pelletier (2004, 01). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics 2*(1), 84–108.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–862.

Creal, D., S. J. Koopman, and A. Lucas (2008). A general framework for observation driven time-varying parameter models. Technical Report 08-108/4, Tinbergen Institute.

Ding, Z. and C. W. Granger (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics 73*(1), 185–215.

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

Engle, R. F. and G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. *Handbook of Statistics 14*, 475–497.

Engle, R. F. and S. Manganelli (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business Economic Statistics 22*(4), 367–381.

Fernández, C. and M. F. J. Steel (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association 93*(441), 359–371.

Giot, P. and S. Laurent (2003). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics 18*(6), 641–663.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review 35*(3), 705–730.

Hansen, P. R. and Z. Huang (2016). Exponential garch modeling with realized measures of volatility. *Journal of Business & Economic Statistics 34*(2), 269–287.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.

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

Harvey, A. and D. Palumbo (2023). Score-driven models for realized volatility. *Journal of Econometrics 237*(2, Part B), 105448.

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

Harvey, A. C. and G. Sucarrat (2014). EGARCH models with fat tails, skewness and leverage. *Computational Statistics and Data Analysis 76*, 320–338.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives 3*(2), 73–84.

Martens, M. and D. van Dijk (2007). Measuring volatility with the realized range. *Journal of Econometrics 138*(1), 181–207. 50th Anniversary Econometric Institute.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Opschoor, A., M. van der Wel, and D. van Dijk (2018). Combining high-frequency data and score-driven models for volatility forecasting. *Journal of Econometrics 202*(1), 118–140.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business 53*(1), 61–65.

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