RQUF2667881 VOL 00, ISS 00

Testing and modelling time varying (a)symmetric tails

Dario Palumbo

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left are hyperlinked to the location of the query in the paper.

The title and author names are listed here as they will appear in your paper and the Table of Contents. Please check that this information is correct and let us know if any changes are needed. Also check that affiliations, funding information and conflict of interest statements are correct.

Please review your paper as a whole for typos and essential corrections. Note that we cannot accept a revised manuscript at this stage of production or major corrections, which would require Editorial review and delay publication.

Taylor & Francis is committed to increasing accessibility in line with WCAG and all figures and images will now include image descriptions (alternative text). This has been generated using AI technologies. If you would like to learn more about alt-text, please visit our Author Services website.

For more information on making corrections to the text using Adobe Acrobat, please visit https://authorservices.taylorandfrancis.com/publishing-your-research/moving-through-production/correct-article-proofs-with-adobe/

AUTHOR QUERIES

QUERY NO.	QUERY DETAILS
Q1	To improve accessibility, all images we publish now require descriptive text. AI-generated image descriptions have been created for all Figures. Please review for accuracy and relevance, making corrections as necessary.
Q2	Figure 2 was not cited in the text so a citation has been inserted. Please provide a correction if this is inaccurate.
Q3	The CrossRef database (www.crossref.org/) has been used to validate the references. Mismatches between the original manuscript and CrossRef are tracked in red font. Please provide a revision if the change is incorrect. Do not comment on correct changes.
Q4	Please provide missing city for "Embrechts *et al.* (1997)" references list entry.
Q5	Please provide missing volume number/page numbers for "Harvey (2013)" references list entry.

Quantitative Finance, 2026
https://doi.org/10.1080/14697688.2026.2667881

Testing and modelling time varying (a)symmetric tails

DARIO PALUMBO*†‡

†Department of Economics, Ca' Foscari University of Venice, Fondamenta S. Giobbe, 873, 30100 Cannaregio,Venezia, Italy
‡Homerton College, University of Cambridge, Hills Road, Cambridge CB2 8PH, United Kingdom

(*Received 7 February 2025; accepted 19 March 2026*)

The occurrence of extreme observations in a time series depends on the heaviness of the distribution's tails. This paper proposes a score-driven framework for detecting and modelling time-varying tail behaviour. The framework is based on the *t* conditional distributions and is extended to allow for asymmetric tails with distinct dynamic behaviour. In addition, the paper introduces a novel Lagrange Multiplier test to detect the presence of dynamics in the tail index parameters. The paper examines the properties of the test and demonstrates that it is more effective than existing methodologies at detecting tail variation. The framework is then applied to the tail behaviour of market returns from Equity Indices and Credit Default Swaps. The implications of neglecting dynamic tail features are assessed in terms of conditional density forecasts. The paper shows that allowing for a dynamic tail index, where appropriate, improves the forecasting accuracy of expected shortfalls and value-at-risk.

Keywords: Heavy tailed distributions; Extreme events; Score-driven models; Tail index; Lagrange multiplier test; Financial markets

JEL Classifications: C12, C18, C51, C52, C46, C58, G12

1. Introduction

The analysis of time series focuses on identifying the time-varying features of the underlying data-generating process. It has been empirically shown that the unconditional distributions of market returns are heavy-tailed, with evidence of volatility clustering and long memory. These features can be partly explained if the second moment of the conditional distribution of the data is time-varying. However, this alone is insufficient to explain how the occurrence of extreme events can vary over time It is important to accurately account for potential variations in the tails size when forecasting probability distributions of financial returns, particularly for the purpose of minimising portfolio risks and monitoring the stability of financial markets.

The occurrence of extreme events in financial data is described by tail risk. Modelling time varying tails is notably a difficult task, which becomes even harder without knowing when the tail index is actually time varying. The main contribution of this paper is to introduce a methodology to accurately identify and capture the presence of dynamic variations in the tails of time series distributions over time, distinct from scale variations. Furthermore, the paper introduces a new dynamic model capable of separating the dynamics of the upper tail from those of the lower tail.

Given the challenges of modelling the tails of the conditional distribution, it is necessary to test for the presence of dynamics before attempting to model them, in order to avoid spurious results. For this reason, the paper introduces a formal test to detect the presence of dynamic tails.

The concept of tail risk can be decomposed into two elements: the time variation in the overall length of the distribution's tails, and the relative difference in size between the upper and lower tails, which defines asymmetry. If estimated statically through moving windows exercises is possible to notice that, as the scale varies over time, also the degrees of freedom are time-varying. The accuracy of their estimation depends on the sample size and on the overall level of the tail index.† Moreover, if estimating asymmetric tails with a fitted static asymmetric *t* distribution (AST) of Zhu and Galbraith (2010) the relative magnitude and variation of the two degrees of freedom parameters can vary much more than

*Corresponding author. Email: dario.palumbo@unive.it

† For a *t* dostribution for example it is difficult to estimate accurately differences in the tail index level for degrees of freedom around 10-15 since the change in the likelihood function tends to flatten around those values.

in the symmetric case, with periods where the lower tail is heavier than the upper tail and vice versa.

These variations observed in the tail index parameters of the two tails, and in their relative asymmetry, can be consequence of incorrect estimations due to small sample issues or be actual dynamic movements, which imply in turn time variations in the higher moments of the distribution. Various observation-driven models have been proposed to directly model the higher moments of the conditional distribution of the data, focussing particularly on skewness to describe asymmetry, as in Harvey and Siddique (1999), and kurtosis for the heaviness of the tails, as in Brooks *et al.* (2005). However, as highlighted by Hansen (1994), to have valid quasi-maximum likelihood properties when modelling conditional moments, it is necessary to impose stricter restrictions on even higher conditional moments.† These conditions can be difficult to be satisfied by the data generating process in presence of heavy tails. Moreover, the moments being modelled must always exist.‡ For these reasons, Hansen (1994) suggested that the solution should be to model the shape parameters of the conditional densities directly, and outlined a general framework for doing so using an ARCH-type dynamic model.

Another approach to measuring tail variations is through extreme value theory. As described by Embrechts *et al.* (1997), this theory approximates the unconditional distribution of random variables at the lower and upper tails. This approximation allows a focus directly on the distribution of observations in the tails beyond a given threshold, which can be approximated by a generalised Pareto distribution or linked to the tail index parameter via a power law. Based on this theory, Quintos *et al.* (2001) developed formal tests to detect structural breaks in the tail index of the unconditional distribution of data, which Werner and Upper (2004) and Galbraith and Zernov (2004) used to analyse German bond futures returns and US equity returns, respectively. In this framework, the occurrence of extreme events can be modelled by assigning dynamics directly to the tail index parameter, as in Wagner (2005). However, because the estimation of the model parameters depends only on the observations beyond a given threshold, it is necessary to have long time series to accurately describe its dynamics. The problem with this approach is that§ the parameters governing the dynamics of the tail, as well as other time-varying features, may not be stable over such a long period. To address this issue, while examining tail risk in equity indices, Allen *et al.* (2012), Kelly (2014), and Kelly and Jiang (2014) developed a dynamic power law model that focuses on both the time-series and cross-sectional dimensions of the available data, exploiting the information from all stocks traded on an index.

To model the tail index, this paper suggests using models from the recent score-driven literature developed by Creal *et al.* (2013) and Harvey (2013). The motivation arises from the fact that score-driven models, besides allowing for a wider choice of conditional distributions, focus on providing dynamics directly to the parameters of the conditional distribution rather than their moments. The score that drives the dynamics is a continuous function of the observations, with an adaptive response that gives higher weights to observations at the extremes of the distribution than to those near the median. An earlier example of a score-driven framework used for modelling the tail index parameter can be found in Lucas and Zhang (2016), which developed an exponentially weighted moving average (EWMA) model for both the tail index and the skewness parameter, imposing strongly persistent time-varying behaviour. In addition, Blazsek and Monteros (2017) considered a score-driven model for the degrees of freedom of a t distribution fitted to equity returns. More recently D'Innocenzo *et al.* (2024) introduced a score-driven dynamic model for modelling dynamic tail and scale through a generalised pareto distribution in the EVT setting, derived from its asymptotic properties.

The main issue with all the aforementioned dynamic tail index models is that, to our knowledge, no formal diagnostic tools have been introduced in a dynamic setting to assess the actual presence of a dynamic tail index, thereby justifying the use of these models.

Building on this literature, the present work focuses on modelling a dynamic tail index within a score-driven framework assuming a conditional t distribution, as in Lucas and Zhang (2016). This distribution has a separate parameter to define the shape of its tails and can be further generalised to include another parameter to describe its skewness, as shown by Harvey and Lange (2017) in the case of the generalised t distribution which has as a special case the t. The paper empirically investigates the convergence properties of the model for different average values of the tail index parameter. In addition, a novel test is introduced aimed at detecting whether the tail index parameter is dynamic. The methodology is based on the Lagrange Multiplier (LM) test, which was introduced in score-driven models by Harvey (2013) and Harvey and Thiele (2016) in the context of time-varying correlation. Our test focuses on the residual correlation of the fitted scores with respect to the shape parameter of the conditional distribution under the null hypothesis of static dynamics. This test differs from that of Quintos *et al.* (2001), as it focuses on the conditional distribution of the data. Other types of LM tests for general parameter instabilities and structural breaks in score-driven frameworks have been considered by Calvori *et al.* (2017). However, as our test explicitly accounts for the cross-correlation between the scores with respect to the scale and tail index parameters under the alternative hypothesis of dynamic behaviour, the LM test displays greater power in detecting dynamics in tail index parameters. We also provide a power and size comparison with a simplified version of the LM test based on the Box-Ljung test statistic.

The final contribution of this paper is to extend the conditional t distribution to its asymmetric version, incorporating different independent time-varying tail parameters for each tail to capture asymmetry. The motivation for this is that, in

† For example, following the seminal work of Lee and Hansen (1994), the GARCH(1,1) model requires the fourth moment of the conditional distribution to exist and be finite.

‡ For example, if the variability in the data is too extreme, the tail index might be so low that it cannot guarantee the existence of skewness, kurtosis, or even variance, as in the case of a Cauchy distribution.

§ As demonstrated for the dynamics of second moments in GARCH-type models by Lamoureux and Lastrapes (2002) and Engle and Mustafa (1992).

the presence of asymmetric data, a symmetric model would incorrectly estimate the quantiles of the conditional distribution, likely underestimating the thickness of the heavier tail. In the score-driven literature, dynamic asymmetry has been introduced previously by Lucas and Zhang (2016) and more recently by Delle Monache *et al.* (2024), both of whom modelled the skewness parameter of the distribution. The issue with restricting variation to a single skewness parameter is that the relative proportion of mass across the distribution is kept constant, whereas a fully asymmetric model allows separate estimation of the thickness of each tail (see Harvey and Lange 2017). This approach not only allows for greater heterogeneity in the distribution's shape but also more parsimoniously captures the contribution of each tail to the overall asymmetry of the distribution.

The idea of dynamic asymmetry in a score-driven framework has only been studied in few cases. In the static tails framework, Thiele (2020) models a dynamic scale in the presence of the AST distribution of Zhu and Galbraith (2010), and Harvey and Lange (2017) with the generalised-t distribution. On the other hand Massacci (2017), who, following a peak-over-threshold approach from the extreme value theory approach, proposes a time-varying tail index model for the tails of the conditional distribution of the data, assuming they are conditionally Laplace-distributed. To our knowledge, this study is the first to introduce an adaptive model for the asymmetry of the full conditional distribution by independently modelling the two tail index parameters.

Finally, the paper verifies the empirical relevance of both the symmetric and asymmetric specifications in modelling market returns for equity indices and credit default swap (CDS) rates. The analysis shows that tail movements in equity indices are not particularly persistent and are driven either by the lower or upper tail, depending on whether leverage is taken into account. In contrast, for CDS rates, both tails vary independently over time with very persistent movements. The impact of tail variations on density forecasts is also assessed in terms of quantiles, testing the accuracy of the models in predicting expected shortfalls when tail dynamics are included or neglected.

The paper is structured as follows: Section 2 introduces the modelling framework. Section 3 presents the theory behind the formal test for detecting time variation in the tail index parameters, which is then analysed through simulations and compared with other relevant tests in the literature. Section 2.2 briefly introduces an extension of the statistical framework based on the generalised asymmetric t distribution. Section 4 presents the empirical results from fitting dynamic tail score-driven models to equity index and CDS returns, and analyses the out-of-sample quantiles of the forecasted conditional distributions in comparison with standard models. Section 5 provides the concluding remarks.

2. Statistical framework: score-driven dynamic tail(s) model

To set up the modelling frameworks for a dynamic tail index, we decided to focus on the simple location and scale Student's t distribution, where the degrees of freedom, η, determine the tail index. However, this approach could be applied to any location and scale, or location/scale, distribution where a shape parameter directly affects the tail index. Assuming that y_t is conditionally t-distributed, the conditional density is then,

$$f_t(\varepsilon_t) = K(\eta)\left(1+\frac{\varepsilon_t^2}{\eta}\right)^{-\frac{(\eta+1)}{2}} \qquad \eta > 0 \text{ and } -\infty < \varepsilon_t < \infty$$

$$K(\eta) = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\eta\pi}\,\Gamma\left(\frac{\eta}{2}\right)} \qquad \text{for } t = 1,\dots,T$$

where $\Gamma(.)$ is the gamma function, and $\varepsilon_t = (y_t - \mu)/\varphi$ are the residuals, with μ location and φ scale. The degrees of freedom parameter η governs the tail index.†

Harvey (2013) introduced the Beta-t-EGARCH model for modelling the scale, where the scale is expressed using an exponential link function, such that $\varphi_{t|t-1} = e^{\lambda_{t|t-1}}$, and its dynamics are captured by a score-driven model for $\lambda_{t|t-1}$. The conditional score and its information matrix with respect to the dynamic scale parameter $\lambda_{t|t-1}$ are given by

$$\frac{\partial \ln f_t}{\partial \lambda} = (\eta+1)\,b_t - 1, \quad \mathcal{I}_{\lambda\lambda} = \frac{2\eta}{3+\eta}$$

where $b_t = \frac{\varepsilon_t^2/\eta}{\varepsilon_t^2/\eta+1}$ is distributed as a Beta $(\frac{1}{2}, \frac{\eta}{2})$. The dynamics of the scale parameter $\lambda_{t|t-1}$ is then driven by the standardised conditional score $u_t^{\lambda} = \frac{\partial \ln f_t}{\partial \lambda}\mathcal{I}_{\lambda\lambda}^{-1}$. While providing asymptotic normality results for the ML estimator under correct specification, Harvey (2013) also shows how the model is more effective at modelling volatility compared to standard GARCH models and is robust to the presence of outliers in the observed series. Moreover, the model can be extended to include a dynamic location parameter.

In the present study, we assume dynamics for both the conditional scale, $\varphi_{t|t-1}$, and the tail index parameter, $\eta_{t|t-1}$, while the location μ is assumed to be 0.‡ As with the scale parameter, to ensure that $\eta_{t|t-1}$ remains strictly positive, an affine exponential link function of the form $\eta = \eta_L + e^{\eta_s \vartheta}$ can be used. As in Lucas and Zhang (2016), η_L is a lower bound for the tail index parameter. This can be used to restrict the parameter to be, for example, greater than two, ensuring the existence of variance or greater than one in empirical studies if one would evaluate expected shortfall under a time varying tail. η_s is a fixed parameter that allows us to either model η directly (for $\eta_s = 1$) or its inverse, $\bar{\eta} = 1/\eta$ (for $\eta_s = -1$).§ This approach is often useful in deriving analytical results, as

† The degrees of freedom parameter, η, is a proxy for the tail index, as defined by the CDF decomposition $\bar{F}_Y(y) = cL(y)y^{-\eta}$, where $\bar{F}Y$ is the survival function, c is a non-negative constant, and $L(y)$ is a slowly varying function such that $\lim k- > \infty \frac{L(ky)}{L(y)} = 1$. A lower tail index implies longer and fatter tails, and a higher occurrence of extreme events. A distribution with a given tail index, η, has only $k < \eta$ finite moments.

‡ This is consistent with the empirical evidence on financial series like market returns.

§ This general setup of the link function nests several specifications. For instance, if we instead model the inverse of the degrees of freedom, $\bar{\eta}$, with a logistic function that restricts it to $0 < \bar{\eta} < 1$, such

Colour online, B/W in print

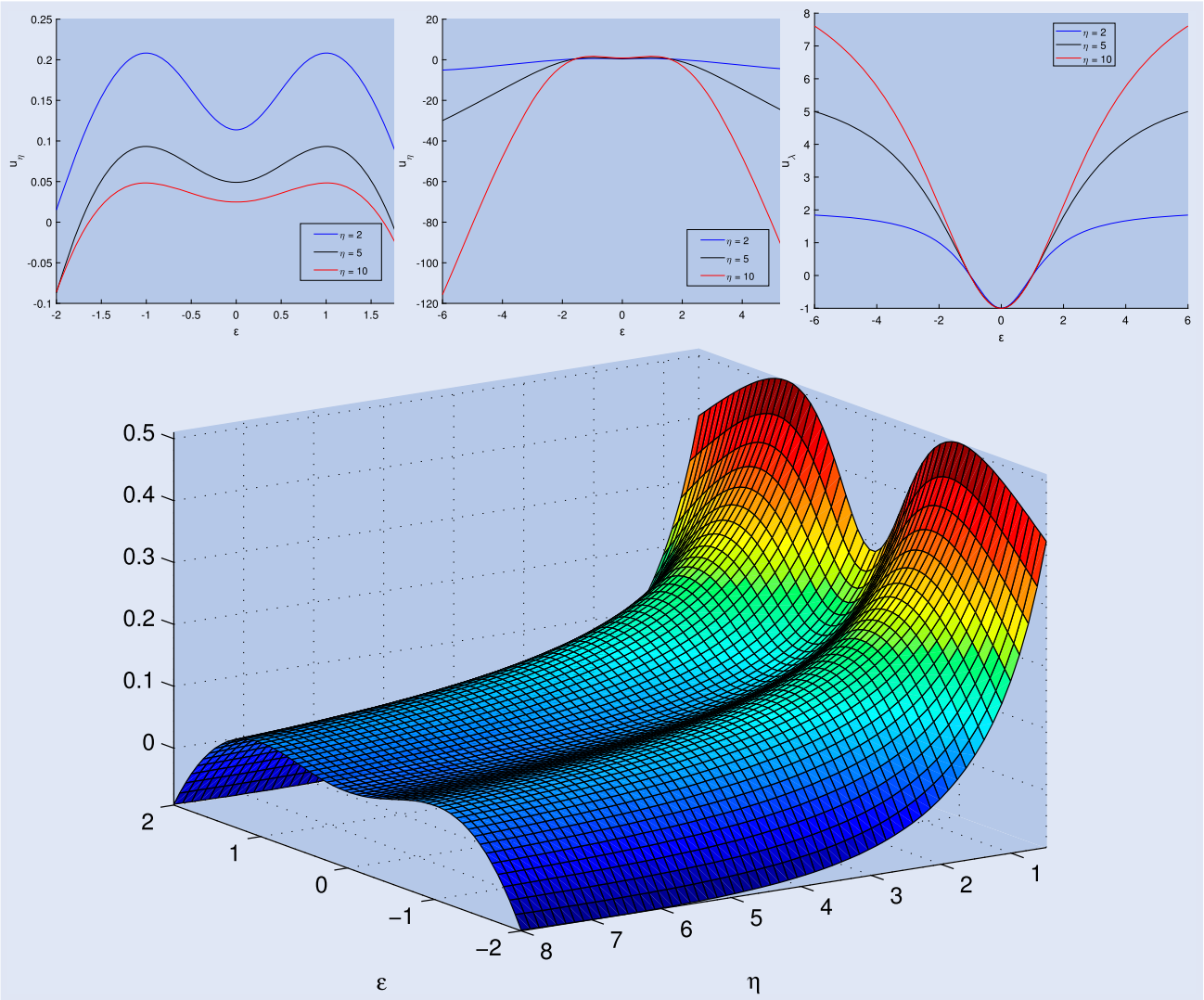


Figure 1. The figure provide the plot of the raw score with respect to ϑ (Top Left and Top Middle) and with respect to λ (Top Right) against different residuals values ε_t, for $\upsilon = 2$ and $\eta = 2$, $\eta = 6$, $\eta = 10$. On the bottom we have a three dimensional surface of score of the unbounded tail index parameter for $-2 < \varepsilon < 2$, and $1/2 < \eta < 8$.

shown by Harvey and Lange (2017) in the case of the Generalised t score-driven model. The conditional score and its information matrix with respect to ϑ can be obtained as

$$\frac{\partial \ln f_t}{\partial \vartheta} = \eta_s \frac{(\eta_{t|t-1} - \eta_L)}{2} \left[\psi\left(\frac{\eta_{t|t-1}+1}{2}\right) - \psi\left(\frac{\eta_{t|t-1}}{2}\right) + \ln(1-b_t) + \frac{1}{\eta_{t|t-1}} \frac{\partial \ln f_t}{\partial \lambda} \right] \tag{1}$$

$$\mathcal{I}_{\vartheta\vartheta} = \eta_s^2 \frac{(\eta_{t|t-1} - \eta_L)^2}{4} \left[\psi'\left(\frac{\eta_{t|t-1}}{2}\right) - \psi'\left(\frac{\eta_{t|t-1}+1}{2}\right) - \frac{2(\eta_{t|t-1}+5)}{\eta_{t|t-1}(\eta_{t|t-1}+1)(\eta_{t|t-1}+3)} \right], \tag{2}$$

where $\psi(x)$ and $\psi'(x)$ are the digamma and trigamma functions, respectively. It is noteworthy that the score with respect to λ appears in the last term of the score with respect to ϑ, highlighting the close relationship between the scale and the tail index parameters.

From figure 1, it can be observed that for values of ε_t close to the median, the response of the score with respect to ϑ tends to increase as $\eta_{t|t-1}$ falls, while the score with respect to λ remains unchanged. For large positive and negative values of ε_t, however, the score with respect to ϑ becomes unbounded, with its response increasing in magnitude as η increases. In contrast, the score with respect to λ decreases and becomes bounded for very low values of $\eta_{t|t-1}$. This result is consistent because, as the degrees of freedom increase, observations far from the median become more informative about tail behaviour. Depending on the heaviness of the fitted distribution at any given time, these observations are discounted accordingly. Ultimately, since both ε_t and $\eta_{t|t-1}$ vary over time, it is more helpful to consider the overall joint score response, which decreases and stabilises as $\eta_{t|t-1}$ becomes large.

A notable distinction from the conditional score with respect to the scale parameter λ, is that the score with respect to the tail parameter ϑ is unbounded for large values of ε_t. In

as $\bar{\eta} = \frac{\exp 2\vartheta}{1+\exp 2\vartheta}$, then $\eta = 1 + e^{-2\vartheta}$, which is our specification with $\eta_L = 1$ and $\eta_s = -2$.

principle this feature seems to go against the robustness principles of score-driven models, see Harvey (2013), Harvey and Luati (2014). However, we need to acknowledge the fact that a model for time varying tail should gain information precicely from the extreme observations that a *winsorized* score, like the one for the scale parameter, wouldn't be able to capture due to its robustness towards outliers. This feature can creates difficulties in the derivation of the invertibility properties of the filter $\vartheta_{t|t-1}$. For these purpose, one could consider substitute the score u_t^{ϑ} with a bounded function and derive the invertibility properties following the results of Blasques *et al.* (2023).

The Dynamic Scale-Tail model can be specified in the following way. Assume that y_t are conditionally distributed as

$$y_t = \mu + \varepsilon_t e^{(\lambda_{t|t-1})}, \qquad \varepsilon_t \mid \mathcal{F}_{t-1} \sim t\left(\eta_{t|t-1}\right), \quad t = 1, \ldots, T \tag{3}$$

where μ is a static location parameter and $\lambda_{t|t-1}$ the time varying parameter representing the logarithm of scale. Then ε_t are conditionally distributed with a t distribution with location 0 and scale 1 and a time varying scale index parameter $\eta_{t|t-1}$. Then a first-order score-driven dynamics for the scale and tail index parameters can be described by the following score-driven model

$$\begin{cases} \lambda_{t+1|t} = (1-\phi_\lambda)\,\omega_\lambda + \phi_\lambda \lambda_{t|t-1} + \kappa_\lambda u_t^{\lambda} \\ \vartheta_{t+1|t} = (1-\phi_\vartheta)\,\omega_\vartheta + \phi_\vartheta \vartheta_{t|t-1} + \kappa_\vartheta u_t^{\vartheta} \end{cases} t = 1, \ldots, T, \tag{4}$$

where $\boldsymbol{\psi}_\lambda = (\omega_\lambda, \phi_\lambda, \kappa_\lambda)^\intercal$ and $\boldsymbol{\psi}_\vartheta = (\omega_\vartheta, \phi_\vartheta, \kappa_\vartheta)^\intercal$ are the sets of parameters governing the dynamics of $\lambda_{t|t-1}$ and $\vartheta_{t|t-1}$, respectively. Here ω_λ and ω_ϑ are the intercepts, or long run means of the processes, ϕ_λ and ϕ_ϑ are the parameters governing the persistency of the processes while κ_λ and κ_ϑ define the impacts of the score innovations over time. Then the score innovations u_t^{λ} are the standardised conditional scores with respect to λ, as described for the Beta-t-EGARCH, while $u_t^{\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta} \mathcal{I}_{\vartheta\vartheta}^{-1}$, where $\mathcal{I}_{\vartheta\vartheta}$ is the static information matrix with respect to ϑ described in 2. The full static information matrix with respect to the two parameters is then given by,

$$\mathbf{I}\begin{pmatrix} \lambda \\ \vartheta \end{pmatrix} = \begin{bmatrix} \mathcal{I}_{\lambda\lambda} & \mathcal{I}_{\lambda\vartheta} \\ \mathcal{I}_{\lambda\vartheta} & \mathcal{I}_{\vartheta\vartheta} \end{bmatrix} \tag{5}$$

where,

$$\mathcal{I}_{\lambda\vartheta} = \eta_s \left(\eta_{t|t-1} - \eta_L\right)\left(\frac{1}{\left(\eta_{t|t-1}+3\right)} - \frac{1}{\left(\eta_{t|t-1}+1\right)}\right)$$

where its elements are independent of λ. Following Harvey (2013), to guarantee stationarity we restrict $|\phi_\lambda|, |\phi_\vartheta| < 1$ while $\kappa_\lambda, \kappa_\vartheta > 0$.

In Section 3 of the Appendix, we conduct an important analysis of the implications of bounding the tail index parameter by η_L on the score response. We identify that bounding the tail index can result in significant distortions to the score response. These distortions can ultimately affect the fit, as they make it more difficult for the dynamic parameter to move away from the bound once it is close. Moreover, we observe that, since the score function naturally pushes the dynamic parameter away from very low values, the chances of the parameter falling below the bound and remaining there are much lower when the tail index is unbounded compared to when it is bounded. As a result, for the remainder of this paper, we assume that $\eta_L = 0$ and $\eta_s = 1$.

2.1. *Extending the statistical framework: asymmetric tails modelling*

As in Harvey and Lange (2017), skewness in the t distribution can be easily introduced by defining negative and positive residuals as follows:

$$\varepsilon_t = \begin{cases} \varepsilon_t^- = \dfrac{y_t - \mu}{2\,(1-\alpha)\,\varphi}, & y_t \le \mu \\ \varepsilon_t^+ = \dfrac{y_t - \mu}{2\alpha\varphi}, & y_t > \mu \end{cases}$$

where the parameter α, with $0 < \alpha < 1$, governs the skewness; for $\alpha = 1/2$, the distribution is symmetric. The distribution can be further generalised to its asymmetric version as follows:

$$f_t(y_t) = \begin{cases} f_{1t}(y_t) = \dfrac{K_{12}}{\varphi}\left(1 + \dfrac{\left|\varepsilon_t^-\right|^2}{\eta_1}\right)^{-\frac{(\eta_1+1)}{2}}, & y_t \le \mu \\ f_{2t}(y_t) = \frac{K_{12}}{\varphi}\left(1 + \frac{\left|\varepsilon_t^+\right|^2}{\eta_2}\right)^{-\frac{(\eta_2+1)}{2}}, & y_t > \mu \end{cases} \tag{6}$$

Each η_i governs the shape of the left and right sides of the distribution. Here, $K_{12} = 1/[\alpha/K_1 + (1-\alpha)/K_2]$, with $K_i = K(\eta_i)$ for $i = 1, 2$. The distribution is symmetric if $\eta_1 = \eta_2$.

In the case of an asymmetric distribution, the score is more complex and differs for the left and right tails, as do the corresponding information matrices. For the dynamic scale parameter λ, we have

$$\frac{\partial \ln f_t}{\partial \lambda} = \begin{cases} \left(1 + \eta_{1t|t-1}\right) b_{1t} - 1, & y_t \le \mu \\ \left(1 + \eta_{2t|t-1}\right) b_{2t} - 1, & y_t > \mu \end{cases},$$

$$\mathcal{I}_{\lambda\lambda} = \begin{cases} \dfrac{2\eta_{1t|t-1}}{3 + \eta_{1t|t-1}}, & y_t \le \mu \\ \dfrac{2\eta_{2t|t-1}}{3 + \eta_{2t|t-1}}, & y_t > \mu \end{cases}$$

where $b_{1t} = \frac{|\varepsilon_t^-|^{\nu_1}/\eta_{1t|t-1}}{|\varepsilon_t^-|^{\nu_1}/\eta_{1t|t-1}+1}$ and $b_{2t} = \frac{|\varepsilon_t^+|^{\nu_2}/\eta_{2t|t-1}}{|\varepsilon_t^+|^{\nu_2}/\eta_{2t|t-1}+1}$. Then, the score becomes

$$u_t^{\lambda} = \left[\left(1 + \eta_{1t|t-1}\right) b_{1t} - 1\right] \frac{3 + \eta_{1t|t-1}}{2\eta_{1t|t-1}} \mathbf{1}_{(\varepsilon_t \le 0)} + \left[\left(1 + \eta_{2t|t-1}\right) b_{2t} - 1\right] \frac{3 + \eta_{2t|t-1}}{2\eta_{2t|t-1}} \left(1 - \mathbf{1}_{(\varepsilon_t \le 0)}\right)$$

which coincides with the asymmetric Student t (AST) score-driven model of Thiele (2020). We can now introduce score-driven dynamics to the tail index parameters as well. This is done by using the link function $\eta_{it|t-1} = \eta_i^{\dagger} + e^{\widetilde{\eta}_i \vartheta_{it|t-1}}$, where $\widetilde{\eta}_i = 1$ when modelling the tail index parameter, and $\widetilde{\eta}_i = -1$ when modelling its inverse. The conditional scores

Colour online, B/W in print

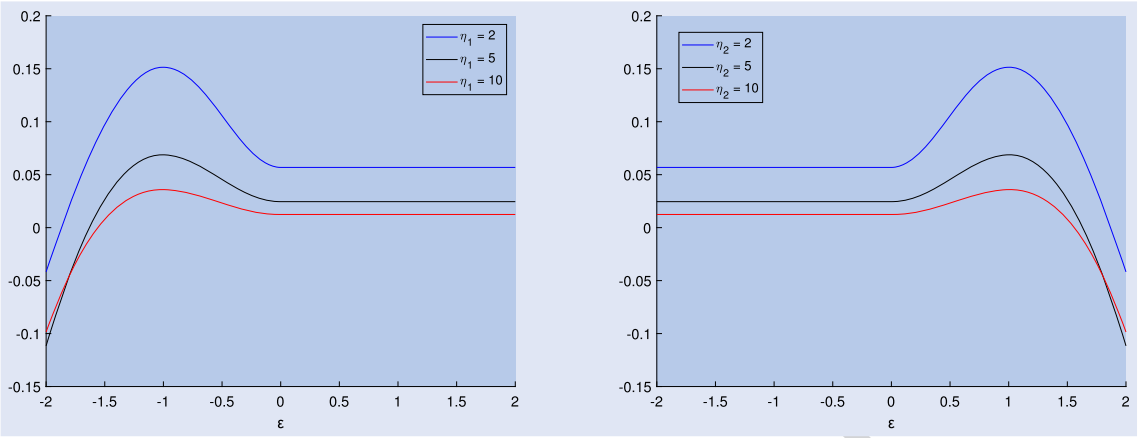


Figure 2. Plot of the score with respect to different residuals values ε_t for $\upsilon_1 = \upsilon_2 = 2$ and $\eta_1 = \eta_2 = 2$, $\eta_1 = \eta_2 = 6$, $\eta_1 = \eta_2 = 10$. The score with respect to η_1 (Left) and the score with respect to η_2 (Right).

with respect to the dynamic tail index parameters are then (figure 2).

$$\frac{\partial \ln f_t}{\partial \vartheta_1} = \begin{cases} \widetilde{\eta}_1 \dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{2} \left[\alpha^{+}\tau_{1t} + \ln\left(1 - b_{1t}\right) + \dfrac{1}{\eta_1}\left(\dfrac{\partial \ln f_{1t}}{\partial \lambda} + 1\right)\right], & y_t \leq \mu \\ \widetilde{\eta}_1 \dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{2} \alpha^{+}\tau_{1t}, & y_t > \mu \end{cases}, \quad (7)$$

$$\frac{\partial \ln f_t}{\partial \vartheta_2} = \begin{cases} \widetilde{\eta}_2 \dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{2} \left(1 - \alpha^{+}\right)\tau_{2t}, & y_t \leq \mu \\ \widetilde{\eta}_2 \dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{2} \left[\left(1 - \alpha^{+}\right)\tau_{2t} + \ln\left(1 - b_{2t}\right) + \dfrac{1}{\eta_2}\left(\dfrac{\partial \ln f_{2t}}{\partial \lambda} + 1\right)\right], & y_t > \mu \end{cases}, \quad (8)$$

The information quantities are as follows:

$$\mathcal{I}_{1\vartheta\vartheta} = \begin{cases} \widetilde{\eta}_1^2 \dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{2} \left(\left[\left(1 - \alpha_t^{+}\right)\tau_{1t} + \alpha_t^{+}\left(1 - \alpha_t^{+}\right)\dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{2}\tau_{1t}^2 - \dfrac{\alpha_t^{+}}{\widetilde{\eta}_1}\dfrac{\partial \tau_{1t}}{\partial \vartheta_1}\right] - \dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{\eta_{1t|t-1}^2} \dfrac{\eta_{1t|t-1} - 3}{\left(1 + \eta_{1t|t-1}\right)\left(3 + \eta_{1t|t-1}\right)}\right), & y_t \leq \mu \\ \widetilde{\eta}_1^2 \dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{2} \left[\alpha_t^{+}\left(1 - \alpha_t^{+}\right)\dfrac{\left(\eta_{1t|t-1} - \eta_1^{\dagger}\right)}{2}\tau_{1t}^2 - \dfrac{\alpha_t^{+}}{\widetilde{\eta}_1}\dfrac{\partial \tau_{1t}}{\partial \vartheta_1} - \alpha_t^{+}\tau_{1t}\right], & y_t > \mu \end{cases} \quad (9)$$

$$\mathcal{I}_{2\vartheta\vartheta} = \begin{cases} \widetilde{\eta}_2^2 \dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{2} \left[\alpha_t^{+}\left(1 - \alpha_t^{+}\right)\dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{2}\tau_{2t}^2 - \dfrac{\left(1 - \alpha_t^{+}\right)}{\widetilde{\eta}_2}\dfrac{\partial \tau_{2t}}{\partial \vartheta_2} - \left(1 - \alpha_t^{+}\right)\tau_{2t}\right], & y_t \leq \mu \\ \widetilde{\eta}_2^2 \dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{2} \left(\left[\alpha_t^{+}\tau_{2t} + \alpha_t^{+}\left(1 - \alpha_t^{+}\right)\dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{2}\tau_{2t}^2 - \dfrac{\left(1 - \alpha_t^{+}\right)}{\widetilde{\eta}_2}\dfrac{\partial \tau_{2t}}{\partial \vartheta_2}\right] - \dfrac{\left(\eta_{2t|t-1} - \eta_2^{\dagger}\right)}{\eta_{2t|t-1}^2} \dfrac{\eta_{2t|t-1} - 3}{\left(1 + \eta_{2t|t-1}\right)\left(3 + \eta_{2t|t-1}\right)}\right), & y_t > \mu \end{cases} \quad (10)$$

The asymmetry mixing parameter α^{+} is then defined as

$$\alpha_t^{+} = \frac{\alpha / K_{1t}}{\alpha / K_{1t} + (1 - \alpha) / K_{2t}}$$

which, as noted by Harvey and Lange (2017), coincides with the probability under the distribution of having a negative observation. The parameters τ_{it} and their derivatives are defined as follows,

$$\tau_{it} = \psi\left(\frac{\eta_{it|t-1} + 1}{2}\right) - \psi\left(\frac{\eta_{it|t-1}}{2}\right) - \frac{1}{\eta_{it|t-1}}$$

$$\frac{\partial \tau_{it}}{\partial \vartheta_i} = \widetilde{\eta}_i \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{2} \left[\psi'\left(\frac{\eta_{it|t-1} + 1}{2}\right) - \psi'\left(\frac{\eta_{it|t-1}}{2}\right) + \frac{2}{\eta_{it|t-1}^2}\right]$$

Then, the dynamic score-driven specification for each $\vartheta_{it|t-1}$ can be specified as

$$\vartheta_{it+1|t} = \left(1 - \phi_{i\vartheta}\right)\omega_{i\vartheta} + \phi_{i\vartheta}\vartheta_{it|t-1} + \kappa_{i\vartheta}u_{it}^{\vartheta} \quad t = 1, \ldots, T$$

where $\boldsymbol{\psi}_{i\vartheta} = (\omega_{i\vartheta}, \phi_{i\vartheta}, \kappa_{i\vartheta})^{\intercal}$ are the parameters governing the dynamics of each tail parameter ϑ_i for $i = 1, 2$. Here, $u_{it}^{\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_i}\mathcal{I}_{i\vartheta\vartheta}^{-1}$ for $i = 1, 2$. Also in this case we apply the same restrictions imposed when modelling the symmetric tail index, $|\phi_{i\vartheta}| < 1$ and $\kappa_{i\vartheta} > 0$.

The score updates provide interesting insights. Existing peak-over-threshold models in the extreme value theory literature focus only on observations that exceed a pre-determined threshold and are therefore considered as 'belonging to the tail' of the distribution. This approach treats 'non-tail observations', or, in the case of asymmetric tails, observations falling in the opposite tail, as missing.† In contrast, the score-driven framework uses score innovations in equations (7) and (8) for each of the tail index parameters, directly affected by the residuals on their respective sides of the distribution. If the residual falls in the opposite tail, the score responds as if the residual appeared at the median of the distribution. Furthermore, because the response depends on both α_t^+ and τ_{it}, it varies depending on the levels of both tail indexes and incorporates information from both tails up to that point. As a result, the score remains time-varying, dependent on both $\eta_{1t|t-1}$ and $\eta_{2t|t-1}$. We acknowledge that, as much as the threshold in a peak-over-threshold model, the value μ at which the two distributions patches together has to be determined, and can be estimated from the data, see Zhu and Galbraith (2010), Harvey and Lange (2017) and Thiele (2020). Nevertheless, the advantage of the proposed model is that it naturally uses more information from observations on both sides of the distribution compared to the extreme value theory approach.

Finally, following Zhu and Galbraith (2010), we can construct the log-likelihood function as follows,

$$
\begin{aligned}
&L\left(\mu, \boldsymbol{\psi}_{\lambda}, \quad \boldsymbol{\psi}_{1\vartheta}, \quad \boldsymbol{\psi}_{2\vartheta}, \quad \alpha\right) \\
&= -\sum_{t=1}^{T} \lambda_{t|t-1} + \sum_{t=1}^{T} \ln K_{12}\left(\eta_{1t|t-1}, \quad \eta_{2t|t-1}\right) \\
&\quad - \sum_{t=1}^{T} \frac{\left(\eta_{1t|t-1}+1\right)}{2} \ln\left(1+\frac{|\varepsilon_t|^2}{\eta_{1t|t-1}}\right) \mathbf{1}_{(\varepsilon_t \leq 0)} \\
&\quad - \sum_{t=1}^{T} \frac{\left(\eta_{2t|t-1}+1\right)}{2} \ln\left(1+\frac{|\varepsilon_t|^2}{\eta_{2t|t-1}}\right) \left(1-\mathbf{1}_{(\varepsilon_t \leq 0)}\right)
\end{aligned}
$$

where $\boldsymbol{\psi}_{\lambda}$, $\boldsymbol{\psi}_{1\vartheta}$, and $\boldsymbol{\psi}_{2\vartheta}$ are the vectors containing the parameters for the dynamic specifications of $\lambda_{t|t-1}$, $\vartheta_{1t|t-1}$, and $\vartheta_{2t|t-1}$. All parameters can be estimated using maximum likelihood.

In order to focus solely on the asymmetry between the two tail indexes of the conditional distribution of the data, we assume $\alpha = 1/2$ for the remainder of the paper.

2.2. A generalisation of the tails' shape: the generalised asymmetric t model

As discussed in Harvey and Lange (2017), the Generalised t distribution introduces greater flexibility in both tail shapes and the modelling of time series,

$$
f_t\left(y_t\right)
$$

$$
= \begin{cases} f_{1t}\left(y_t\right) \\ = \frac{K_{12}}{\varphi}\left(1+\frac{\left|\varepsilon_t^-\right|^{\upsilon_1}}{\eta_1}\right)^{-\frac{(\eta_1+1)}{\upsilon_1}}, & y_t \leq \mu \\ f_{2t}\left(y_t\right) \\ = \frac{K_{12}}{\varphi}\left(1+\frac{\left|\varepsilon_t^+\right|^{\upsilon_2}}{\eta_2}\right)^{-\frac{(\eta_2+1)}{\upsilon_2}}, & y_t > \mu \end{cases} \qquad \eta_1, \eta_2, \upsilon_1, \upsilon_2 > 0
$$

Here, $K_{12} = 1/[\alpha/K_1 + (1-\alpha)/K_2]$, with $K_i = K(\eta_i, \upsilon_i)$ for $i = 1, 2$, where

$$
K\left(\eta_i, \upsilon_i\right) = \frac{\upsilon_i}{2\eta_i^{1/\upsilon_i}} \frac{1}{B\left(1/\upsilon_i, \eta_i/\upsilon_i\right)}
$$

In this general distribution, each of η_i and υ_i governs the shape of the left and right sides of the distribution. Specifically, η_i controls the left and right tail indices. The distribution is symmetric if $\eta_1 = \eta_2$ and $\upsilon_1 = \upsilon_2$. It can exhibit fat tails when $\upsilon > 1$ and heavy but not fat tails when $0 < \upsilon < 1$. For $\eta \to \infty$, the distribution converges to a Generalised Error Distribution (GED) with shape parameter υ, which becomes a Laplace distribution when $\upsilon = 1$ and a normal distribution when $\upsilon = 2$. For $\upsilon = 2$, it becomes a t distribution with η degrees of freedom, as introduced in 2.1. This level of flexibility makes the Generalised t distribution an appealing alternative to the asymmetric t distribution. However, given the focus of this paper, we demonstrate our testing approach using the simpler asymmetric t distribution. Nonetheless, we have derived tests also for the Generalised t distribution, following the same lines as the developments in the following sections, albeit with more involved calculations. These results are aimed at future research aimed at investigating the relationship between time-varying tails in a more general setting and can be obtained upon request.

3. Detecting time varying dynamics in tail index parameters

3.1. The LM approach

Testing techniques for detecting the presence of dynamics in the parameters of a score-driven model have been presented for dynamic correlation by Harvey and Thiele (2016). Following their approach, in the case of a single time-varying parameter, denoted by ϑ, with dynamics as in equation (4) driven by its unstandardised conditional score,† the conditional score is given by

$$
u_t^{\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}, \quad t = 1, \ldots, T,
$$

where f_t denotes the conditional distribution of the t-th observation, y_t, at time t.

† For example, at time t, the residual $\varepsilon_t > 0$ would be considered a 'missing observation' when modelling the lower tail parameter.

† For simplicity, we limit our exposition of the LM test statistic to the case of unstandardised scores. However, when working with standardised scores, if the information matrix with respect to the time-varying parameters is only dependent on shape parameters like the tail index, under the null hypothesis of a static tail index, the information matrix elements are fixed scalars. Thus, the 'update' part of equation (4) can be rewritten as $\kappa_{\vartheta}\hat{u}_t^{\vartheta} = \kappa_{\vartheta}\mathcal{I}_{\vartheta\vartheta}^{-1}\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}} = \hat{\kappa}_{\vartheta}u_t^{\vartheta}$. This applies to both the scale and tail index in our case.

A test for detecting dynamics in an otherwise static tail index model, which tests the null hypothesis that $H_0 : \bar{\eta} = e^{\omega_\vartheta}$ versus the alternative hypothesis that $H_1 : \eta_{t|t-1} = e^{\vartheta_{t|t-1}}$, can be based on the Portmanteau or Box and Pierce (1970) statistic as follows,

$$Q_u(P) = T \sum_{j=1}^{P} r^2_{\vartheta u}(j) \sim \chi^2_P, \tag{11}$$

where $r_{\vartheta u}(j)$ denotes the sample autocorrelation of u^ϑ_t at lag j. When u^ϑ_t corresponds to the *true* conditional score evaluated at the true parameter value under H_0, the score sequence is i.i.d. and the statistic $Q_u(P)$ has an *exact* χ^2_P null distribution. This result provides the theoretical benchmark for Portmanteau-type tests.

In practice, however, the score must be evaluated using estimated parameters. In that case, $Q_u(P)$ is no longer exactly chi-square distributed in finite samples, but retains a χ^2_P distribution asymptotically under standard regularity conditions. Rejection of H_0 for some lag order P indicates residual autocorrelation in the score sequence and hence misspecification of the static tail index model.

Since equation (4) is not identifiable under the null hypothesis $\phi_\vartheta = \kappa_\vartheta = 0$, the Box-Pierce test can be derived as an LM test reformulating the dynamics of $\vartheta_{t|t-1}$ in its finite $MA(P)$ approximate representation. Then the Portmanteau test is the interpreted as an LM test under the null hypothesis that $\kappa_{\vartheta 0} = \kappa_{\vartheta 1} = \ldots = \kappa_{\vartheta P-1} = 0$, against the alternative hypothesis $\kappa_{\vartheta i} \geq 0$, for $i = 0, \ldots, P-1$, in the model,

$$\vartheta_{t+1|t} = \omega_\vartheta + \kappa_{\vartheta 0} u^\vartheta_t + \cdots + \kappa_{\vartheta P-1} u^\vartheta_{t+1-P}, \quad t = 1, \ldots, T.$$

From the seminal work of Hosking (1980, 12), this approach to use the approximate finite MA representation of the first-order model is standard in the literature of dynamics detection and Portmanteau type of tests, see Li and Mak (1994) and Ling and Li (1997) for conditional heteroscedasticity model as well as Harvey (2013) and Harvey and Thiele (2016) for score-driven models. The LM test statistic is then defined as

$$LM_u(P) = \frac{1}{T} \begin{bmatrix} \mathbf{0}^\intercal & \partial \ln L / \partial \boldsymbol{\kappa}^\intercal_\vartheta \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} & \boldsymbol{\Psi}_{\boldsymbol{\theta\kappa}} \\ \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} & \boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \partial \ln L / \partial \boldsymbol{\kappa}_\vartheta \end{bmatrix}, \tag{12}$$

where $\boldsymbol{\Psi}_{\mathbf{ij}^\intercal} = E[\frac{\partial \ln L}{\partial \mathbf{i}} \frac{\partial \ln L}{\partial \mathbf{j}^\intercal}]$, with $\mathbf{i}$ and $\mathbf{j}$ representing either the vectors $\boldsymbol{\kappa}$ or $\boldsymbol{\theta}$, where $\boldsymbol{\kappa} = (\kappa_{\vartheta 0}, \kappa_{\vartheta 1}, \ldots, \kappa_{\vartheta P-1})$ and $\boldsymbol{\theta}$ is the vector of all other parameters, which in this case consists solely of ω_ϑ. Under these conditions, we show in the appendix that

$$LM_u(P) = \frac{1}{T} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}^\intercal_\vartheta} \boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta} = T \sum_{j=1}^{P} r^2_{\vartheta u}(j) \tag{13}$$

This is the same result obtained by Harvey and Thiele (2016) in the case of dynamic correlation. However, the LM test statistic simplifies to equation (13) only if $\vartheta_{t|t-1}$ is the only time-varying parameter under the alternative hypothesis and there are no other time-invariant parameters to be estimated in the conditional distribution of the data. This is infeasible in practice since, given the dependence structure identified in the static information matrix in 5, in principle it is not possible to model independently $\lambda_{t|t-1}$, the log-scale parameter, and $\vartheta_{t|t-1}$ if the other parameter is actually time varying. Then the LM test statistic becomes,†

$$LM_u(P) = \frac{1}{T} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}^\intercal_\vartheta} \boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta} + \frac{1}{T} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}^\intercal_\vartheta} \left[\boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}} \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} \left(\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} - \boldsymbol{\Psi}^\intercal_{\boldsymbol{\kappa\theta}} \boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}} \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} \right)^{-1} \boldsymbol{\Psi}^\intercal_{\boldsymbol{\kappa\theta}} \boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}} \right] \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta}, \tag{14}$$

Given that the model under the alternative is a score-driven model where both the scale and the tail index are time-varying the model under the null of static dynamics for $\vartheta_{t|t-1}$ is then a Beta-t-EGARCH model, defined in Harvey (2013) as a score-driven model for dynamic scale assuming a conditional t distribution. The invertibility and asymptotic properties of the ML estimator for this model have been already definied by Harvey (2013) under correct specification, and can be easily extended with the results of Blasques *et al.* (2022). Given that we need only the asymptotic normality of the ML estimator of the model under the null to derive the asymptotic distribution of an LM test statistic, a test for detecting dynamics in the tail index $\eta_{t|t-1}$ can be constructed as defined in the proposition,

PROPOSITION 3.1 *If the data generating process y_t is*

$$y_t \mid \mathcal{F}_{t-1} \sim \mathrm{t}\left(\varphi_{t|t-1}, \eta_{t|t-1}\right), \qquad t = 1, \ldots, T$$

and the dynamic scale $\varphi_{t|t-1} = \exp(\lambda_{t|t-1})$ is fitted through maximum likelihood using a restricted model with static degrees of freedom η with dynamics,

$$\lambda_{t+1|t} = (1 - \phi_\lambda)\, \omega_\lambda + \phi_\lambda \lambda_{t|t-1} + \kappa_\lambda u^\lambda_t, \quad t = 1, \ldots, T.,$$

the Lagrange Multiplier test for the presence of dynamics in $\eta_{t|t-1} = \eta_L + \exp(\eta_s \vartheta_{t|t-1})$ under the null hypothesis

$$H_0 : \kappa_{\vartheta 0} = \kappa_{\vartheta 1} = \ldots = \kappa_{\vartheta P-1} = 0 \quad \text{against}$$
$$H_1 : \kappa_{\vartheta i} > 0, i = 0, \ldots, P-1,$$

in the dynamic equation,

$$\vartheta_{t+1|t} = \omega_\vartheta + \kappa_{\vartheta 0} u^\vartheta_t + \cdots + \kappa_{\vartheta P-1} u^\vartheta_{t+(P-1)}, \quad t = 1, \ldots, T.,$$

takes the form

$$LM_u(P) = T\gamma_1 \sum_{j=0}^{P-1} \left(r_{\vartheta u}(j+1) + d_{j\lambda} \right)^2 + T \left(\gamma_2 + z_1 \frac{\mathcal{I}^2_{\lambda\vartheta}}{\mathcal{I}^2_{\vartheta\vartheta}} \right) \left(\sum_{j=0}^{P-1} a^j d_{j\lambda} + \sum_{j=0}^{P-1} a^j r_{\vartheta u}(j+1) \right)^2$$

† See Harvey and Thiele (2016) for details.

$$+ T\,(z_2 + z_3)\,\frac{\mathcal{I}^2_{\lambda\vartheta}}{\mathcal{I}^2_{\vartheta\vartheta}}\left(\sum_{j=0}^{P-1} a^j d_{j\lambda} + \sum_{j=0}^{P-1} a^j r_{\vartheta u}(j+1)\right)$$
$$\left(\sum_{j=0}^{P-1} \phi^j_\lambda d_{j\lambda} + \sum_{j=0}^{P-1} \phi^j_\lambda r_{\vartheta u}(j+1)\right)$$
$$+ T z_4 \frac{\mathcal{I}^2_{\lambda\vartheta}}{\mathcal{I}^2_{\vartheta\vartheta}}\left(\sum_{j=0}^{P-1} \phi^j_\lambda d_{j\lambda} + \sum_{j=0}^{P-1} \phi^j_\lambda r_{\vartheta u}(j+1)\right)^2$$

where $r_{\vartheta u}(j+1)$ are the sample autocorrelations of the fitted scores with respect to ϑ while the quantities $d_{j\lambda} = \frac{1}{T}\sum u^\lambda_t \frac{\partial \lambda_{t|t-1}}{\partial \kappa_{\vartheta j}} / \mathcal{I}_{\vartheta\vartheta}$ also both computed under the null. Then $a = \phi_\lambda - \kappa_\lambda \mathcal{I}_{\lambda\lambda}$, the constants γ_1, γ_2, z_1, z_2, z_3, z_4 are derived in the Appendix and depend only on the parameters of the restricted model under the null $\boldsymbol{\theta}_0 = (\omega_\lambda, \phi_\lambda, \kappa_\lambda, \omega_\vartheta)^\intercal$ and the number of lags P, while $\mathcal{I}_{\lambda\lambda}$, $\mathcal{I}_{\lambda\vartheta}$, $\mathcal{I}_{\vartheta\vartheta}$ are the information quantities with respect to the parameters λ and ϑ of the model under the null.

When all the parameters of the model under the null, $\boldsymbol{\theta}_0$, are evaluated at the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{ML} = (\hat\omega_\lambda, \hat\phi_\lambda, \hat\kappa_\lambda, \hat\omega_\vartheta)^\intercal$, and the usual regularity and consistency conditions for $\hat{\boldsymbol{\theta}}_{ML}$ are satisfied, the LM statistic $LM_u(P)$ has an asymptotic χ^2_P null distribution, with P degrees of freedom.

We refer to this test statistic as the *full* $LM_u(P)$ and the proof of its derivation can be found in the Appendix. We can see that, as the basic test statistic in equation (11), it also concerns squared sum of the autocorrelations with respect to u^ϑ_t, weigted by the exponentially decaying weights a^i and ϕ^i_λ. The additional components $d_{j\lambda}$ capture the impact on u^λ_t of the presence of dynamics in η and are represented as the sample covariance between u^λ_t and $\frac{\partial \lambda_{t|t-1}}{\partial \kappa_{\vartheta j}}$. This last quantity can be computer recursively from the data as described in the appendix. Expanding the square of the first term, we can see that the basic $Q_u(P)$ is contained in the full specification of the test and remains one of its components.

Given the complexity of the *full* statistic $LM_u(P)$, and the fact that under the null hypothesis H_0 the score innovations u^ϑ_t would need to behave as an i.i.d. sequence driven by the true (and hence infeasible) DGP parameters, we propose a simpler and fully operational alternative for detecting time variation in $\eta_{t|t-1}$. The idea is to employ the Portmanteau statistic $Q_u(P)$, but computed from *fitted* scores constructed in a way that removes the dependence on the additional dynamic parameters.

Specifically, we define

$$\hat u^{\vartheta\dagger}_t = \frac{\hat{\tilde\eta}^\dagger}{2}\left[\psi\left(\frac{\hat{\tilde\eta}^\dagger + 1}{2}\right) - \psi\left(\frac{\hat{\tilde\eta}^\dagger}{2}\right) + \ln(1 - \hat b^\dagger_t) + \frac{\hat{\tilde\eta}^\dagger + 1}{\hat{\tilde\eta}^\dagger}\hat b^\dagger_t - \frac{1}{\hat{\tilde\eta}^\dagger}\right], \tag{15}$$

where

$$\hat b^\dagger_t = \frac{\hat\varepsilon^2_t / \hat{\tilde\eta}^\dagger}{1 + \hat\varepsilon^2_t / \hat{\tilde\eta}^\dagger}, \quad \hat\varepsilon_t = \frac{y_t}{\hat\varphi_{t|t-1}}.$$

The residuals $\hat\varepsilon_t$ are obtained from a preliminary Beta–*t*–EGARCH fit, and $\hat{\tilde\eta}^\dagger$ is the tail-index estimate from a static Student–*t* distribution fitted to $\hat\varepsilon_t$. This two-step construction "cleans" the fitted scores $\hat u^{\vartheta\dagger}_t$ from the influence of the dynamic components of the model, yielding a feasible statistic that is close in spirit to the idealised version based on the true score. We refer to the resulting statistic as the *simple* $Q_{u^\dagger}(P)$ test. In empirical work, we use the standard Box–Ljung modification,

$$Q^*_{u^\dagger}(P) = T(T+2)\sum_{j=1}^{P}(T-j)^{-1} r^2_{u^\dagger}(j).$$

In light of Proposition 3.1, and following the general results of Hansen (2021) on quadratic criterion-based tests, the limiting null distribution of both $Q_{u^\dagger}(P)$ and $Q^*_{u^\dagger}(P)$ is expected to take the form of a *weighted chi-square distribution*,† i.e.

$$\sum_{j=1}^{P} \lambda_j \chi^2_{1,j},$$

where the weights λ_j depend on nuisance parameters associated with the first-step estimation of $\hat\varphi_{t|t-1}$ and $\hat{\tilde\eta}^\dagger$. Because these nuisance quantities are unknown and difficult to estimate consistently, the resulting asymptotic distribution is non-pivotal, and analytic critical values are not operational for applied work. For this reason we suggest to rely on *block bootstrap* methods, which provide a fully feasible approximation to the null distribution of the test statistics. Several approaches have been proposed in the literature, including Lahiri (2003), Politis and White (2004), and Hall *et al.* (1995). When heavy-tailed parameters are estimated, the bootstrap's performance is highly sensitive to block-length selection (Lahiri 1999, Cavaliere *et al.* 2018). We therefore recommend the single (BOB) and double (DBOB) block bootstrap procedures of Horowitz *et al.* (2006), developed precisely for Box–Pierce-type autocorrelation tests. Notably the DBOB procedure is particularly robust to the choice of block length.

Following the reasoning in Hansen (2021), a theoretically first-order valid bootstrap would require re-estimating both the Beta–*t*–EGARCH model and the static *t* distribution within each bootstrap replication, so that the bootstrap analogue of the statistic correctly reflects the joint sampling variation of the cleaned scores and the first-step estimators. However, repeatedly estimating a heavy-tailed volatility model is computationally burdensome. For this reason, and in line with common practice for Portmanteau-type tests, we also consider a simplified bootstrap in which the first-step parameters are kept fixed at their full-sample estimates. This approximation does not fully replicate the asymptotic error propagation implied by the theory, but our Monte Carlo evidence indicates that, for the sample sizes relevant in our

† This result follows from the fact that $Q^*_{u^\dagger}(P)$ is asymptotically a quadratic form in an estimated score vector. By Lemma 1 and the general discussion in Hansen (2021, pp. 2–4), together with the examples in Section 4, the limiting null distribution of such statistics is a weighted chi-square with non-pivotal weights depending on nuisance parameters.

application, the resulting size distortions are negligible. Consequently, the simplified bootstrap offers a computationally attractive and empirically accurate alternative, while the re-estimation bootstrap remains available when full first-order validity is required.

REMARK 3.1 Instead of fixing P, it may be selected using a consistent information criterion as in Escanciano and Lobato (2009). Under the exact χ^2_P limit, only the first lag is selected with probability one in large samples, implying that the asymptotic null distribution reduces to χ^2_1. However, to avoid the computationally intensive nature of bootstrap resampling, the information criterion may also be used to verify whether the empirical distribution of the simple $Q_{u^\dagger}(P)$ or $Q^*_{u^\dagger}(P)$ is close to the χ^2_P benchmark, in which case standard chi-square critical values may be used.

Overal, when the tail is highly persistent, i.e. when the parameter ϕ_ϑ in equation (4) is close to one, larger values of P may result in more powerful tests. Then, in these cases another alternative approach suggested by Harvey (2013) is the test proposed by Nyblom (1989), which is a general test for parameter constancy against a random walk alternative, based on the LM principle. In the present context, the statistic is based on the same scores as the Box-Pierce test and can be written as

$$N_{u^\dagger} = \frac{1}{T^2\sigma^{2\dagger}_{\vartheta u}} \sum_{j=1}^{T} \left(\sum_{k=j}^{T} u_k^{\vartheta\dagger} \right)^2 .$$

where $\sigma^{2\dagger}_{\vartheta u} = \mathcal{I}^{\dagger}_{\vartheta\vartheta}$ computed with $\hat{\tilde{\eta}}^{\dagger}$. Under the null hypothesis of parameter constancy, $N_{u^\dagger}$ has approximately a Cramer-von Mises distribution. Although the Nyblom test is generally regarded as a test against a random walk alternative, it can also be interpreted as a test against a very persistent, but stationary, alternative.†

In the following sections, we will investigate the performance of both the *simple* $Q_{u^\dagger}(P)$ test and the $LM_u(P)$ test and their alternatives.

3.2. Tests simulation study

Here, the power and size of the tests are assessed under various assumptions for parameters. For this purpose, we have designed a simulation study. We have generated $N = 1000$ samples of length $T = 500$, 1000, 2000 assuming that the data generating process is conditionally distributed with a t distribution with dynamic scale, $\varphi_{t|t-1} = e^{\lambda_{t|t-1}}$, and dynamic degrees of freedom, $\eta_{t|t-1} = e^{\vartheta_{t|t-1}}$. The dynamics of the two parameters are modelled using an exponential link function in a score-driven framework with a dynamics described in equation (4) with $\omega_\lambda = -4.7$, $\phi_\lambda = 0.985$, $\kappa_\lambda = 0.03$, while ω_ϑ, ϕ_ϑ and κ_ϑ vary between simulations. The exact specifications are described in figures 3–5. In particular, we have $\omega_\vartheta = \log 2$ and $\log 8$. For the size of the test, we repeat the simulations with the same parameters just assuming $\phi_\vartheta = \kappa_\vartheta = 0$.

To perform the tests, we first fit the Beta-t-EGARCH model, as described in Harvey (2013). From this we obtain the fitted parameters, the fitted scores $\hat{u}^{\lambda}_t$ with respect to the scale parameter, the fitted scores with respect to the tail parameter $\hat{u}^{\vartheta}_t$ under the null of static η and with these we compute the full LM test statistic for various P.

At the same time, we also obtain the fitted scale $\hat{\varphi}_{t|t-1}$, and the fitted residuals, as $\hat{\varepsilon}_t = y_t e^{-\hat{\lambda}_{t|t-1}}$. Then we re-estimate the $\hat{\eta}$ parameter with a static t distribution fitted on the $\hat{\varepsilon}_t$ to compute the *clensed* fitted scored $\hat{u}^{\vartheta\dagger}_t$ with respect to ϑ under the null hypothesis of static tail described in equation (15). With these we construct the simple Box-Pierce test, referred to as the *simple* Q^* test statistic here presented in its more robust Box-Ljung version, and we compare the results of the two tests.

For the block bootstrap, we aim to adopt a block length of order $T^{1/3}$, as recommended in the literature on resampling for dependent data (see, e.g. Lahiri 2003, Hall *et al.* 1995, Politis and White 2004). In particular, we set

$$\ell = \lceil 2T^{1/3} \rceil ,$$

which provides a practical rule-of-thumb consistent with the asymptotically optimal rate in the literature.

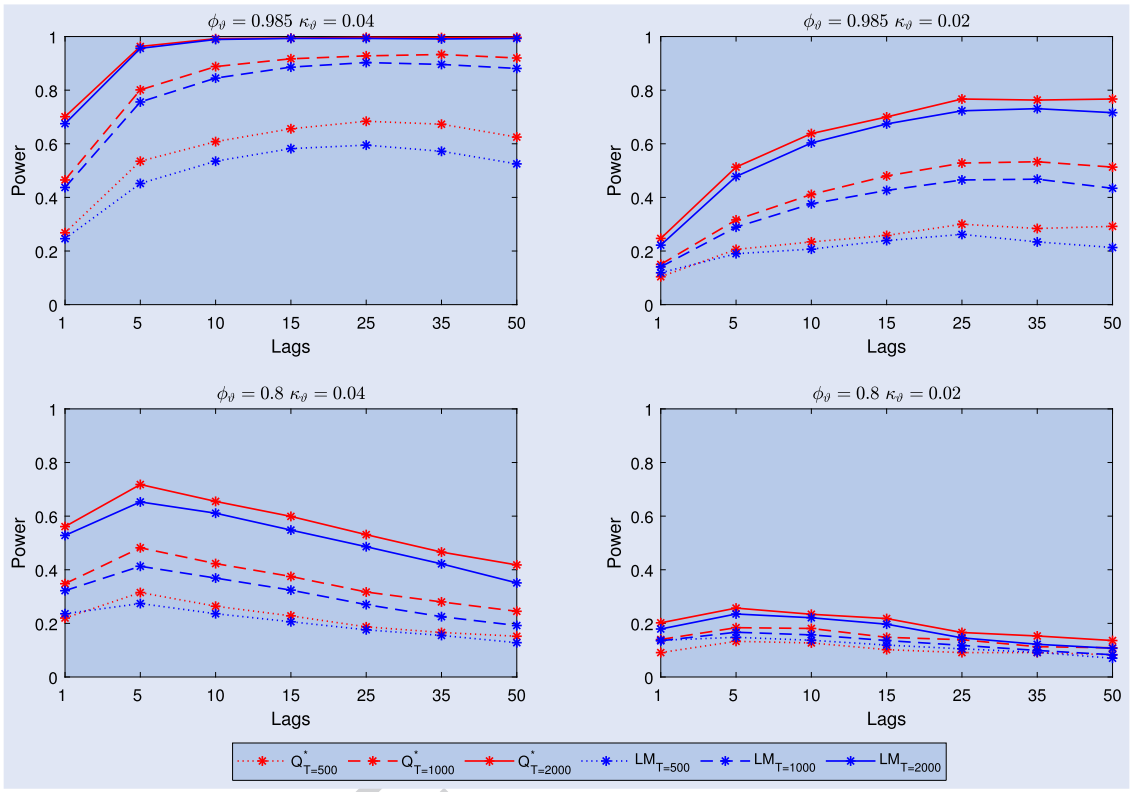
In any case, we present in figure 5 the results of the empirical size of the full LM test and of the simple Q^* test with $T = 2000$ computed using the standard χ^2_P critical values. From these is possible to notice that, in particular for lower values of the tail index parameter $\eta = \exp(\omega_\vartheta)$, the empirical size of the tests assuming a χ^2_P distribution is close to the nominal significance levels even at small sample sizes. When we increase the value of the tail index parameter the simple Q^* test tends to over reject slightly at higher lags with respect to the full LM test.

The increase in empirical size as the number of lags P in Box-Pierce type tests is a known problem in the literature, see Kan and Wang (2010), and this discrepancy is only partly improved by correcting the size with the implementation of the block bootstrap procedure‡ . A valid alternative solution to implement can be the correction suggested by Dufour and Roy (1985), but given that the size issue in our case is not too severe, we leave it as a viable implementation for future research.

REMARK 3.2 These results indicate that, for the data-generating processes considered in our simulations, the weighted chi-square distribution governing the null behaviour of the simple Q^* test is numerically close to the corresponding χ^2_P distribution. In other words, the eigenvalues λ_j appearing in the weighted chi-square representation are sufficiently close to one for the purpose of inference at conventional significance levels. As a consequence, while block bootstrap procedures remain available as a theoretically robust alternative, they are not strictly necessary in our empirical implementation. In particular, when the tail index takes relatively low values, standard chi-square critical values provide an accurate

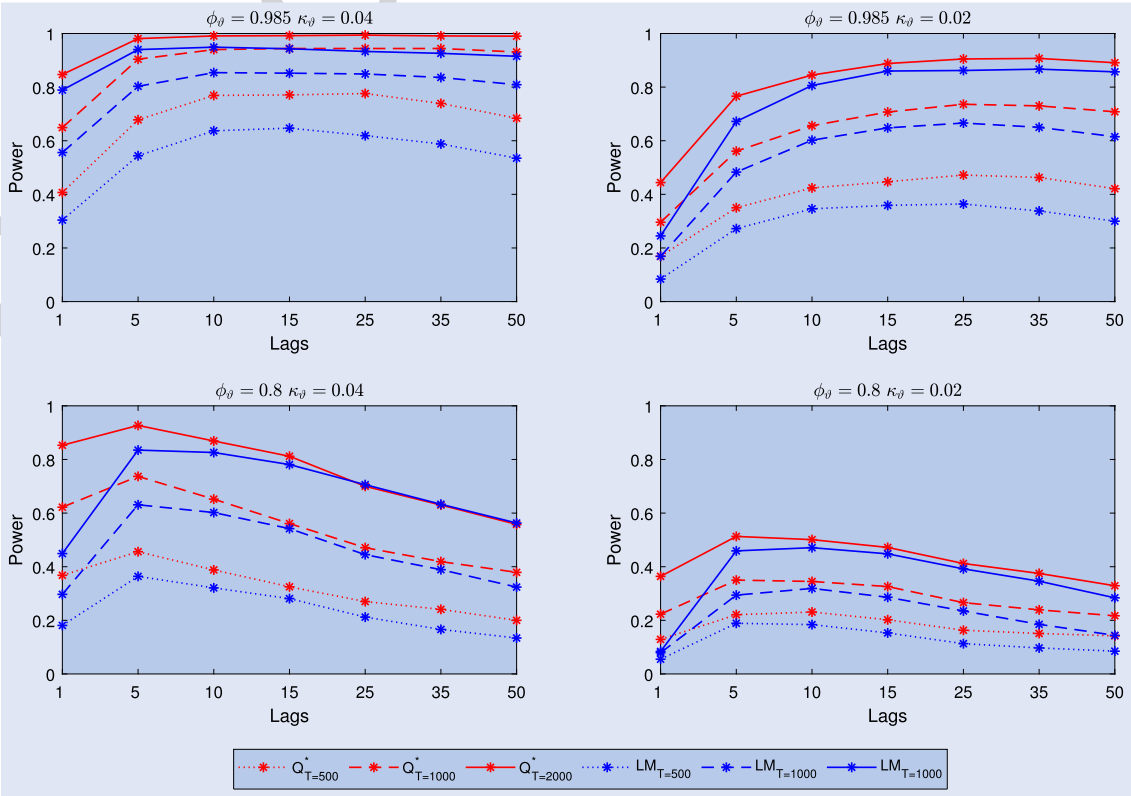
† See, for example, Harvey and Streibel (1998) and Harvey and Thiele (2016).

‡ In additional simulations, not reported for brevity, we verified that the empirical size differences are robust to alternative choices of block size (e.g. $\ell = \lceil T^{1/3} \rceil$ and $\ell = \lceil 3T^{1/3} \rceil$).



Colour online, B/W in print

Figure 3. Plot of the empirical Power of the *full* LM test and the *simple* Q^* test for different lags, obtained from $N = 1000$ simulations of the Dynamic Scale-Tail model with sample size $T = 500$, 1000, 2000 for $\exp(\omega_\vartheta) = 2$.



Colour online, B/W in print

Figure 4. Plot of the empirical Power of the *full* LM test and the *simple* Q^* test for different lags, obtained from $N = 1000$ simulations of the Dynamic Scale-Tail model with sample size $T = 500$, 1000, 2000 for $\exp(\omega_\vartheta) = 8$.

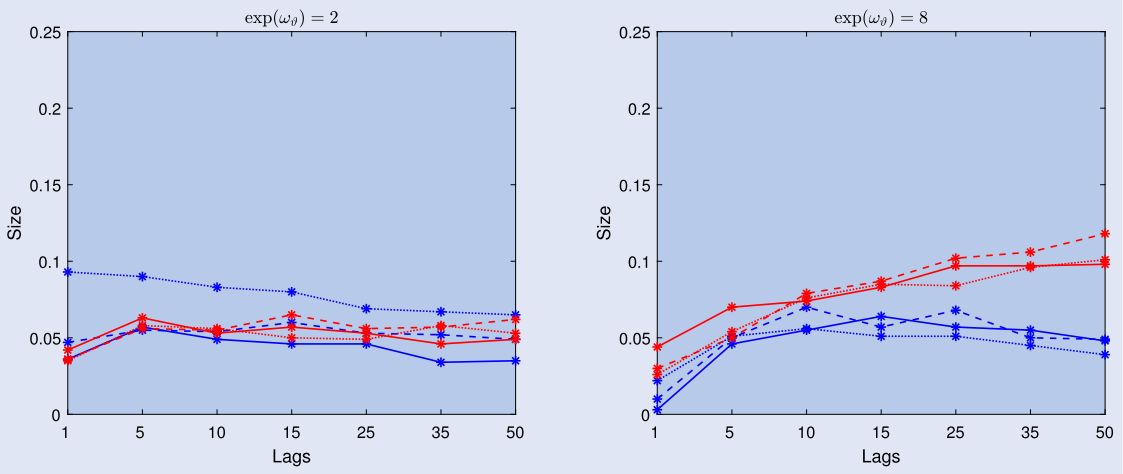


Figure 5. Plot of the empirical Size of the *full* LM test and the *simple* Q^* test for different lags, obtained from $N = 1000$ simulations of the Dynamic Scale-Tail model with sample size $T = 500$, 1000, 2000.

and computationally efficient approximation for the proposed tests.

In figures 3–5 we can see that the power of the test tends to decrease as the sample size T decreases. Since ω_ϑ is the unconditional mean of $\vartheta_{t|t-1}$ we can see that, while the average tail index decreases, the two tests tend to perform quite similarly, as the tail index increases the difference between the two tests increases. In particular, the full *LM* test tends to be slightly more conservative than its *simple* counterpart. This result can be explained by the fact that as the degrees of freedom increases the score of the likelihood with respect to ϑ tends to flatten. Therefore, as η_0 approaches 10 it is difficult to estimate the exact values of η which would maximise the likelihood. This uncertainty drives certainly the difference in sizes since the actual distribution of the simple Q^* test depends mostly on the re-estimation of the tail index parameter in the secondary stage, while the full *LM* test relies on the ML estimation of the parameters and scores under the restricted model. This issue can also be observed in the results in tables 1–3 where it is possible to see that as ω_ϑ increases, the standard errors of the estimates of the parameters governing the dynamics of $\vartheta_{t|t-1}$ also increase, making the estimates less accurate. In general, we can see that the dynaic model is quite reliable in estimating the correct dynamics of $\eta_{t|t-1}$ when $\omega_\vartheta \leq \log 15$.

In addition, We can also notice that fitting the dynamic tail also improves the fit of the scale. This can be seen from the results of the Box–Ljung $Q^*_{u^\dagger}(P)$ test performed on the fitted score with respect to the scale parameter, $\hat{u}^\lambda_t$, which tends to fall when fitting the joint model, in particular for true values of $\omega_\vartheta \leq \log 8$.

Overall, While ω_ϑ is low, both tests tend to lose power quickly as the persistency of the process falls, i.e. when ϕ_ϑ and κ_ϑ fall. On the other hand, when ω_ϑ is high, even for low values of the tail index, both tests tend to retain power at detecting dynamics.

Given these results, we can assert that, in detecting dynamics in the tail index parameter, in terms of power, the simple Q^* performs similarly, and in some cases also better, than the full LM test and remains a viable simpler alternative to implement.

3.3. Test comparison

In this section, we compare the performance of the full LM test and the simple LM test with the GAS-LM test developed by Calvori *et al.* (2017), another test designed for dynamic parameters in the score-driven literature. The GAS-LM test is also based on fitted scores with respect to the dynamic parameter being tested, under the null hypothesis of static dynamics, as well as the scores with respect to the other static parameters of the model estimated under the null hypothesis. Calvori *et al.* (2017) demonstrate that the test generally performs well, particularly in the presence of strong unobserved mean-reverting dynamics, and that it has significantly higher power than other competitors, such as those developed by Andrews (1993) and Muller and Petalas (2010).

We compare the empirical power of the GAS-LM test against the power of the full LM and the simple Q^* test, assuming that all three tests are performed with $P = 1$, $P = 20$ and P^* selected using the automatic algorithm of Escanciano and Lobato (2009). In addition, we also include the Nyblom alternative in the comparison as a benchmark, as it is often regarded as a general test for parameter instability.

To make this comparison, we performed a series of simulations of the same model used in 3.2, with the same dynamic specification for the parameter $\lambda_{t|t-1}$. For the tail-index dynamics, we consider two values of the unconditional mean, $\omega_\vartheta \in \{\log 2, \log 8\}$, and parameterise the other coefficients of the governing the dynamics as

$$\kappa_\vartheta = \sqrt{c/50}, \quad \phi_\vartheta = \sqrt{1 - c/50}, \quad c \in \{0, 1, \ldots, 25\}.$$

In this way, as c varies, we can describe models for the tail index with various levels of persistency from high to low, including a static static tail index model for $c = 0$. In the same way, we can assess the performance of the tests as the impact of the score dynamics, represented by κ_ϑ, increases while maintaining the persistence of the process. For each specification, we performed $N = 1000$ simulations for both $T = 500$ and $T = 1000$. In these simulations all the tests tend to maintain the nominal size level, a part from the GAS-LM. This may be due to the fact that the test, although theoretically an *LM* test, numerically evaluates the scores and is impacted by

Table 1. Estimation results and residual correlation of the fitted scores $\hat{u}_t$ after fitting to the data a Beta-t-EGARCH and a Dynamic Scale and Tail DCS model on 1000 simulations of length $T = 2000$ generated by a conditional t distribution with dynamic scale and tail with different ω_ϑ assumptions.

	$\omega_\vartheta = \log 2, \phi_\vartheta = 0.99, \kappa_\vartheta = 0.25$				$\omega_\vartheta = \log 8, \phi_\vartheta = 0.99, \kappa_\vartheta = 0.01$				$\omega_\vartheta = \log 15, \phi_\vartheta = 0.99, \kappa_\vartheta = 0.005$				$\omega_\vartheta = \log 30, \phi_\vartheta = 0.98, \kappa_\vartheta = 0.005$			
	Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail		
$T = 2000$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$
$Q(1)$	4633	1017	0.760	0.021	2901	3182	0.972	0.033	7684	16476	2588	0.133	12895	9559	2875	0.076
	(0.031)	(0.313)	(0.383)		(0.089)	(0.074)	(0.324)		(0.006)	(0.000)	(0.108)		(0.000)	(0.002)	(0.090)	
$Q(5)$	20204	4879	3978	0,031	11947	7664	4054	0,048	21428	31076	10623	0.201	32203	18083	12047	0.116
	(0.462)	(0.961)	(0.979)		(0.715)	(0.672)	(0.965)		(0.175)	(0.006)	(0.763)		(0.024)	(0.089)	(0.719)	
$Q(10)$	38057	9924	8122	0.027	22697	12907	7896	0.061	29593	42149	18570	0.241	43636	23515	22063	0.143
	(0.000)	(0.447)	(0.617)		(0.012)	(0.229)	(0.639)		(0.001)	(0.000)	(0.046)		(0.000)	(0.009)	(0.015)	
$Q(15)$	53361	15204	12331	0.029	32464	18058	11934	0.062	36102	53139	25901	0.259	50814	28798	31550	0.149
	(0.000)	(0.437)	(0.654)		(0.006)	(0.260)	(0.684)		(0.002)	(0.000)	(0.039)		(0.000)	(0.017)	(0.007)	
$Q(25)$	78734	25757	20858	0.028	48790	28550	20291	0.07	48775	75217	40019	0.269	63117	39127	48588	0.149
	(0.000)	(0.421)	(0.701)		(0.003)	(0.283)	(0.731)		(0.003)	(0.000)	(0.029)		(0.000)	(0.036)	(0.003)	
$Q(35)$	100182	36224	29937	0.025	62381	39088	29502	0.083	60.259	96736	52598	0.264	74225	49320	65097	0.154
	(0.000)	(0.411)	(0.711)		(0.003)	(0.291)	(0.730)		(0.005)	(0.000)	(0.028)		(0.000)	(0.055)	(0.001)	
$Q(50)$	128365	51868	43896	0.033	81934	55127	42569	0.081	77820	129303	70.945	0.274	89406	64611	87234	0.156
	(0.000)	(0.401)	(0.716)		(0.003)	(0.287)	(0.763)		(0.007)	(0.000)	(0.027)		(0.001)	(0.080)	(0.001)	
$\hat{\eta}$	1903	–			8795	–			20756	–			59181	–		
	(0.286)				(15198)				(33762)				(88537)			
$\exp(\hat{\omega}_\vartheta)$	–	2048			–	8828			–	19858			–	37361		
		(0.179)				(0.355)				(0.596)				(0.638)		
$\hat{\phi}_\vartheta$	–	0.984			–	0.921			–	0.934			–	0.919		
		(0.030)				(0.262)				(0.216)				(0.208)		
$\hat{\kappa}_\vartheta$	–	0.025			–	0.011			–	0.003			–	0.003		
		(0.008)				(0.006)				(0.008)				(0.006)		

Table 2. Estimation results and residual correlation of the fitted scores $\hat{u}_t$ after fitting to the data a Beta-t-EGARCH and a Dynamic Scale and Tail DCS model on 1000 simulations of length $T = 1000$ generated by a conditional t distribution with dynamic scale and tail with different ω_ϑ assumptions.

	$\omega_\vartheta = \log 2\, \phi_\vartheta = 0.99\, \kappa_\vartheta = 0.25$				$\omega_\vartheta = \log 8\, \phi_\vartheta = 0.99\, \kappa_\vartheta = 0.01$				$\omega_\vartheta = \log 15\, \phi_\vartheta = 0.99\, \kappa_\vartheta = 0.005$				$\omega_\vartheta = \log 30\, \phi_\vartheta = 0.98\, \kappa_\vartheta = 0.005$			
	Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail		
$T = 1000$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$
$Q(1)$	2799	1051	0.755	0.021	2205	0.983	1748	0.076	2516	3463	2192	0.15	5742	4547	1329	0.08
	(0.094)	(0.305)	(0.385)		(0.138)	(0.322)	(0.186)		(0.113)	(0.063)	(0.139)		(0.017)	(0.033)	(0.249)	
$Q(5)$	14551	5051	3828	0.033	9412	5150	6868	0.112	8541	8232	7955	0.222	18834	12075	5378	0.129
	(0.731)	(0.958)	(0.980)		(0.820)	(0.964)	(0.883)		(0.774)	(0.629)	(0.822)		(0.332)	(0.474)	(0.932)	
$Q(10)$	27307	10.259	8099	0.038	18034	10.445	12705	0.14	14743	13141	13693	0.241	27362	20471	9725	0.141
	(0.002)	(0.418)	(0.619)		(0.054)	(0.402)	(0.241)		(0.142)	(0.216)	(0.187)		(0.002)	(0.025)	(0.465)	
$Q(15)$	38637	15447	12313	0.036	25247	15582	18159	0.148	20.468	18316	19861	0.264	33299	28426	13964	0.151
	(0.001)	(0.420)	(0.655)		(0.047)	(0.410)	(0.254)		(0.155)	(0.246)	(0.177)		(0.004)	(0.019)	(0.528)	
$Q(25)$	58617	25626	21154	0.037	39688	25834	28506	0.162	31982	28596	30.165	0.266	43907	44002	22605	0.171
	(0.000)	(0.428)	(0.684)		(0.031)	(0.417)	(0.285)		(0.159)	(0.281)	(0.218)		(0.011)	(0.011)	(0.601)	
$Q(35)$	75253	36019	30.291	0.039	52428	36327	38080	0.172	43278	38884	39999	0.272	54043	59672	31011	0.173
	(0.000)	(0.421)	(0.695)		(0.029)	(0.407)	(0.331)		(0.159)	(0.299)	(0.258)		(0.021)	(0.006)	(0.661)	
$Q(50)$	97089	51714	43777	0.05	69943	51655	52148	0.159	59242	54354	54245	0.254	68553	83185	43558	0.172
	(0.000)	(0.407)	(0.720)		(0.033)	(0.409)	(0.390)		(0.174)	(0.312)	(0.316)		(0.042)	(0.002)	(0.728)	
$\hat{\eta}$	1938	–			10.323	–			35463	–			94880	–		
	(0.495)				(20.535)				(70.487)				(126017)			
$\exp(\hat{\omega}_\vartheta)$	–	2120			–	10421			–	24091			–	43993		
		(0.340)				(0.654)				(0.897)				(0.855)		
$\hat{\phi}_\vartheta$	–	0.974			–	0.947			–	0.923			–	0.913		
		(0.077)				(0.175)				(0.232)				(0.224)		
$\hat{\kappa}_\vartheta$	–	0.024			–	0.007			–	0.000			–	0.002		
		(0.013)				(0.014)				(0.011)				(0.008)		

Table 3. Estimation results and residual correlation of the fitted scores $\hat{u}_t$ after fitting to the data a Beta-t-EGARCH and a Dynamic Scale and Tail DCS model on 1000 simulations of length $T = 500$ generated by a conditional t distribution with dynamic scale and tail with different ω_ϑ assumptions.

	$\omega_\vartheta = \log 2\, \phi_\vartheta = 0.99\, \kappa_\vartheta = 0.25$				$\omega_\vartheta = \log 8\, \phi_\vartheta = 0.99\, \kappa_\vartheta = 0.01$				$\omega_\vartheta = \log 15\, \phi_\vartheta = 0.99\, \kappa_\vartheta = 0.005$				$\omega_\vartheta = \log 30\, \phi_\vartheta = 0.98\, \kappa_\vartheta = 0.005$			
	Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail			Beta-t-EGARCH	Dynamic Scale and Tail		
$T = 500$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% Rej $r_{u\vartheta}(P)$
$Q(1)$	1699	1097	0.833	0.02	1149	0.846	3484	0.109	1541	2096	7235	0.17	2472	3817	2568	0.12
	(0.192)	(0.295)	(0.361)		(0.284)	(0.358)	(0.062)		(0.214)	(0.148)	(0.007)		(0.116)	(0.051)	(0.109)	
$Q(5)$	9610	5080	4223	0.051	5850	4955	12588	0.155	5791	7518	28958	0.23	9400	12876	9861	0.18
	(0.889)	(0.954)	(0.975)		(0.950)	(0.974)	(0.626)		(0.908)	(0.836)	(0.204)		(0.781)	(0.576)	(0.766)	
$Q(10)$	18330	10.190	8492	0.061	12315	10.393	21432	0.179	10.786	14327	49986	0.241	15491	24211	17460	0.18
	(0.050)	(0.424)	(0.581)		(0.265)	(0.407)	(0.018)		(0.374)	(0.159)	(0.000)		(0.115)	(0.007)	(0.065)	
$Q(15)$	26038	15292	12739	0.056	17779	15637	28435	0.195	16022	21417	67902	0.249	20.140	34947	24059	0.173
	(0.038)	(0.431)	(0.622)		(0.274)	(0.407)	(0.019)		(0.381)	(0.124)	(0.000)		(0.167)	(0.003)	(0.064)	
$Q(25)$	40.342	25508	21927	0.067	28976	26176	40.786	0.193	26207	34763	95830	0.241	30.107	55668	37284	0.185
	(0.027)	(0.434)	(0.640)		(0.265)	(0.398)	(0.024)		(0.397)	(0.093)	(0.000)		(0.220)	(0.000)	(0.054)	
$Q(35)$	53253	35617	30.991	0.064	39467	36782	51242	0.178	36229	47508	117822	0.237	39457	75138	48140	0.168
	(0.025)	(0.439)	(0.662)		(0.277)	(0.386)	(0.038)		(0.411)	(0.077)	(0.000)		(0.277)	(0.000)	(0.069)	
$Q(50)$	70.635	51067	44514	0.066	54049	52250	65948	0.169	50.419	65881	144210	0.225	53312	101293	63758	0.164
	(0.029)	(0.432)	(0.692)		(0.322)	(0.387)	(0.065)		(0.457)	(0.065)	(0.000)		(0.348)	(0.000)	(0.091)	
$\hat{\eta}$	2077	–			25047	–			63319	–			124676	–		
	(0.746)				(64449)				(106959)				(143219)			
$\exp(\hat{\omega}_\vartheta)$	–	2275			–	14335			–	31182			–	49877		
		(0.464)				(1027)				(1077)				(0.944)		
$\hat{\phi}_\vartheta$	–	0.947			–	0.928			–	0.916			–	0.907		
		(0.161)				(0.224)				(0.258)				(0.211)		
$\hat{\kappa}_\vartheta$	–	0.018			–	0.000			–	-0.002			–	0.000		
		(0.031)				(0.022)				(0.016)				(0.016)		

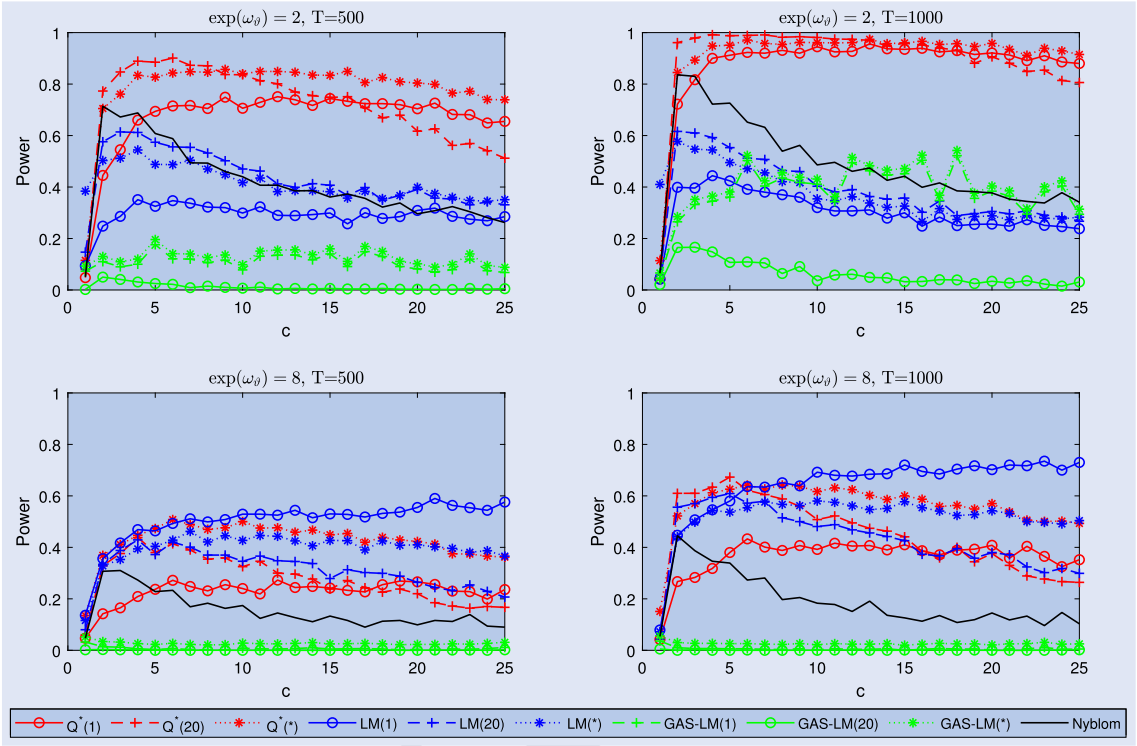


Figure 6. Size adjusted power comparison under simulation of the *full* LM test, the *simple Q* test, the GAS-LM test of Calvori *et al.* (2017) and the Nyblom test for 1 lag, 20 lags and a number of lags determined by the criterion of Escanciano and Lobato (2009). The comparisons are performed under different assumptions of persistency in the true dynamics of the tail index parameter $\vartheta_{t|t-1}$ as well as assuming an average tail index value of either $\exp(\omega_\vartheta) = 2$ or $\exp(\omega_\vartheta) = 8$ and time series lengths of both $T = 500$ and $T = 1000$.

changes in all model parameters at the same time. In particular, due to their close relationship, it could struggle to numerically disentangle the changes in the scores with respect to the scale parameter λ from the changes in the scores with respect to the tail parameter ϑ. For these reasons, instead of evaluating the test directly from the model under the null, we have evaluated the test on the scores of a static t distribution fitted on the $\hat{\varepsilon}_t = y_t e^{-\hat{\lambda}_{t|t-1}}$ evaluated from standardising the observation from the previously fitted scale from a Beta-t-EGARCH model, as we have done for the simple Q^* test. This has improved the empirical size level significantly. However, there were still size distortions reported between 0.1 and 0.2. For these reasons, in our comparisons, all reported power levels are size adjusted.

The results of the tests shown in figure 6 reveal that the performances of the full *LM* and simple Q^* tests tend to diverge as c decreases. Both tests tend to have the highest power with low values of c, and they lose it slightly as it increases. In particular, in the case of low tail index the simple test is significantly more powerful for higher c values, even in the case of a small sample size T. For a higher tail index, the two tests are more comparable in size, even if the most powerful test is the full *LM* with $P = 1$ also for higher values of c.

The GAS-LM(1) test tends to fail to capture the presence of a dynamic tail index parameter across almost all parameter specifications, apart from high values of c for $\omega_\vartheta = \log 2$. The GAS-LM(*) becomes more competitive for low values of the tail index, becoming rapidly comparable in power to the full *LM* test as c decreases. None of the GAS-LM tests have power for an $\omega_\vartheta = \log 8$. The Nyblom test has a similar behaviour to the simple and full test, rapidly gaining power for low c and gradually losing it for higher c values. It can be seen that it has a comparable and sometimes even higher power than the full *LM* test for low values of the tail index. On the other hand, it has lower power than both the simple and full tests for higher values of the tail index parameter.

In summary, we can conclude that the GAS-LM(*) test is less suitable at evaluating time varying tails in a scale-tail score-driven model. However, the full LM and simple Q^* are viable alternatives, with a preference overall for the simple test, in particular in cases of a low tail index parameter.

3.3.1. Test comparison: asymmetric tails. In the previous section, we derived the exact formulation of the LM test to detect dynamics in the tail index of the distribution when the tails are symmetric. In the case of the asymmetric tail extension introduced in Section 6, the derivation becomes much more complicated, particularly in deriving the matrix $\boldsymbol{\Psi}_{\vartheta\vartheta}$. It is straightforward to verify that, in this case, it would also be composed of an alteration of the simple Q^* test with a correction, as in equation (14). Given our findings on the performance of the simple Q^* in the symmetric case, we are inclined to believe that it could also be appropriate in the asymmetric case.

REMARK 3.3 When modelling the individual tails of the Asymmetric t distribution, the score with respect to the dynamic tail index parameter of each tail depends on the data only when the corresponding observation falls within that tail. Although the scores may exhibit some contemporaneous dependence through the structure of the joint distribution,

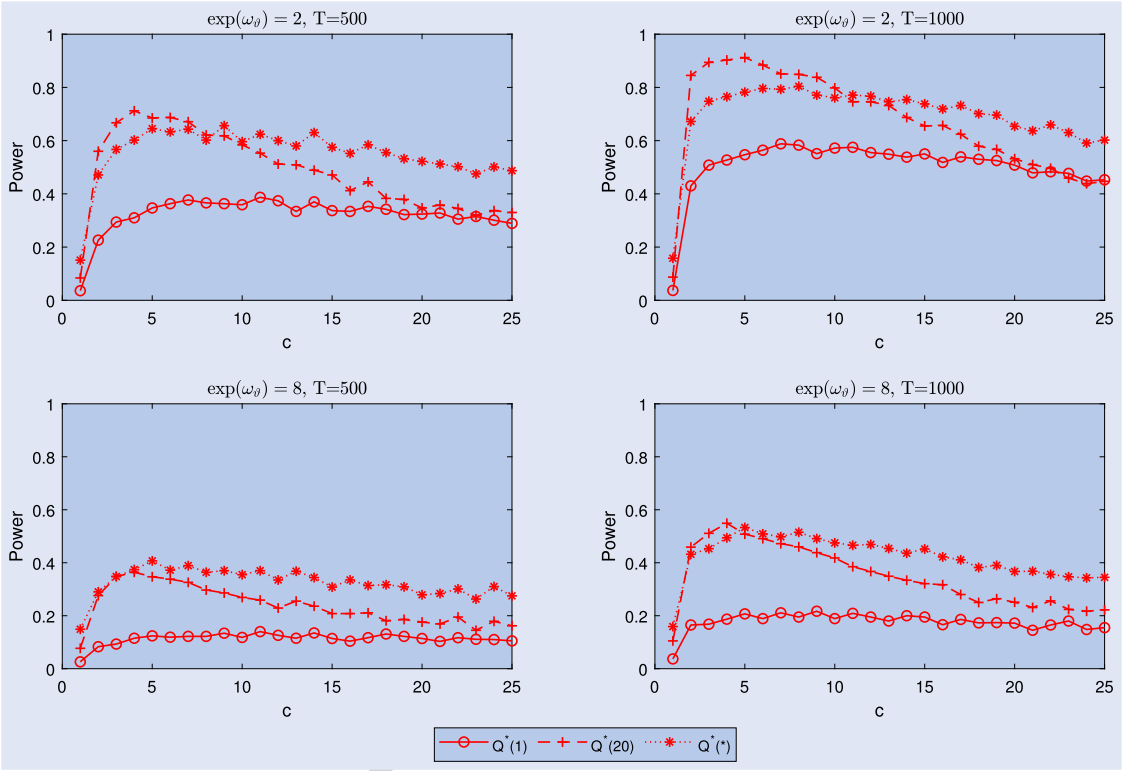


Figure 7. Size adjusted power comparison Upper tail When only upper tail is dynamic under simulation of the *simple* Q test for 1 lag, 20 lags and a number of lags determined by the criterion of Escanciano and Lobato (2009). The comparisons are performed under different assumptions of persistency in the true dynamics of the tail index parameter $\vartheta_{t|t-1}$ as well as assuming an average tail index value of either $\vartheta = 2$ or $\vartheta = 8$ and time series lengths of both $T = 500$ and $T = 1000$.

their informational content remains tail-specific. As a consequence, the simple Q^* test applied separately to each tail index parameter effectively relies on a reduced subset of observations, which can lead to lower statistical power compared to a test applied on a symmetric tail index parameter, which exploits the full sample.

We confirmed the assertion in this remark by comparing the power of the *simple* Q^* test applied to a single tail, both when only that tail and when both tails are dynamic. The results, shown in figures 7 and 8 for the upper tail (with similar results for the lower tail), indicate that when the tested tail is the only dynamic tail, the power of the simple Q^* test behave similarly to the case of a symmetric dynamic tail but with slightly lower power. When the average degrees of freedom are lower, its power becomes significantly larger, as seen in the symmetric tails case. However, as the average degrees of freedom increases, the power of the simple Q^* decreases significantly.

Given these results, we still consider the simple Q^* test a feasible option for testing the presence of dynamics in tail parameters, even in the asymmetric case.

4. Empirical results

To investigate the effectiveness of the new dynamic tail model across different types of data series, we considered returns from Equity Indexes and Credit Default Swaps (CDS), both known for their extreme fluctuations over time.

For the equity indexes, we analysed the daily log returns of the Dow Jones, Nikkei, and Hang Seng indexes. The data were collected from Yahoo Finance, covering the period from 1 February 1987 to 29 April 2016. For the CDS, we examined the daily log returns of the 5-year Italy, 5-year Portugal, and 10-year France CDS sovereign debt rates. These data were sourced from Bloomberg, spanning from 1 August 2007 to 28 February 2018.

From table 4, we can observe that all six series exhibit high sample kurtosis, with the Equity Indexes showing higher kurtosis than the CDS series. Two of the CDS series are right − skewed, and their sample standard deviation is four times higher than that of the Equity Indexes, which are all left-skewed. In all cases, there are signs of residual correlation at lag 20 in the squared returns, as well as at higher lags, indicating features of long memory, as is common with financial time series.

To estimate the Dynamic Scale-Tail score-driven model, we first fitted both series with a beta-t-EGARCH score-driven model, assuming both conditional symmetric and asymmetric t distributions. Then, using the fitted residuals, we computed the scores under the null hypothesis, $\hat{u}t^{\vartheta\dagger}$, as defined in equation (15), and performed the simple Q^* test.† Where appropriate, we then fitted the general Dynamic Scale-Tail

† Another reason for preferring the simple Box-Ljung version of the test is that, aside from its simplicity and effectiveness, it allows for direct comparison with the Box-Ljung test performed on the fitted scores to detect residual correlation after fitting the Dynamic Tail model.

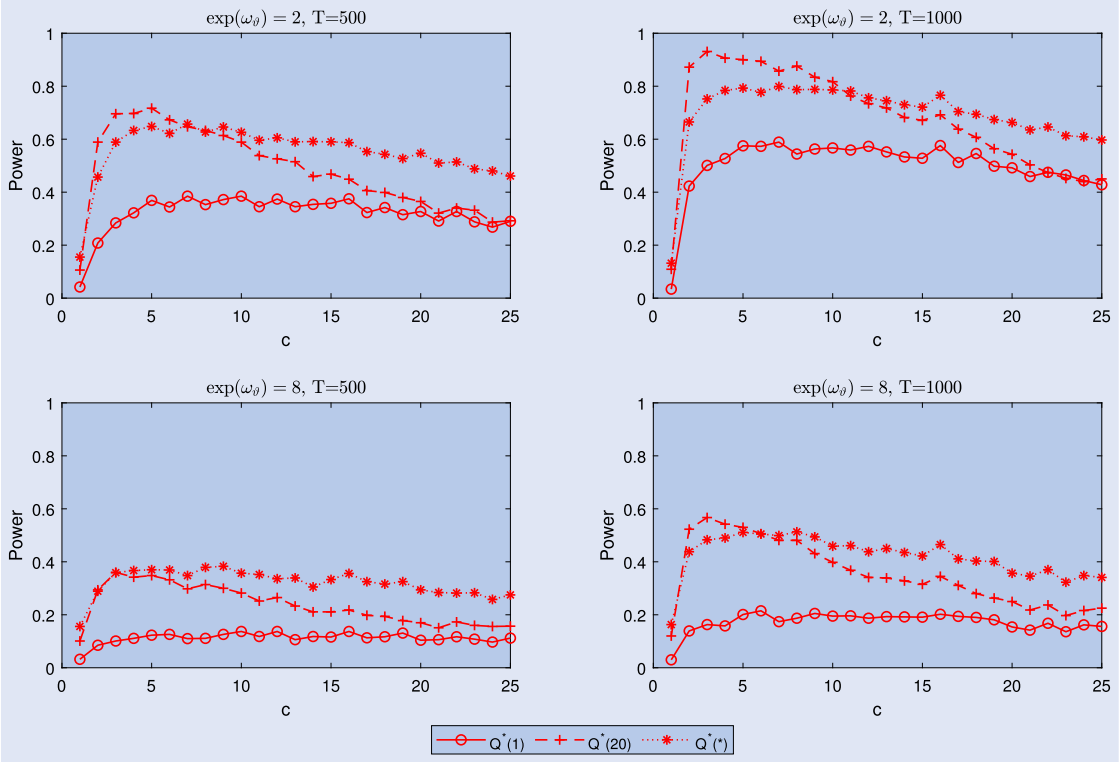


Figure 8. Power comparison Upper tail When both tail are dynamic under simulation of the *simple Q* test for 1 lag, 20 lags and a number of lags determined by the criterion of Escanciano and Lobato (2009). The comparisons are performed under different assumptions of persistency in the true dynamics of the tail index parameter $\vartheta_{t|t-1}$ as well as assuming an average tail index value of either $\vartheta = 2$ or $\vartheta = 8$ and time series lengths of both $T = 500$ and $T = 1000$.

Table 4. Descriptive statistics.

	N	Mean	St. Dev.	Skewness	Kurtosis	Min.	Max.	$Q_{y_t^2}(1)$	$Q_{y_t^2}(20)$
Dow Jones Index	7293	0.000	0.015	− 0.246	10.330	− 0.161	0.132	547.850	3544.937
Nikkei Index	7293	0.000	0.011	− 1.693	44.696	− 0.256	0.105	80.671	773.295
Han Seng Index	7293	0.000	0.017	− 2.349	60.262	− 0.405	0.172	135.572	279.996
5y Italy CDS	2761	0.001	0.044	0.262	18.409	− 0.437	0.429	79.434	429.104
5y Portugal CDS	2761	0.001	0.042	− 0.219	18.913	− 0.543	0.280	24.722	621.996
10y France CDS	2761	0.001	0.043	0.260	35.979	− 0.589	0.542	398.766	464.718

score-driven model.‡ Additionally, we included the results from fitting GARCH and GARCH-*t* models in the comparison. All estimations were performed using maximum likelihood.§

4.1. Modelling equity index dataseries

In fitting the Beta-*t*-EGARCH model to both the Dow Jones Index and the Italy 5Y CDS returns series, we assumed a two-component dynamics for $\lambda_{t|t-1}$, as described in Harvey (2013, pp. 91–92), to capture the long-memory feature of returns volatility.† This approach helps to remove any residual correlation in the fitted scores with respect to λ, which could otherwise affect the accuracy of estimating the tail index parameter dynamics through the scores for the tail index parameter, u_t^{ϑ}. We retain this specification for the scale in the subsequent dynamic scale-tail models we fit to the series.

Given the length of the series, the question arises whether we should account for possible leverage effects. The challenge in doing so is that introducing an asymmetric response to negative returns in the scale dynamics could influence the

‡ When fitting a dynamic tail index parameter, the score with respect to the scale parameter $\lambda_{t|t-1}$ should also be standardised by its static information quantity $\mathcal{I}_{\lambda\lambda}$, as this would also be time-varying.

§ Since estimating the general Dynamic Scale-Tail model is complex, we improved the accuracy of parameter estimates by first fitting a Dynamic Tail score-driven model to the standardised data, assuming the tail index parameter to be dynamic while keeping the scale constant and set to 1. We then used the estimated parameters, in combination with the parameter estimates from the Beta-*t*-EGARCH score-driven model, as starting values for the full Dynamic Scale-Tail score-driven model.

† Both the full and simple tests assume that the Beta-*t*-EGARCH scale specification under the null has only one single component, as described in Proposition 3.1. In principle, if the scale specification would be more complex the derivation of the statistics should be modified accordingly. Nevertheless, as a simple detection device of dynamics in the tail parameter which assumes a first-order specification in the scale is enough. This because the resulting test would be more conservative than when exploiting the full extend of the dependency between scale and tail taking into account of additional parameters that drive the dynamics of the scale.

behaviour of our dynamic asymmetric tail model, as outlined in Harvey (2013). For this reason, we present the results without and with leverage effects in tables 5 and 6, respectively,‡ where the leverage effect is added to the dynamics of the components of $\lambda_{t|t-1}$ as follows:

$$\lambda_{t|t-1} = \omega_\lambda + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}$$
$$\lambda_{i,t+1|t} = \phi_{i,\lambda}\lambda_{i,t|t-1} + \kappa_{i,\lambda}u_t^\lambda + \kappa^*_{i,\lambda}\text{sgn}(-y_t)(u_t^\lambda + 1)$$
$$i = 1, 2$$

In table 5, we see that without leverage, the simple Q^* test rejects the null of static degrees of freedom in both the symmetric and asymmetric cases. However, in the asymmetric case, the Dow Jones and Hang Seng indexes only show dynamic behaviour in the lower tail parameter η_1, suggesting a dynamic lower tail and a static upper tail. There are weak rejections for the upper tail in the case of the Nikkei. This result aligns with the findings of Mazur and Pipień (2018), who identified the left tail of returns as more variable and consistently heavier than the right tail. The Dynamic Scale-Tail model is then fitted accordingly. The estimated ϕ_j parameters for the tail dynamics are not highly persistent, with the negative asymmetric tail showing less persistence than the symmetric tail. The Box-Ljung test results for the fitted scores in table 5 suggest that the model fits the dynamic parameters well, removing all correlation in the simple Q^* test up to lag 50, except for the upper tail of the Nikkei series.† The improved fit from using the Beta-*t*-EGARCH model with dynamic tails is confirmed by the higher likelihood and lower information criteria, with a slight preference for symmetric dynamic tails in the Nikkei and Hang Seng indexes.

As shown in figure 9, the fitted symmetric parameter $\eta_{t|t-1}$ for the Dow Jones Index fluctuates between 1 and 6.5, mostly remaining around 6.5. This result is similar to those found by Blazsek and Monteros (2017), Ayala *et al.* (2017) for the S&P 500, and Massacci (2017) for the shape parameter behaviour of small firms, where a 'floor' around a constant number was observed. The parameter falls below 1 only during the 'Black Friday' market crash of November 1989. Figure 9 also shows the plot of the fitted parameter $\eta_{1t|t-1}$ for the lower tail, which averages around 6, slightly heavier than the symmetric parameter, with nearly identical but slightly more pronounced fluctuations.

To analyse the impact of significant market events on lower tail movements, we plotted the inverse of the symmetric tail parameter $\bar{\eta}_{t|t-1}$ against the fitted time-varying scale $\exp(\lambda_{t|t-1})$ in figure 10. The results indicate that the heaviness of the lower tail in the returns distribution corresponds to many notable market events, although these movements are not necessarily linked to volatility. As expected, there are instances where high volatility coincides with a heavier lower tail, such as during 'Black Monday'. However, the majority of extreme movements in the lower tail occur when volatility changes the least. Examples include the 'Black Friday' crash in November 1989, the November 1991 crash following Congress' vote on credit card rates, and the February 2007 crash at the beginning of the subprime crisis, when Greenspan suggested a possible US recession. These unexpected extreme events drove the market down while volatility, as measured by the Beta-*t*-EGARCH model, did not change significantly. Conversely, events like the Lehman Brothers default were fully accounted for by market volatility, leaving the lower tail heaviness largely unaffected.

Introducing an asymmetric response in the scale parameter through the leverage component in the Beta-*t*-EGARCH model removes all remaining residual correlation in the fitted scores with respect to $\lambda_{t|t-1}$. In figure 6, we observe that, for the Dow Jones and Nikkei indexes in the asymmetric tails case, the simple Q^* test shows residual correlation only in the fitted scores for the upper tail parameter η_2. This is likely because the inclusion of the leverage term allows the scale to capture most extreme negative movements, but neglects some positive movements, which should ideally be modelled separately. All the estimated $\kappa^*_{i,\lambda}$ values are positive, and the leverage impact is mainly confined to the less persistent component of $\lambda_{t|t-1}$, confirming the findings of Harvey and Lange (2018). The symmetric dynamic tail model exhibits similar fitted dynamics and paths for $\eta_{t|t-1}$ compared to the model without leverage. However, the fitted $\eta_{2t|t-1}$ is more persistent and follows a different path from $\eta_{1t|t-1}$ in the case without leverage. The path of its inverse shows fewer extreme movements, partly due to its long-run average of approximately $\exp(\omega_{\eta 2}) = 10.014$, which occurs at different times than $\bar{\eta}_{1t|t-1}$. These periods still coincide with low volatility. Finally, figure 6 shows a significant improvement in likelihood and information criteria for the asymmetric Scale-Tail model with leverage, which captures more residual correlation in the fitted scores for $\lambda_{t|t-1}$, further enhanced by the inclusion of dynamic tail parameters.

These results can be explained by the fact that, as the tail index parameter of the conditional distribution falls, the score with respect to the scale becomes more bounded, preventing extreme scale movements unless they are persistent in the series (see Harvey (2013)). This robustness to outliers in the score with respect to the scale parameter is a renown feature in the score-driven literature. However, when the tails of the conditional distribution are allowed to vary, repeated sudden extreme events tend to shift the tail index rather than the scale, making the tail heavier and allowing for more extreme events. This phenomenon is more apparent in the asymmetric case, where, for example, if the leverage effect on the scale is not accounted for, the lower tail index captures these rapid, non-persistent falls in the series, which are neglected by the scale and detected through residual correlation in the fitted scores for the tail index parameter. In this way, the model effectively distinguishes between scale and tail movements, whether they are occasional or persistent. In the case of Index Returns, tail movements tend to be rarely persistent, suggesting that the dynamic tail model could be even

‡ The estimated parameters for each model are presented in the online appendix.

† Notably, the simple Beta-*t*-EGARCH with two components, either symmetric or asymmetric, fails to entirely remove residual correlation in the fitted scores for the scale parameter $\lambda t \mid t-1$ in the case of the Dow Jones.

Table 5. Goodness of fit and results of the simple $Q_u^*(P)$ test on both the fitted scores with respect to the scale parameter $Q_{u\lambda}(P)$ and with respect to the tail parameter $Q_{u\eta}(P)$ for different lags P. When a model with a static tail parameter $\bar{\eta}_i$ is fitted we report the results of the simple $Q_u^*(P)$ test. If we fit a model with dynamic tails $\hat{\eta}_{it|t-1}$ we report the result of the Box-Ljing test applied to the fitted scores with respect to the tail parameter. In the case of $Q_{u\lambda}(*)$ and $Q_{u\eta}(*)$ the lags P are determined by the algorithm of Escanciano and Lobato (2009). The reported significance levels are ' ****' for 0.01' ***' for 0.05 and ' ****' for 0.1. In case of asymmetric tails we report instead the results for the test on the fitted scores of each tail lower $Q_{u\eta_1}(P)$ and upper $Q_{u\eta_2}(P)$. The models compared for each series are a Beta-t-EGARCH with two components for scale with: (1) a symmetric t with a single static $\hat{\eta}$, (2) a symmetric t with a single dynamic $\eta_{t|t-1}$, (3) an asymmetric t with static $\hat{\eta_1}$ and $\hat{\eta_2}$, (4) an asymmetric t with a dynamic $\eta_{1t|t-1}$ and a static $\hat{\eta_2}$, then: (5) a GARCH, (6) a GARCH-t, all the models do not assume leverage.

	Model	AIC	BIC	Logl	$Q_{u\lambda}(1)$	$Q_{u\lambda}(10)$	$Q_{u\lambda}(25)$	$Q_{u\eta_1}(1)/Q_{u\eta}(1)$	$Q_{u\eta_1}(10)/Q_{u\eta}(10)$	$Q_{u\eta_1}(25)/Q_{u\eta}(25)$	$Q_{u\eta_1}(*)/Q_{u\eta}(*)$	$Q_{u\eta_2}(1)$	$Q_{u\eta_2}(10)$	$Q_{u\eta_2}(25)$	$Q_{u\eta_2}(*)$
Dow Jones Index	(1)	− 48552.32	− 48510.87	24282.16	7.7655****	20.164***	35.073**	16.508****	60.881****	71.797****	81.997****	–	–	–	–
	(2)	− 48570.71	− 48515.44	24293.35	12.138****	27.903****	42.531***	0.0157	11.969	25.026	3.9775	–	–	–	–
	(3)	− 48557.00	− 48508.64	24285.50	7.4963****	19.977***	34.675**	5.1159***	42.096****	49.446****	32.240****	0.8874	11.045	21.082	0.8874
	(4)	− 48565.64	− 48503.47	24291.82	10.371****	23.894****	38.392***	0.3857	10.036	17.003	4.6949**	–	–	–	–
	(5)	− 48084.74	− 48050.20	24047.37	0.0081	8.5582	13.914	–	–	–	–	–	–	–	–
	(6)	− 48549.40	− 48507.95	24280.70	0.9152	10.770	16.725	–	–	–	–	–	–	–	–
Nikkei Index	(1)	− 42339.48	− 42298.17	21175.74	1.5795	10.533	29.119	128.55****	135.64****	148.31****	133.03****	–	–	–	–
	(2)	− 42368.76	− 42313.67	21192.38	6.1647***	15.808	36.420**	0.8621	1.3990	13.179	0.8621	–	–	–	–
	(3)	− 42342.35	− 42294.15	21178.17	1.6149	10.596	29.460	34.970****	41.131****	62.691****	38.572****	0.0529	9.6197	40.744***	61.165****
	(4)	− 42355.39	− 42279.65	21188.70	5.5660***	15.518	35.091**	1.4050	2.4014	40.845***	1.4050	68.781****	255.63****	273.71****	272.58****
	(5)	− 41946.57	− 41912.14	20.978.28	0.9405	2.3402	14.644	–	–	–	–	–	–	–	–
	(6)	− 42324.81	− 42283.50	21168.41	6.5075***	7.8646	17.692	–	–	–	–	–	–	–	–
Han Seng Index	(1)	− 42460.41	− 42419.04	21236.20	0.3648	16.341**	28.473	306.66****	391.01****	400.40****	391.01****	–	–	–	–
	(2)	− 42527.10	− 42471.95	21271.55	3.5481**	21.831***	33.198	0.6217	4.2195	11.776	0.6217	–	–	–	–
	(3)	− 42461.63	− 42413.37	21237.82	0.3353	16.069**	28.159	202.32****	281.07****	288.59****	284.00****	1.0959	9.9888	21.607	1.0959
	(4)	− 42504.32	− 42442.27	21261.16	2.6622	16.620**	28.877	1.2696	3.5164	8.7684	1.2696	–	–	–	–
	(5)	− 41858.93	− 41824.46	20.934.47	123.02****	129.93****	136.41****	–	–	–	–	–	–	–	–
	(6)	− 42426.92	− 42385.55	21219.46	310.93****	320.74****	326.14****	–	–	–	–	–	–	–	–

Table 6. Goodness of fit and results of the simple $Q_u^*(P)$ test on both the fitted scores with respect to the scale parameter $Q_{u\lambda}(P)$, and with respect to the tail parameter $Q_{u\eta}(P)$ for different lags P. When a model with a static tail parameter, $\bar{\eta}_i$, is fitted we report the results of the simple $Q_u^*(P)$ test. If we fit a model with dynamic tails, $\hat{\eta}_{it|t-1}$, we report the result of the Box-Ljing test applied to the fitted scores with respect to the tail parameter. In the case of $Q_{u\lambda}(*)$ and $Q_{u\eta}(*)$ the lags P are determined by the algorithm of Escanciano and Lobato (2009). The reported significance levels are ' **** for 0.01' ***' for 0.05 and ' ****' for 0.1. In case of asymmetric tails we report instead the results for the test on the fitted scores of each tail, lower $Q_{u\eta_1}(P)$ and upper $Q_{u\eta_2}(P)$. The models compared for each series are, a Beta-t-EGARCH with two components for scale with: (1) a symmetric t with a single static $\hat{\eta}$, (2) a symmetric t with a single dynamic $\eta_{t|t-1}$, (3) an asymmetric t with static $\hat{\eta}_1$ and $\hat{\eta}_2$, (8) an asymmetric t with a static $\hat{\eta}_1$ and a dynamic $\eta_{2t|t-1}$, then: then: (5) a GARCH, (6) a GARCH-t, all the models do not assume leverage.

	Model	AIC	BIC	Logl	$Q_{u\lambda}(1)$	$Q_{u\lambda}(10)$	$Q_{u\lambda}(25)$	$Q_{u\eta_1}(1)/Q_{u\eta}(1)$	$Q_{u\eta_1}(10)/Q_{u\eta}(10)$	$Q_{u\eta_1}(25)/Q_{u\eta}(25)$	$Q_{u\eta_1}(*)/Q_{u\eta}(*)$	$Q_{u\eta_2}(1)$	$Q_{u\eta_2}(10)$	$Q_{u\eta_2}(25)$	$Q_{u\eta_2}(*)$
Dow Jones Index	(1)	−48827.89	−48772.62	24421.94	1.1524	6.5813	24.527	9.0555****	24.128****	39.796***	15.059****	–	–	–	–
	(2)	−48847.99	−48778.91	24434.00	0.4927	4.9871	22.250	0.4149	7.2472	28.040	0.4149	–	–	–	–
	(3)	−48838.80	−48776.62	24428.40	1.3760	7.6157	25.230	0.1320	13.721	22.094	3.8908	11.664****	49.854****	64.010****	45.902****
	(8)	−48853.46	−48777.46	24437.73	2.1555	8.1703	25.910	–	–	–	–	0.5513	1.9845	4.2017	0.5513
	(5)	−48353.04	−48304.68	24183.52	0.4964	12.994	19.454	–	–	–	–	–	–	–	–
	(6)	−48731.91	−48676.64	24373.95	0.1322	11.345	18.971	–	–	–	–	–	–	–	–
Nikkei Index	(1)	−42566.17	−42511.08	21291.08	2.0046	10.968	30.511	33.931****	41.196****	53.201****	33.931****	–	–	–	–
	(2)	−42589.03	−42520.18	21304.51	3.7955**	13.894	32.723	1.8589	4.4287	12.951	1.8589	–	–	–	–
	(3)	−42570.86	−42508.89	21294.43	1.9122	11.090	30.632	8.4578****	11.059	25.325	8.4578****	7.2656****	44.600****	61.284****	102.94****
	(8)	−42585.08	−42495.57	21305.54	3.2917**	13.162	33.543	2.0292	3.2806	11.547	2.0292	2.8528**	7.0687	13.502	2.8528**
	(5)	−42224.89	−42176.69	21119.45	0.2715	4.1718	18.809	–	–	–	–	–	–	–	–
	(6)	−42510.24	−42455.15	21263.12	1.8218	4.3488	17.193	–	–	–	–	–	–	–	–
Han Seng Index	(1)	−42617.56	−42562.40	21316.78	1.3486	20.447***	32.529	172.70****	219.77****	228.89****	219.77****	–	–	–	–
	(2)	−42666.31	−42597.37	21343.16	1.7041	21.318***	33.302	0.3409	2.4187	11.551	0.3409	–	–	–	–
	(3)	−42617.69	−42555.64	21317.85	1.3150	20.310***	32.290	167.54****	218.52****	226.88****	222.91****	76.422****	116.45****	125.13****	112.88****
	(8)	−42659.20	−42569.57	21342.60	2.5443	18.468***	31.097	0.3533	1.8418	7.5997	0.3533	0.5687	4.5706	17.868	0.5687
	(5)	−42136.57	−42088.31	21075.28	48.028****	58.489****	62.901****	–	–	–	–	–	–	–	–
	(6)	−42583.08	−42527.92	21299.54	83.560****	93.443****	98.338****	–	–	–	–	–	–	–	–

Colour online, B/W in print

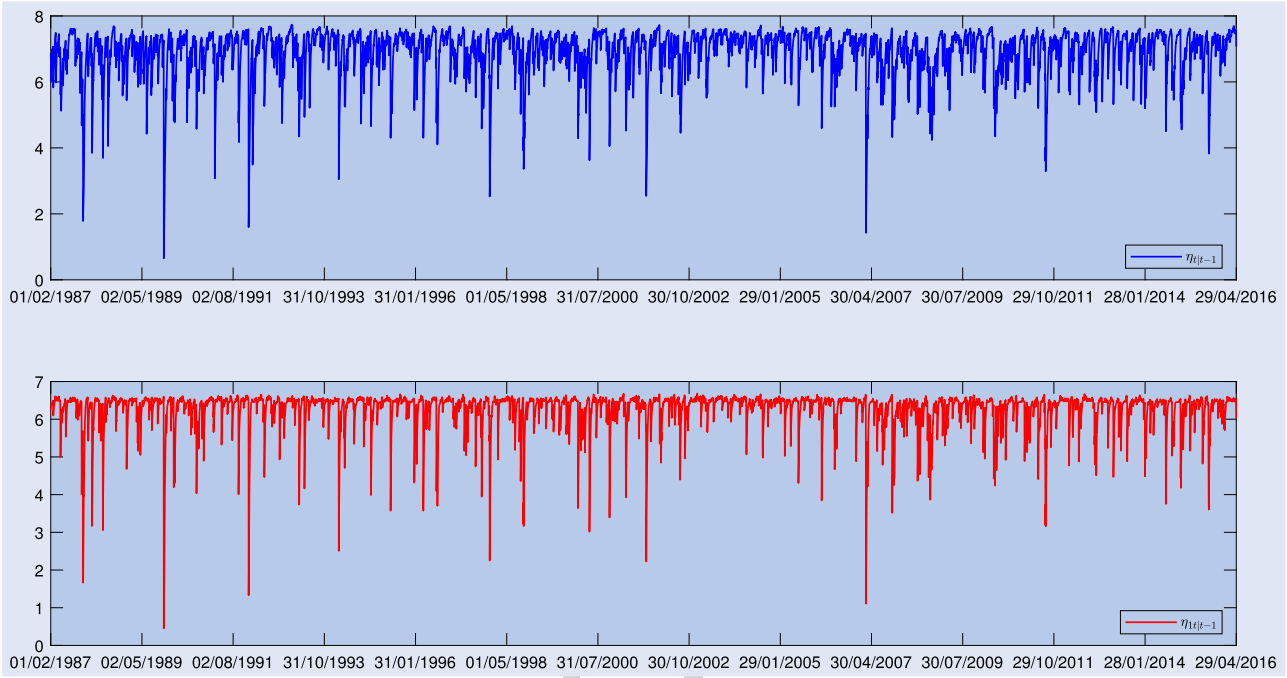


Figure 9. Fitted degrees of freedom of the Dow Jones Index Returns in the symmetric case (Top), $\hat{\eta}_{t|t-1}$, and for the Lower Tail in the asymmetric case (Bottom), $\hat{\eta}_{1t|t-1}$.

Colour online, B/W in print

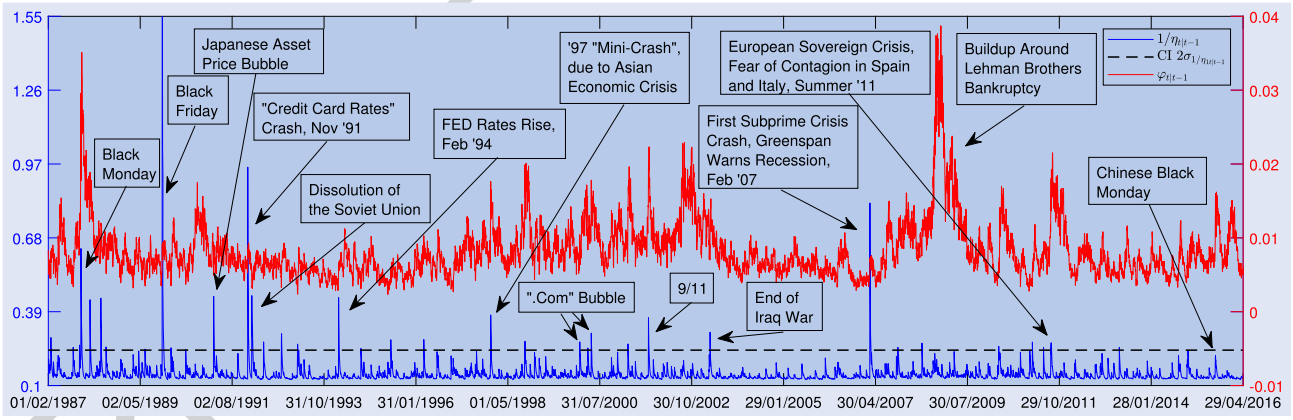


Figure 10. Plot of the fitted scale $\exp(\lambda_{t|t-1})$ without leverage, and the inverse of the fitted degrees of freedom for the Lower Tail $\bar{\eta}_{1t|t-1}$ of the Dow Jones Index Returns in the asymmetric case, with the confidence bounds for one and two standard deviations from the mean.

more effective for series exhibiting more frequent extreme events.

4.2. *Modelling CDS dataseries*

We fitted the scale of the CDS series using a Beta-*t*-EGARCH model with single-component dynamics. This was sufficient to remove most of the residual correlation in the fitted scores for $\lambda_{t|t-1}$ up to lag 25 across all data series. As seen in the results in table 7, the simple Q^* test rejects the null hypothesis of static degrees of freedom in both the symmetric and asymmetric cases. This suggests that both tails are dynamic for Italy and Portugal, while only the upper tail is detected as dynamic for France.

In all cases, the three fitted dynamic tail index parameters are highly persistent, nearly $I(1)$. The improved fit of the Dynamic Scale-Tail specification is confirmed by higher likelihood values and lower information criteria.

Figures 12 and 11 show that the fitted tail index parameters for the 5-year Italy CDS are more persistent than those for the Index Returns.† In the asymmetric case, the parameters tend to move closely together, fluctuating between 4 and 1. In both cases, they remain below 2 for extended periods early in the sample, particularly around the 2008 Lehman default. As seen in figure 11, the symmetric tail index parameter also drops below 1 during this period, suggesting that the conditional distribution might, at times, lack finite variance.‡ This outcome highlights the limitations of dynamic models, such as GARCH, which focus on moments of the conditional distribution, as they may become invalid in such contexts.

† This is because the CDS returns exhibit more frequent and extreme movements than the index returns.

‡ These low values coincide with periods when the CDS market is quite illiquid, with many consecutive zero returns, resulting in a very heavy-tailed conditional distribution. However, the total number of zeros in the sample is less than 5%.

Table 7. Goodness of fit and results of the simple $Q_u^*(P)$ test on both the fitted scores with respect to the scale parameter $Q_{u\lambda}(P)$, and with respect to the tail parameter $Q_{u\eta}(P)$ for different lags P. When a model with a static tail parameter, $\bar{\eta}_i$, is fitted we report the results of the simple $Q_u^*(P)$ test. If we fit a model with dynamic tails $\hat{\eta}_{it|t-1}$, we report the result of the Box-Ljing test applied to the fitted scores with respect to the tail parameter. In the case of $Q_{u\lambda}(*)$ and $Q_{u\eta}(*)$ the lags P are determined by the algorithm of Escanciano and Lobato (2009). The reported significance levels are ' ****' for 0.01' ***' for 0.05 and ' ****' for 0.1. In case of asymmetric tails we report instead the results for the test on the fitted scores of each tail, lower $Q_{u\eta_1}(P)$ and upper $Q_{u\eta_2}(P)$. The models compared for each series are, a Beta-t-EGARCH with two components for scale with: (1) a symmetric t with a single static $\hat{\eta}$, (2) a symmetric t with a single dynamic $\eta_{t|t-1}$, (3) an asymmetric t with static $\hat{\eta_1}$ and $\hat{\eta_2}$, (9) an asymmetric t with dynamic $\eta_{1t|t-1}$ and $\eta_{2t|t-1}$, (5) a GARCH, (6) a GARCH-t, all the models do not assume leverage.

	Model	AIC	BIC	Logl	$Q_{u\lambda}(1)$	$Q_{u\lambda}(10)$	$Q_{u\lambda}(25)$	$Q_{u\eta_1}(1)/Q_{u\eta}(1)$	$Q_{u\eta_1}(10)/Q_{u\eta}(10)$	$Q_{u\eta_1}(25)/Q_{u\eta}(25)$	$Q_{u\eta_1}(*)/Q_{u\eta}(*)$	$Q_{u\eta_2}(1)$	$Q_{u\eta_2}(10)$	$Q_{u\eta_2}(25)$	$Q_{u\eta_2}(*)$
5y Italy CDS	(1)	− 10.985.92	− 10.950.38	5498.96	0.3133	4.5195	11.323	17.288****	37.877****	74.174****	70.364****	–	–	–	–
	(2)	− 11074.42	− 11027.04	5545.21	0.0279	5.0688	13.882	0.1655	5.6789	16.714	0.1633	–	–	–	–
	(3)	− 10.986.21	− 10.944.75	5500.11	0.3150	4.3879	11.270	0.4580	14.231	68.103****	68.811****	4.3717***	25.710****	48.099****	4.3453***
	(9)	− 11036.65	− 10.971.50	5529.33	0.1075	4.5801	13.677	0.2345	2.7825	14.826	0.2295	1.1554	5.7061	11.842	1.1331
	(5)	− 10.312.19	− 10.282.57	5161.09	0.3648	2.5576	15.023	–	–	–	–	–	–	–	–
	(6)	− 10.877.34	− 10.841.80	5444.67	0.4137	4.3456	23.729	–	–	–	–	–	–	–	–
5y Portugal CDS	(1)	− 11027.44	− 10.991.90	5519.72	2.6428	14.449	28.222	0.0029	20.151***	49.616****	70.575****	–	–	–	–
	(2)	− 11036.24	− 10.988.86	5526.12	2.4055	14.633	28.750	0.1868	5.6003	12.336	0.1000	–	–	–	–
	(3)	− 11031.28	− 10.989.81	5522.64	2.4766	14.376	28.008	0.7220	8.7461	62.437****	54.534****	0.0351	14.092	39.650***	61.966****
	(9)	− 11035.01	− 10.969.85	5528.50	3.0443**	14.724	28.551	1.2325	7.1536	48.421****	0.9588	0.0045	3.7363	10.809	0.0139
	(5)	− 10.637.80	− 10.608.19	5323.90	0.6681	4.0382	17.773	–	–	–	–	–	–	–	–
	(6)	− 11023.66	− 10.988.12	5517.83	0.6764	9.4588	30.334	–	–	–	–	–	–	–	–
10y France CDS	(1)	− 11717.13	− 11681.59	5864.56	2.2383	10.323	24.215	20.877****	33.088****	64.868****	63.708****	–	–	–	–
	(2)	− 11750.06	− 11702.67	5883.03	0.8498	8.7395	26.183	1.7330	8.1603	21.556	1.7330	–	–	–	–
	(3)	− 11716.71	− 11675.25	5865.36	2.2152	10.703	24.298	0.0471	14.323	26.802	6.7706***	0.9413	9.1862	51.603****	50.530****
	(9)	− 11729.54	− 11676.23	5873.77	1.7711	11.555	26.563	–	–	–	–	0.0427	4.5258	20.528	0.0427
	(5)	− 10.597.25	− 10.567.63	5303.62	0.0054	1.4434	12.004	–	–	–	–	–	–	–	–
	(6)	− 11401.45	− 11365.91	5706.72	0.3947	1.5763	13.517	–	–	–	–	–	–	–	–

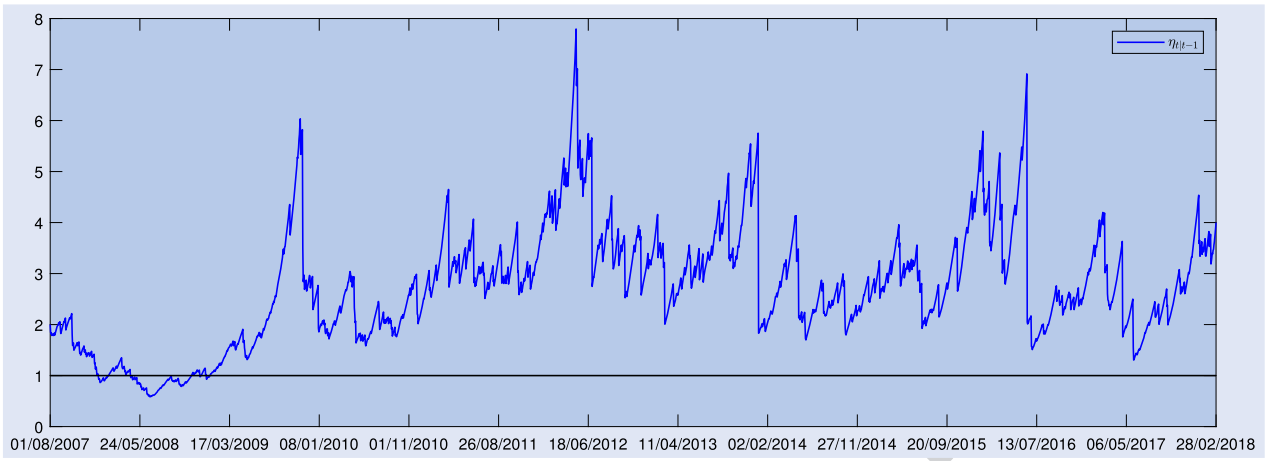


Figure 11. Plot of the fitted estimated degrees of freedom, $\eta_{t|t-1}$ for the symmetric model.

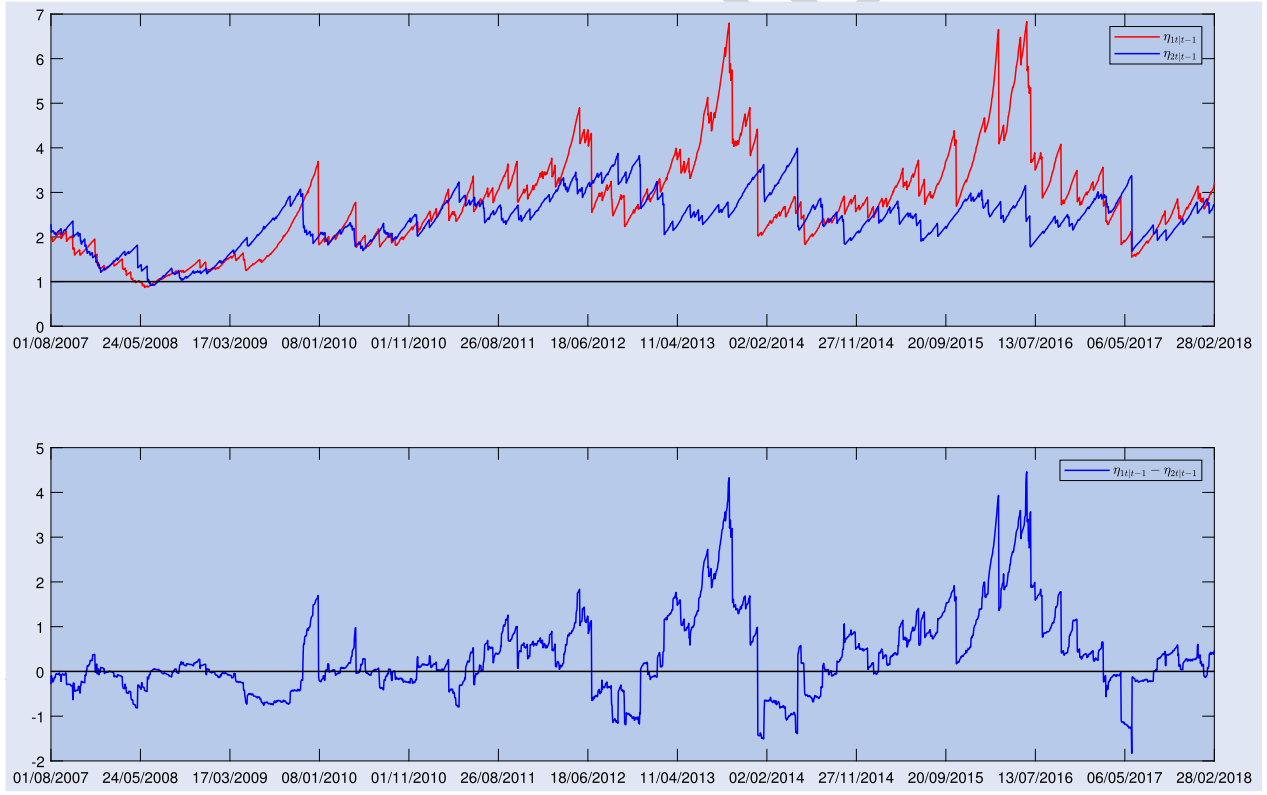


Figure 12. The top figure shows a plot of the fitted estimated degrees of freedom, $\eta_{2t|t-1}$ and $\eta_{1t|t-1}$, for the upper and lower tail in the asymmetric model. On the bottom figure shows the spread between the upper and lower tail dynamic degrees of freedom of the asymmetric distribution, $\eta_{2t|t-1} - \eta_{1t|t-1}$.

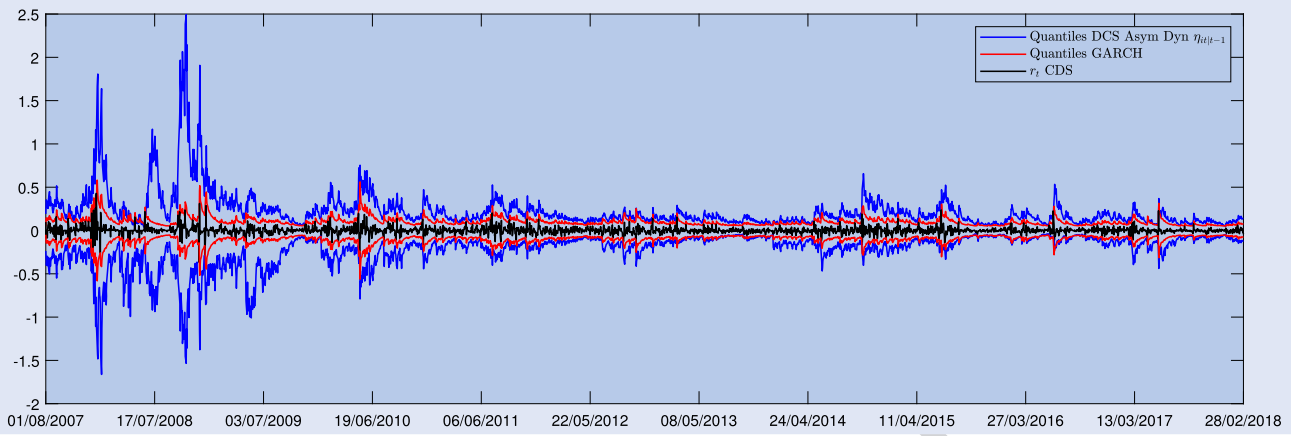
Moreover, the tail index parameter falls below 1 only for brief periods, despite not being bounded. As explained in the appendix, the score naturally pushes the tail index parameter away from extremely low values unless there is a large number of consecutive extreme observations. The tail index parameter for the symmetric distribution tends to move the most, closely following the movements of the lower tail, though it occasionally exceeds the values of either tail index parameter in the asymmetric specification.

Finally, if we express the relative comparison of the tail parameters of the asymmetric distribution as a spread, $\eta_{2t|t-1} - \eta_{1t|t-1}$, it becomes informative of periods of financial distress for the country. Indeed, periods when the spread turns negative† coincide with times of economic and political turbulence in Italy, during which CDS rates increased rapidly.

4.3. Conditional distribution modelling under dynamic tails

The inclusion of dynamic tails directly affects the modelling of the conditional distribution of the data. This effect can also be appreciated by examining its quantile levels. Figure 13 illustrates the upper and lower 0.5% quantiles fitted by the GARCH model and the asymmetric Dynamic Tail DCS model

† This indicates periods when the lower tail is closer than the upper tail to being Gaussian, suggesting more extreme positive returns.



Colour online, B/W in print

Figure 13. Plot of the 0.5% upper and lower quantiles of the conditional distribution of the returns series for the Italian 5y CDS Rates, when fitted by a GARCH model and a score-driven with dynamic asymmetric tails.

for the Italian 5-year CDS returns data series. It is evident that the returns data frequently reaches the upper and lower quantiles estimated by the GARCH model. This indicates that there have been numerous instances within the dataset where returns exhibit movements that should occur with a probability of 0.5%. Specifically, across the entire sample, 1.36% of the data breaches the GARCH upper quantile and 1.19% breaches the lower quantile. In contrast, for the asymmetric Dynamic Tail DCS model, only 0.03% of observations exceed both the upper and lower tails. This suggests that the GARCH model is considerably less conservative than the Dynamic Tail Index model, underestimating the frequency of extreme events. This discrepancy is also evident in the event of 15 July 2008, two months before the Lehman Brothers' bankruptcy, when the 5-year Italy CDS surged from 21.167 to 32.5 in a single day. The GARCH model estimated the probability of such an event at 0.06%, whereas the static Beta-*t*-EGARCH DCS model assigned a probability of 0.57%, and the symmetric and asymmetric Dynamic Tail DCS models estimated probabilities of 2.22% and 1.18%, respectively.

To determine whether these significant differences in the conditional distribution are also observable out-of-sample, we conducted a density forecasting exercise. We obtained one-step-ahead point and density forecasts for the best performing models for the Dow Jones Index and the 5-year Italy CDS data series over the final two years of each sample. These would be, for the Dow Jones Index returns a dynamic asymmetric tail model, with two components for scale, with a static lower tail and a dynamic upper tail, while for the Italy CDS returns a dynamic asymmetric tail model, also with two components dynamic scale, with both upper and lower tail dynamic. The out-of-sample periods contain a total of 730 out-of-sample observations, between 01/05/2014–29/04/2016 for the Dow Jones Index and between 01/03/2016–28/02/2018 for the CDS. The forecasts were generated by re-estimating the models using all available data up to the date of each forecast.

For the evaluation we look at the values of the predictive likelihood and the average CRPS of Gneiting and Raftery (2007) across the forecasted period for each model.

In addition, we define the one-step-ahead p-lower and p-upper Value-at-Risk (VaR) as follows,

$$\mathrm{VaR}_{1p}\left(y_{T+1}\right) = \inf\left\{x \in \mathbb{R} : \mathbb{P}\left(y_{T+1} \leq x | \mathcal{F}_T\right) \geq p\right\},$$
$$\mathrm{VaR}_{2p}\left(y_{T+1}\right) = \sup\left\{x \in \mathbb{R} : \mathbb{P}\left(y_{T+1} \geq x | \mathcal{F}_T\right) \leq 1-p\right\}$$

For a symmetric distribution centred at zero, these correspond to $\mathrm{VaR}_{1,1-p}(y_{T+1}) = F^{-1}_{Y_{T+1|T}}(p)$ and $\mathrm{VaR}_{2p}(y_{T+1}) = -F^{-1}_{Y_{T+1|T}}(p)$, where $F_{Y_{T+1|T}}(\cdot)$ is the one-step-ahead forecasted conditional cumulative distribution function (CDF) of y_t. In the case of an asymmetric distribution centred at zero, with one-step-ahead forecasted conditional CDFs $F_{1Y_{T+1|T}}(\cdot)$ and $F_{2Y_{T+1|T}}(\cdot)$ for the left and right tails of y_t, respectively, we have $\mathrm{VaR}_{1p}(y_{T+1}) = F^{-1}_{1Y_{T+1|T}}(p)$ and $\mathrm{VaR}_{2p}(y_{T+1}) = -F^{-1}_{2Y_{T+1|T}}(p)$.

On the other hand, we define the one-step-ahead lower and upper Expected Shortfall (ES)† as,

$$\begin{aligned} ES_{1p}\left(y_{T+1}\right) &= E\left[-y_{T+1} | y_{T+1} \leq \mathrm{VaR}_{1p}\left(y_{T+1}\right), \mathcal{F}_T\right] \\ &= -\frac{1}{p}\int_{-\infty}^{\mathrm{VaR}_{1p}(y_{T+1})} x f_{Y_{T+1|T}}(x)\, \mathrm{d}x \\ ES_{2p}\left(y_{T+1}\right) &= E\left[y_{T+1} | y_{T+1} \geq \mathrm{VaR}_{2p}\left(y_{T+1}\right), \mathcal{F}_T\right] \\ &= \frac{1}{p}\int_{\mathrm{VaR}_{2p}(y_{T+1})}^{\infty} x f_{Y_{T+1|T}}(x)\, \mathrm{d}x \end{aligned}$$

Given these measures of one-step-ahead VaR and ES over the out-of-sample forecasting period, we then define the hit processes:

$$h_{1T+1}\left(p\right) = \mathbf{1}_{\left(y_{T+1} < \mathrm{VaR}_{1p}(y_{T+1})\right)}$$

† In principle our dynamic tail index models doesn't restrict the degrees to any values and if they fall below 1, as at the beginning of the CDS sample, the expectation for the ES couldn't be computed. If one would like in practice to be able to compute ES exactly in all the cases, it is still possible in our general framework to set the tail lowerbound $\eta_L = 1$. However, the results in this section are still valid even without imposing this restriction since we have already showed that, in the out-of-sample period selected for forecasting the CDS distribution, the tail index never falls below 1, not even in the forecasts.

Table 8. Results for the Dow Jones Index of the unconditional coverage likelihood ratio tests of Christoffersen (1998) for the upper an power tail as well as the results for the unconditional backtest of Du and Escanciano (2017) to evaluate the upper and lower one-step-ahead ES accuracy. ∗, ∗∗, ∗ ∗ ∗ define rejections with confidence levels of 0.1, 0.05, and 0.01 respectively.

		Symmetric Tails				Asymmetric Tails			
		Christoffersen		Du-Escanciano		Christoffersen		Du-Escanciano	
	Quantile	Lower Tail	UpperTail	Lower Tail	Upper Tail	Lower Tail	UpperTail	Lower Tail	Upper Tail
Static Tails	0.15	0.1327	0.2151	0.1325	− 1.362	0.1327	0.5933	− 0.360	− 1.729**
	0.1	0.0002	0.3885	0.2586	− 2.448***	0.1386	0.1386	− 0.310	− 2.813****
	0.05	0.0640	9.3262****	0.2161	− 3.028****	0.1842	8.1267****	− 0.351	− 3.364****
	0.01	0.3723	5.4598***	− 0.258	− 1.993***	0.2489	3.2899**	− 0.828	− 2.318***
	0.005	0.8974	2.7202**	− 0.956	− 1.572	2.7202**	2.7202**	− 1.253	− 1.657**
	Pred. Lik.	2531.84				2533.78			
	Avg. CRPS	0.6101				0.6053			
		Christoffersen		Du-Escanciano		Christoffersen		Du-Escanciano	
	Quantile	Lower Tail	UpperTail	Lower Tail	Upper Tail	Lower Tail	UpperTail	Lower Tail	Upper Tail
Dynamic Tails	0.15	0.0242	0.4467	0.1412	− 1.289	0.2201	0.5933	− 0.399	− 1.206
	0.1	0.0001	0.2475	0.2295	− 2.324***	0.2476	0.0528	− 0.335	− 2.241***
	0.05	0.0071	7.0252****	0.2118	− 2.973****	0.0073	6.0175***	− 0.378	− 2.926****
	0.01	0.3723	5.4598***	− 0.284	− 1.918**	0.0126	3.2900	− 0.906	− 2.126***
	0.005	0.8974	2.7202**	− 1.245	− 1.442	2.7202**	2.7202**	− 1.277	− 1.657**
	Pred. Lik.	2529.97				2534.60			
	Avg. CRPS	0.5882				0.5100			

$$h_{2T+1}(p) = \mathbf{1}_{(y_{T+1} > \text{VaR}_{2p}(y_{T+1}))}$$

Using these, we can construct the unconditional coverage likelihood ratio tests of Christoffersen (1998) and Kupiec (1995) for VaR violations in both the upper and lower tails.† This corresponds to the null hypothesis H_0 : $E[h_{iT+1}(p)] = \mathbb{P}(h_{iT+1}(p) = 1) = p$. For each individual tail, the test statistic follows a $\chi^2(1)$ distribution.

To evaluate the ES for both tails, we employed the unconditional backtest of Du and Escanciano (2017). These tests are constructed from the cumulative violation processes,

$$\begin{aligned} H_{1T+1}(p) &= \frac{1}{p}\int_0^p h_{1T+1}(q)\, \mathrm{d}q \\ &= \frac{1}{p}(p - PIT_{T+1})\,\mathbf{1}_{(y_{T+1} < \text{VaR}_{1p}(y_{T+1}))} \\ H_{2T+1}(p) &= \frac{1}{p}\int_{1-p}^1 h_{2T+1}(q)\, \mathrm{d}q \\ &= \frac{1}{p}(1 - p - PIT_{T+1})\,\mathbf{1}_{(y_{T+1} > \text{VaR}_{2p}(y_{T+1}))} \end{aligned}$$

where $PIT_{T+1} = F_{Y_{T+1|T}}(y_{T+1})$ are the conditional one-step-ahead probability integral transforms (PIT) computed on the out-of-sample data. Du and Escanciano (2017) demonstrate that testing the correct specification of the ES simplifies to testing whether the mean of $H_{iT+1}(p)$ equals $p/2$, which can be assessed using the t-statistics

$$t_{ES} = \frac{\overline{H_i(p)} - p/2}{\sqrt{v_{ES}(p)/T_f}}$$

where $\overline{H_i(p)}$ is the sample mean of $H_{iT+1}(p)$, $v_{ES} = \text{Var}(H_{iT+1}(p)) = p(\frac{1}{3} - \frac{p}{4})$, and T_f is the number of out-of-sample observations.‡

The results of the unconditional coverage tests in tables 8 and 9 indicate that, for models with fixed tails, some quantile levels are significantly misspecified compared to models with dynamic tails. Specifically, this mis-specification is evident in the upper tail of the Dow Jones Index and in both the upper and lower tails of the 5-year Italy CDS. This occurs because the time variation in the quantiles for fixed tail models depends solely on variations in the conditional variance, which tends to spike in the presence of extreme events. Consequently, these models overestimate the quantiles closer to the median at the expense of those in the tails.

In the case of the Dow Jones Index (see table 8), the lower tail quantiles are accurately captured by all four models in terms of both unconditional coverage and unconditional backtests. However, the upper tail quantiles are significantly better captured by the dynamic tail models. Furthermore, when employing asymmetric dynamic tails, all unconditional backtests improve, and almost all violations of the unconditional coverage tests become insignificant, except for the 0.05% quantile. Additionally, the predictive likelihood out-of-sample is higher for the asymmetric tails models, with the highest likelihood observed for the dynamic tails model and the lowest average CRPS score.

For the 5-year Italy CDS (see table 9), the predictive likelihood results are similar across models as well as the average CRPS score, with a slight preference for the symmetric dynamic tail index model. However, there are a few rejections

† We have also considered the independence test of Christoffersen (1998), however the low number of rejections was insufficient to discriminate between models, therefore the results are not reported.

‡ We have also considered the conditional backtest of Du and Escanciano (2017), but the low number of rejections was insufficient to discriminate between models and the results are not reported.

Table 9. Results for the 5y Italy CDS of the unconditional coverage likelihood ratio tests of Christoffersen (1998) for the upper an power tail, as well as the results for the unconditional backtest of Du and Escanciano (2017) to evaluate the upper and lower one-step-ahead ES accuracy. $*, **, ***$ define rejections with confidence levels of 0.1, 0.05, and 0.01 respectively.

		Symmetric Tails				Asymmetric Tails			
		Christoffersen		Du-Escanciano		Christoffersen		Du-Escanciano	
	Quantile	Lower Tail	UpperTail	Lower Tail	Upper Tail	Lower Tail	UpperTail	Lower Tail	Upper Tail
Static Tails	0.15	0.7934	2.0289	− 1.787**	− 0.940	0.4615	2.0289	− 1.319	− 0.615
	0.1	3.1714**	0.5618	− 1.924**	− 0.442	2.3084	1.2809	− 1.456	− 0.012
	0.05	3.5189**	0.0657	− 1.983***	− 0.764	2.8498**	0.9173	− 1.533	− 0.014
	0.01	5.4598***	0.8229	− 1.266	− 0.568	5.4598***	1.8024	− 1.211	1.3263
	0.005	0.8974	0.1238	− 0.419	-0.037	0.8974	0.1238	− 0.263	3.2489****
	Pred. Lik.	1695.18				1695.77			
	Avg. CRPS	0.7756				0.7752			
		Christoffersen		Du-Escanciano		Christoffersen		Du-Escanciano	
	Quantile	Lower Tail	UpperTail	Lower Tail	Upper Tail	Lower Tail	UpperTail	Lower Tail	Upper Tail
Dynamic Tails	0.15	0.6161	2.6896	− 1.380	− 0.465	0.2200	4.3058***	− 0.945	− 1.028
	0.1	1.0076	0.3885	− 1.649**	0.3757	1.0076	1.2809	− 1.124	− 0.443
	0.05	3.5189**	1.1554	− 1.541	0.6966	2.2577	0.6082	− 0.508	− 0.156
	0.01	0.8229	0.3723	− 0.721	0.5428	0.3723	1.8024	0.4878	1.5806
	0.005	0.8974	0.0327	− 0.026	1.0180	0.4496	0.1238	0.0468	4.8936****
	Pred. Lik.	1696.58				1695.15			
	Avg. CRPS	0.7660				0.7798			

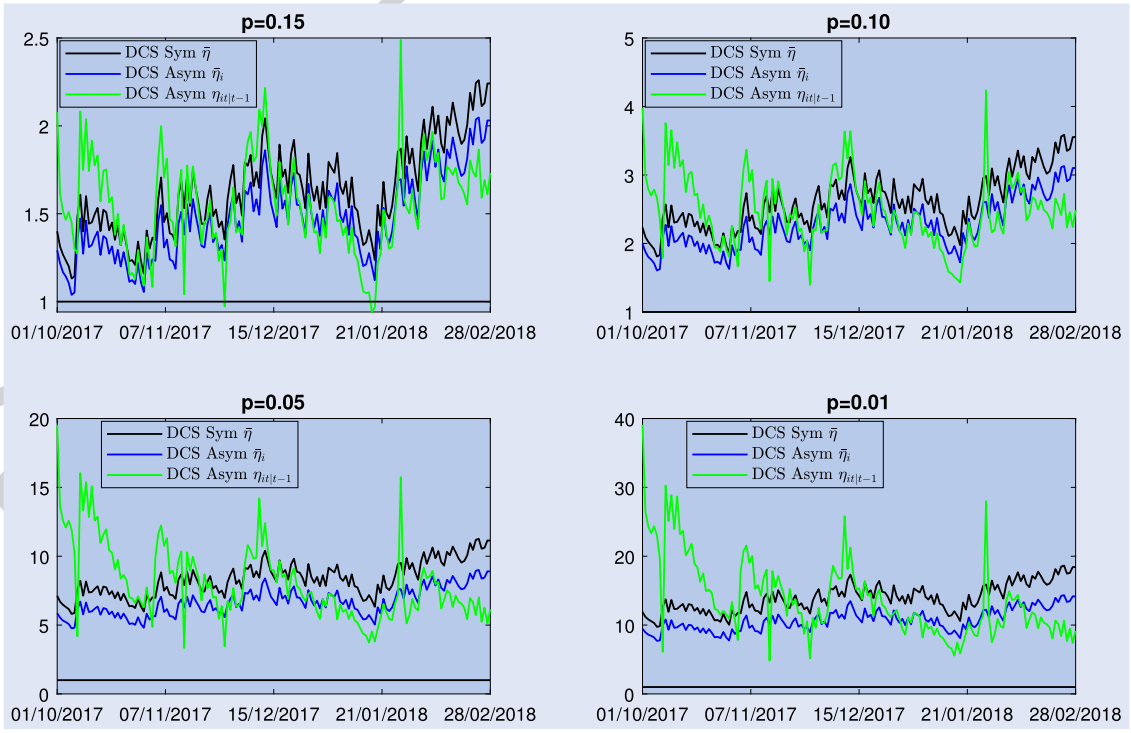


Figure 14. Plot of the ratios of the Expected Shortfall above the lower 10% (Top Left), 5% (Top Right), 1% (Bottom Left) and 0.5% (Bottom Right) quantiles of the one-step-ahead forecasted conditional distribution of the 5y Italian CDS Rate Returns from fitting a symmetric score–driven Beta-t-EGARCH model, an asymmetric score-driven Beta-t-EGARCH model, an symmetric dynamic tail score-driven EGARCH model over the one-step-abead forecasted Expected Shortfall from a GARCH model.

Colour online, B/W in print

of the unconditional coverage and backtests at higher quantiles for the static tail models, predominantly concentrated in the lower tail. Conversely, when dynamic tails are included, these rejections largely disappear across almost all quantiles for both tails, with only a few rejections at the 10% significance level in the lower tail for the symmetric tail model and at the 5% significance level in the upper tail for the asymmetric model.

As a further illustration of the results, figures 14 and 15 display the lower and upper out-of-sample Expected Shortfall (ES), respectively, expressed as ratios relative to the ES forecasted by a GARCH model. These figures demonstrate that

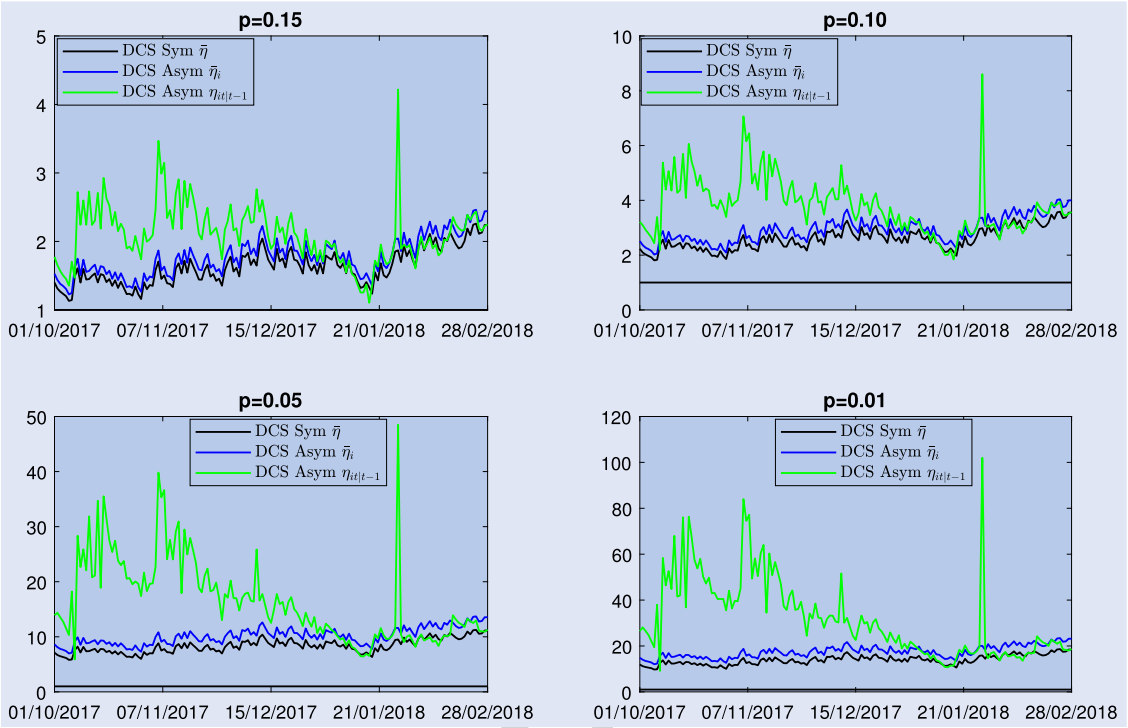


Figure 15. Plot of the ratios of the Expected Shortfall below the upper 10% (Top Left), 5% (Top Right), 1% (Bottom Left) and 0.5% (Bottom Right) quantiles of the one-step-ahead forecasted conditional distribution of the 5y Italian CDS Rate Returns from fitting a symmetric score–driven Beta-*t*-EGARCH model, an asymmetric score-driven Beta-*t*-EGARCH model, an symmetric dynamic tail score-driven EGARCH model over the one-step-abead forecasted Expected Shortfall from a GARCH model.

the GARCH model underestimates the length of the tails in both cases. Specifically, for the forecasted 10% quantile, the ES estimates from the GARCH model are approximately half those forecasted by the dynamic tails models. For the 0.1% quantile, the ES estimates range from five to more than thirty-five times higher in the dynamic tails models compared to the GARCH model.

In general, the ES ratios from the asymmetric fixed tail model tend to be higher than those from the symmetric fixed tail model for the upper tail and lower for the lower tail, highlighting the asymmetry in the out-of-sample distribution. Conversely, for the asymmetric dynamic tail model, the ES ratios exhibit considerable variability across the out-of-sample period, with the largest fluctuations occurring at the 0.01% quantile. For the lower tail, most ES ratios are lower than those of the asymmetric fixed tail DCS model, whereas for the upper tail, the ratios fluctuate rapidly, moving both above and below depending on the time periods.

5. Conclusion

This study develops a testing and modelling framework to capture the variation in the occurrence of extreme events in time series, which can be described by fluctuations in the tail index of the conditional distribution. The paper introduces a dynamic score-driven model for the tail index parameter, assuming the data are generated by a conditional symmetric or asymmetric *t* distribution.

An LM test is introduced to detect dynamics in the tail index parameters, based on the autocorrelation of the score under the null hypothesis of no variability. A closed-form solution for the test is derived under the assumption of symmetric tails. The power and size of the full LM test are compared with a simple Box-Ljung test performed on the fitted scores under the null of no variation. The results show that the simple Box-Ljung test is a viable alternative to the full LM test with similar rejection probabilities in most of the cases, and with even higher power in the case of a low tail index. The comparison is extended to include the GAS-LM test of Calvori *et al.* (2017) and the Nyblom test. The GAS-LM test, performs better when the tail index is low and less persistent. However, in most other cases, it is outperformed by the LM and Box-Ljung tests. The Nyblom test applied to the fitted scores with respect to the tail parameter, remains still a viable competitive alternative in the presence of a low tail index. These results remain consistent for the simple Box-Ljung test when the framework is extended to asymmetric tails.

The dynamic tail score-driven model's efficiency in estimating the dynamic tail index parameters is also assessed under various parameter assumptions. As expected, estimation accuracy decreases as sample sizes shrink, especially when the true tail index is around 30 or larger. Conversely, the model performs well when the true tail index is smaller than 15. The analysis, discussed in the appendix, further explores the implications of bounding the tail index parameter to ensure the existence of moments. Bounding can introduce significant distortions in the score response, affecting the model's ability to capture the true dynamics of the tail index.

Both the symmetric and asymmetric models, as well as the tests, are applied to Equity Index and CDS returns data. For Equity Indexes, the test detects dynamics in the symmetric tail index parameter. However, assuming an asymmetric

distribution, dynamics are detected only in either the upper or lower tail index, depending on whether a leverage term is included in the scale dynamics. In all cases, the fitted dynamic tail indexes are not highly persistent and are generally bounded from above, rarely falling below 1. For CDS returns, dynamics are detected in both the symmetric and asymmetric tail index parameters, with all three parameters displaying highly persistent dynamics, moving from 5-7 down to below 1 occasionally. The analysis of the spread between the upper and lower tail index parameters in the asymmetric case reveals considerable variation in the relative heaviness of the two tails over time.

Moreover, an out-of-sample analysis of forecasted quantiles and Expected Shortfalls demonstrates that the dynamic tail score-driven models are more accurate at capturing out-of-sample densities. They are also less conservative than GARCH models in forecasting tail lengths, predicting higher probabilities of extreme events, with significant evidence of asymmetries and time variation in magnitudes depending on the time period.

These findings and the methodology to capture tail behaviour are highly relevant for both empirical researchers and practitioners and are worth applying in other contexts. Additionally, it would be interesting to investigate whether, in the asymmetric case, an increase in the heaviness of one tail could dynamically affect the size of the other, as seen in the empirical example of Massacci (2017). In a similar vein, exploring the possibility of a factor structure could reveal common or specific factors driving tail movements. These factors could potentially be linked to observable explanatory variables related to market structure or the broader economy. Finally, from a systemic risk perspective, studying an asymmetric multivariate framework with several assets would be useful for assessing potential dynamic tail relationships across assets or countries, offering new insights into the concept of tail association.

Acknowledgments

I am grateful for fruitful discussions, comments and suggestions made by Andrew Harvey, Alexander Mayer, Enrico Ghiorzi, Davide delle Monache, Daniele Massacci, Neil Shephard and Andrè Lucas, as well as the editor and the two anonymous referees. I also thank all the participants of the 2019 Cambridge Score Driven Workshop, the 12th International Conference of the ERCIM WG on Computational and Methodological Statistics, the European Economics Association World Congress in 2020, the 9th Italian Congress of Econometrics and Empirical Economics, the Econometric Society World Congress in 2020, the XVXI Quantitative Finance Workshop and of the 2024 Workshop in Emerging Risks in Economics and Finance at the University of Florence. I am responsible for all remaining errors. I am thankful for King's College for having funded my PhD while I have been working on this paper. At the same time I thank the Italian Econometric Association (SIdE), the Unicredit Foundation and Bank of Italy for funding my Carlo Giannini Research Fellowship in Econometrics.

Disclosure statement

No potential conflict of interest was reported by the authors

Supplemental data

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/14697688.2026.2667881.

References

Allen, L., Bali, T. and Tang, Y., Does systemic risk in the financial sector predict future economic downturns? *the Review of Financial Studies¡/DIFdel¿Rev. Financ. Stud.*, 2012, **25**(10), 3000–3036.

Andrews, D.W.K., Test for parameter instability and structural change with unknown change points. *Econometrica*, 1993, **25**, 821–586.

Ayala, A., Blazsek, S. and Escribano, A., Does systemic risk in the financial sector predict future economic downturns? Working paper, 17-08, 2017.

Blasques, F., Gorgi, P., Harvey, A.C., Koopman, S.J. and Lasak, K., A quasi-score driven approach to dynamic model specification. *J. Econom.*, 2023, **236**(2), 591–618.

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

Blazsek, S. and Monteros, L., Dynamic conditional score models of degrees of freedom: Filtering with score-driven heavy tails. *Appl. Econ.*, 2017, **49**(53), 5426–5440.

Box, G.E.P. and Pierce, D.A., Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.*, 1970, **65**(332), 1509–1526.

Brooks, C., Burke, S., Heravi, S. and Persand, G., Autoregressive conditional kurtosis. *J. Finan. Econom.*, 2005, **3**(3), 399–421.

Calvori, F., Creal, D., Koopman, S. and Lucas, A., Testing for parameter instability across different modeling frameworks. *J. Financ. Econom.*, 2017, **15**(2), 223–246.

Cavaliere, G., Rahbek, A. and Taylor, A.R., Bootstrap inference in garch-type models with heavy-tailed errors. *J. Econom.*, 2018, **202**(1), 1–19.

Christoffersen, P., Evaluating interval forecasts. *Int. Econ. Rev.)*, 1998, **39**(4), 841–862.

Creal, D., Koopman, S.J. and Lucas, A., Generalized autoregressive score models with applications. *J. Appl. Econom.*, 2013, **28**, 777–795.

D'Innocenzo, E., Lucas, A., Schwaab, B. and Zhang, X., Modeling extreme events: Time-varying extreme tail shape. *J. Bus Econ. Stat.*, 2024, **42**(3), 903–917.

Delle Monache, D., De Polis, A. and Petrella, I., Modeling and forecasting macroeconomic downside risk. *J. Bus. Econ. Stat.*, 2024, **42**(3), 1010–1025.

Du, Z. and Escanciano, J., Backtesting expected shortfall: Accounting for tail risk. *Manage Sci.*, 2017, **63**(4), 940–958.

Dufour, J.-M. and Roy, R., Some robust exact results on sample autocorrelations and tests of randomness. *J. Econom.*, 1985, **29**(2), 193–221.

Embrechts, P., Klüppelberg, C. and Mikosch, T., *Modelling Extremal Events: for Insurance and Finance*, 1997 (Springer-Verlag).

Engle, R. and Mustafa, C., Implied arch models from option prices. *J. Econom.*, 1992, **52**(1), 289–311.

Escanciano, J. and Lobato, I., An automatic portmanteau test for serial correlation. *J. Econom.*, 2009, **151**(2), 140–149.

Galbraith, J. and Zernov, S., Circuit breakers and the tail index of equity returns. *J. Financ. Econom.*, 2004, **2**(1), 109–129.

Gneiting, T. and Raftery, A.E., Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, 2007, **102**(477), 359–378.

Hall, P., Horowitz, J.L. and Jing, B.-Y., On blocking rules for the bootstrap with dependent data. *Biometrika*, 1995, **82**(3), 561–574.

Hansen, B., Autoregressive conditional density estimation. *Int. Econ. Rev.*, 1994, **35**(3), 705–730.

Hansen, B., Criterion-based inference without the information equality: The weighted chi-square distribution, Working Paper, 2021.

Harvey, A.C., Dynamic models for volatility and heavy tails: With applications to financial and economic time series. *Econom. Soc. Monogr.*, 2013.

Harvey, A. and Lange, R.-J., Volatility modelling with a generalized-t distribution. *J. Time Ser. Anal.*, 2017, **38**(2), 175–190.

Harvey, A. and Lange, R.-J., Modeling the interactions between volatility and returns. *J. Time Ser. Anal.*, 2018, **39**(6), 909–919.

Harvey, A.C. and Luati, A., Filtering with heavy tails. *J. Am. Stat. Assoc.*, 2014, **109**(507), 1112–1122.

Harvey, C. and Siddique, A., Autoregressive conditional skewness. *J. Financ. Quant. Anal.*, 1999, **34**(4), 465–487.

Harvey, A. and Streibel, M., Tests for deterministic versus indeterministic cycles. *J. Time Ser. Anal.*, 1998, **19**(5), 505–529.

Harvey, A. and Thiele, S., Testing against changing correlation. *J. Empir. Financ.*, 2016, **38**, 575–589.

Horowitz, J.L., Lobato, I., Nankervis, J.C. and Savin, N., Bootstrapping the box–pierce q test: A robust test of uncorrelatedness. *J. Econom.*, 2006, **133**(2), 841–862.

Hosking, J.R.M., Lagrange-multiplier tests of time-series models. *J. R. Stat. Soc.: Ser. B (Methodol.)*, 1980, 12, **42**(2), 170–181.

Kan, R. and Wang, X., On the distribution of the sample autocorrelation coefficients. *J. Econom.*, 2010, **156**(2), 135–144.

Kelly, B., The dynamic power law model. *Extremes*, 2014, **17**(4), 557–583.

Kelly, B. and Jiang, H., Tail risk and asset prices. *Rev. Financ. Stud.*, 2014, **27**(10), 2841–2871.

Kupiec, P.H. „*Techniques for Verifying the Accuracy of Risk Measurement Models*. Finance and Economics Discussion Series 95-10, 1995 (Board of Governors of the Federal Reserve System: Washington, DC).

Lahiri, S.N., Theoretical comparisons of block bootstrap methods. *Ann. Stat.*, 1999, **27**(1), 386–404.

Lahiri, S.N., *Resampling Methods for Dependent Data*, 2003 (Springer: New York).

Lamoureux, C. and Lastrapes, W., Persistence in variance, structural change and the garch model. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, 2002, **64**(2), 253–280.

Lee, S. and Hansen, B., Asymptotic theory for the garch(1,1) quasi-maximum likelihood estimator. *Econ. Theory*, 1994, **10**(1), 29–52.

Li, W.K. and Mak, T.K., On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *J. Time Ser. Anal.*, 1994, **15**(6), 627–636.

Ling, S. and Li, W.K., On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Am. Stat. Assoc.*, 1997, **92**(439), 1184–1194.

Lucas, A. and Zhang, X., Score-driven exponentially weighted moving averages and value-at-risk forecasting. *Int. J. Forecast.*, 2016, **32**(2), 293–302.

Massacci, D., Tail risk dynamics in stock returns: „Links to the macroeconomy and global markets connectedness. *Manage. Sci.*, 2017, **63**(9), 2773–3145.

Mazur, B. and Pipień, M., Time-varying asymmetry and tail thickness in long series of daily financial returns. *Stud. Nonlinear Dyn. Econom.*, 2018, **22**(5), 1558–3708.

Muller, U. and Petalas, P., Efficient estimation of the parameter path in unstable time series models. *Rev. Econ. Stud.*, 2010, **77**(4), 1508–1539.

Nyblom, J., Testing for the constancy of parameters over time. *J. Am. Stat. Assoc.*, 1989, **84**(405), 223–230.

Politis, D.N. and White, H., Automatic block-length selection for the dependent bootstrap. *Econom. Rev.*, 2004, **23**(1), 53–70.

Quintos, C., Fan, Z. and Phillips, P., Structural change tests in tail behaviour and the asian crisis. *the Review of Economic Studies¡/DIFdel¿Rev. Econ. Stud.*, 2001, **68**(3), 633–663.

Thiele, S., Modeling the conditional distribution of financial returns with asymmetric tails. *J. Appl. Econom.*, 2020, **35**(1), 46–60.

Wagner, N., Autoregressive conditional tail behavior and results on government bond yield spreads. *Int. Rev. Financ. Anal.*, 2005, **14**, 247–261.

Werner, T. and Upper, C., Time variation in the tail behaviour of bund futures returns. *J. Fut. Markets*, 2004, **24**(4), 387–398.

Zhu, D. and Galbraith, J., A generalized asymmetric student-t distribution with application to financial econometrics. *J. Econom.*, 2010, **157**(2), 297–305.

Fig. 1.	**Alt-Text:** Four graphs showing score functions of residual epsilon against parameters theta, lambda, and eta, with curves for eta 2, 5, and 10.
Fig. 1.	**Extended Description:** The figure shows four statistical graphics related to score functions of a tail index model. The top row contains three two dimensional line graphs. In the first line graph, the horizontal axis is residual epsilon from about -2 to 2 with tick marks at 0.5 intervals, and the vertical axis is score with respect to theta from about -0.05 to 0.25. Three curves are plotted and labeled in a small legend as eta equals 2, eta equals 5, and eta equals 10. All three curves start near zero score at epsilon equal to -2, rise to a local maximum, dip toward a local minimum near epsilon equal to 0, and then rise again toward epsilon equal to 2, with the eta equals 2 curve highest and the eta equals 10 curve lowest. In the second line graph, the horizontal axis is residual epsilon from about -6 to 6 and the vertical axis is score with respect to theta from about -100 to 20. The three eta curves diverge strongly for large positive and negative residuals, with magnitudes increasing as eta increases. In the third line graph, the horizontal axis is residual epsilon from about -6 to 6 and the vertical axis is score with respect to lambda from about -1 to 8. The three eta curves form U shaped trajectories centered near epsilon equal to 0, with the lowest minimum for eta equals 10. All data are approximate. The bottom graphic is a three dimensional surface plot of the joint score as a function of residual epsilon on one axis, eta on the other axis from about 1 to 8, and score on the vertical axis from about 0 to 0.5. The surface forms two ridge like peaks at larger eta values while remaining near zero for small eta and moderate residual epsilon. All data are approximate.
Fig. 2.	**Alt-Text:** Two line graphs showing conditional score functions versus residual epsilon, with 3 curves for tail index values 2, 5, and 10 in each chart.
Fig. 2.	**Extended Description:** The figure shows two line graphs comparing conditional score functions for a score driven tail index model used in the asymmetric Student t specification of Thiele 2020. Both graphs plot smooth curves of the score against a horizontal residual axis labelled by the Greek letter epsilon. The horizontal axis runs from approximately -2 to 2 with tick marks at 0.5 unit intervals. The vertical axis in each graph ranges from about -0.15 to 0.2 with tick marks at 0.05 intervals. In the left graph, three curves represent different values of the first tail index parameter, labelled in the legend as eta 1 equals 2, eta 1 equals 5, and eta 1 equals 10. All three curves start near -0.1 at residual -2, rise to a positive peak around residual -1, then gradually decline and flatten near residual 2, with the eta 1 equals 2 curve lying above the others. In the right graph, three curves represent the second tail index parameter, labelled eta 2 equals 2, eta 2 equals 5, and eta 2 equals 10. These curves begin near zero for residual -2, stay almost flat and positive until around residual 0.5, then rise to a peak between residual 1 and 1.5 before dropping toward negative values near residual 2, with the eta 2 equals 2 curve again highest. All data are approximate.
Fig. 3.	**Alt-Text:** Four line graphs comparing empirical power versus lags for full LM and simple tests in Dynamic Scale Tail simulations with T values.
Fig. 3.	**Extended Description:** The figure shows four line graphs arranged in a 2 by 2 layout, each plotting empirical power against the number of lags for Dynamic Scale Tail simulations. In every graph, the horizontal axis is labeled Lags with tick marks at 1, 5, 10, 15, 25, 35, and 50. The vertical axis is labeled Power with values from 0 to 1, marked at 0, 0.2, 0.4, 0.6, 0.8, and 1. The top left graph is titled with parameter values for phi theta equal to 0.985 and kappa theta equal to 0.04. Several colored and styled lines represent the simple test and the full Lagrange Multiplier test for sample sizes T equal to 500, 1000, and 2000; all lines rise with lags and then level off near power equal to 1. The top right graph uses phi theta equal to 0.985 and kappa theta equal to 0.02, and its lines rise with lags up to moderate power around 0.6 to 0.8 before flattening. The bottom left graph, with phi theta equal to 0.8 and kappa theta equal to 0.04, shows lines that peak around lag 5 and then decline as lags increase. The bottom right graph, with phi theta equal to 0.8 and kappa theta equal to 0.02, displays relatively low power values that slowly decrease with lags. A legend beneath the graphs identifies line styles for the simple test and Lagrange Multiplier test at T equal to 500, 1000, and 2000. All data are approximate.
Fig. 4.	**Alt-Text:** Four line graphs comparing empirical power versus lags for full LM test and simple test from Dynamic Scale Tail simulations.

Fig. 4.	**Extended Description:** The figure shows four line graphs arranged in a 2 by 2 grid, each plotting empirical power against the number of lags for the Dynamic Scale Tail model. In every graph, the horizontal axis is labeled lags and uses values 1, 5, 10, 15, 25, 35, and 50. The vertical axis is labeled power and ranges from 0 to 1 with tick marks at 0, 0.2, 0.4, 0.6, 0.8, and 1. Each graph contains multiple solid, dashed, and dotted lines with star markers that distinguish the full Lagrange Multiplier test from the simple test at sample sizes 500, 1000, and 2000. The top left graph represents parameter setting phi theta equals 0.985 and kappa theta equals 0.04 and shows power rising quickly with lags and then flattening. The top right graph uses phi theta equals 0.985 and kappa theta equals 0.02 and shows a similar but slightly lower set of curves. The bottom left graph uses phi theta equals 0.8 and kappa theta equals 0.04 and shows power peaking at lag 5 or 10 and then decreasing as lags increase. The bottom right graph uses phi theta equals 0.8 and kappa theta equals 0.02 and shows the lowest overall power, with all curves rising from lag 1 to around lag 5 and then slowly declining. A legend below the grid identifies each line style and marker combination with the corresponding test type and sample size. All data are approximate.
Fig. 5.	**Alt-Text:** Two line graphs comparing empirical test size versus lags for Dynamic Scale Tail simulations at 2 and 8 scale levels, with multiple series.
Fig. 5.	**Extended Description:** The figure shows two line graphs placed side by side that summarize simulation results for empirical test size in a Dynamic Scale Tail model. Both graphs share the same horizontal and vertical axis labels. The horizontal axis is labeled Lags and shows tick marks at 1, 5, 10, 15, 25, 35, and 50. The vertical axis is labeled Size and runs from 0 to 0.25 with tick marks at 0, 0.05, 0.1, 0.15, 0.2, and 0.25. Each graph contains several overlaid lines that represent different test procedures from the simulation study, distinguished by dashed, solid, and dotted styles with point markers. The left graph is titled exponential (omega subscript eta) equals sign 2. In this graph, all series start with empirical size values between about 0.04 and 0.09 at lag 1 and remain roughly flat or slightly decline as lags increase to 50. The right graph is titled exponential (omega subscript eta) equals sign 8. In this graph, some series begin near 0.02 to 0.05 at lag 1 and rise steadily toward about 0.1 to 0.18 by lag 50, while others increase more mildly and then level off or slightly decrease. All data are approximate.
Fig. 6.	**Alt-Text:** Four line graphs comparing power versus c for LM, Q, GAS LM and Nyblom tests under exp omega 3 and T values 500 and 1000.
Fig. 6.	**Extended Description:** The figure shows four line graphs arranged in a 2 by 2 grid that compare the power of several tail index tests. Each graph has horizontal axis label c and vertical axis label Power. The vertical scale runs from 0 to 1 in steps of 0.1. The horizontal scale runs from 0 to 25 in steps of 1, with tick marks at 0, 5, 10, 15, 20 and 25. Each graph corresponds to a different combination of exp omega 3, taking values 2 or 8, and time series length T, taking values 500 or 1000, written as titles above the graphs. Within every graph, multiple lines with different point markers represent the simple Q test with 1 lag, the simple Q test with 20 lags, the simple Q test with data driven lags, the full Lagrange Multiplier test with 1 lag, the full Lagrange Multiplier test with 20 lags, the full Lagrange Multiplier test with data driven lags, the Generalized Autoregressive Score Lagrange Multiplier test with 1 lag, the Generalized Autoregressive Score Lagrange Multiplier test with 20 lags, the Generalized Autoregressive Score Lagrange Multiplier test with data driven lags, and the Nyblom test. For each combination of exp omega 3 and T, the curves rise quickly from low power near c equal to 0, then flatten and often decline slowly as c increases. All data are approximate.
Fig. 7.	**Alt-Text:** Four line graphs comparing power versus c for the simple Q test under different values of exp omega g and time series length T.
Fig. 7.	**Extended Description:** The figure shows four line graphs arranged in a 2 by 2 grid that compare the power of the simple Q test for the upper tail. Each graph has horizontal axis label c, ranging from 0 to 25 at steps of 5, and vertical axis label Power, ranging from 0 to 1 at steps of 0.2. Within every graph three red curves appear: a solid line with circle markers labeled Q star (1), a dashed line with + markers labeled Q star (20), and a dotted line with star markers labeled Q star (star). The top left graph is titled exp (omega g) equals 2, T equals 500. In this graph, all three curves rise quickly from near 0 power at c near 0 to peak values around c between 3 and 6, then gradually decline as c increases. The top right graph is titled exp (omega g) equals 2, T equals 1000 and shows the same three curves with generally higher peak power than in the top left, again rising then slowly decreasing with c. The bottom left graph is titled exp (omega g) equals 8, T equals 500 and shows lower power levels overall, with curves that increase at small c and then slowly decline. The bottom right graph is titled exp (omega g) equals 8, T equals 1000 and shows moderate power that increases and then flattens or slightly decreases as c grows. All data are approximate.

Fig. 8.	**Alt-Text:** Four line graphs comparing power of the simple Q test across c values for different exp of omega theta and T settings in the upper tail.
Fig. 8.	**Extended Description:** The figure shows four line graphs arranged in a 2 by 2 grid that illustrate simulation power results for the upper tail of a simple Q test. Every graph has horizontal axis label c with values from 0 to 25 marked at 5 unit intervals, and vertical axis label Power with values from 0 to 1 marked at 0.2 intervals. Each graph contains three red curves representing Q star of 1 lag with circles, Q star of 20 lags with plus signs, and Q star of a data driven number of lags with asterisk markers. The top left graph title reads exponential of omega theta equals 2 comma T equals 500 and shows all three curves rising quickly between c equals 0 and about c equals 3, then flattening and gradually declining as c increases. The top right graph title reads exponential of omega theta equals 2 comma T equals 1000 and shows the same pattern but with higher peak power levels for all three curves. The bottom left graph title reads exponential of omega theta equals 8 comma T equals 500 and shows lower overall power, again peaking around c near 3 and then slowly decreasing. The bottom right graph title reads exponential of omega theta equals 8 comma T equals 1000 and shows somewhat higher power than the corresponding T equals 500 case, with similar hump shaped curves. A shared legend below identifies the three Q star specifications by their markers. All data are approximate.
Fig. 9.	**Alt-Text:** Two line graphs showing fitted degrees of freedom for Dow Jones Index returns from 1987 to 2016, fluctuating around constant upper levels.
Fig. 9.	**Extended Description:** The figure shows two line graphs of fitted degrees of freedom for Dow Jones Index returns over time. The upper graph shows the symmetric fitted parameter. The horizontal axis lists dates from 01 forward slash 02 forward slash 1987 on the left to 29 forward slash 04 forward slash 2016 on the right, with ticks around every 2 to 3 years. The vertical axis runs from 0 to 8 with ticks at 1 unit intervals. The line forms a dense jagged band mostly between about 5 and 8, with many narrow downward spikes that occasionally drop below 2. One deep spike appears near November 1989, matching the Black Friday market crash mentioned in the text. The lower graph shows the parameter for the lower tail. It uses the same horizontal timeline from 1987 to 2016. Its vertical axis runs from 0 to 7 with ticks at 1 unit intervals. The lower tail series forms a compact band just above 6, again with frequent thin downward spikes, some reaching near 1. Fluctuations in the lower graph resemble those in the upper graph but appear slightly more concentrated near 6. Each graph includes a small legend box naming the fitted parameter. All data are approximate.
Fig. 10.	**Alt-Text:** A line graph of Dow Jones lower tail scale versus inverse degrees of freedom, with peaks around Black Monday, Black Friday, and later crises.
Fig. 10.	**Extended Description:** The figure shows a single line graph summarizing lower tail behavior of Dow Jones Index returns over time, with annotated market events. The horizontal axis shows calendar dates from 01/02/1987 on the left to 29/04/2016 on the right. The left vertical axis shows fitted scale values from about 0.10 to about 1.55 with ticks near 0.39, 0.68, 0.97, 1.26, and 1.55. The right vertical axis shows inverse degrees of freedom values from about minus 0.01 to about 0.04 with ticks near 0, 0.01, 0.02, 0.03, and 0.04. A legend in the upper right identifies a solid blue line for the inverse of the tail parameter, a dashed black line for one standard deviation confidence bounds, and a solid red line for the fitted time varying scale. The blue line forms spikes at dates aligned with vertical tick marks, while the red line fluctuates with several sharp peaks and quieter stretches. Rectangular labels with arrows mark events including Black Monday, Black Friday, Japanese Asset Price Bubble, Credit Card Rates Crash November 1991, Dissolution of Soviet Union, Federal Reserve Rates Rise February 1994, Dot Com Bubble, 9/11, End of Iraq War, First Subprime Crisis Crash with Greenspan warning recession February 2007, 97 Mini Crash due to Asian Economic Crisis, European Sovereign Crisis in Spain and Italy Summer 2011, buildup around Lehman Brothers bankruptcy, and Chinese Black Monday. All data are approximate.
Fig. 11.	**Alt-Text:** A line graph of estimated degrees of freedom for the 5 year Italy credit default swap from 2007 to 2018, showing spikes above 7.

Fig. 11.	**Extended Description:** The figure shows a line graph of fitted estimated degrees of freedom for the symmetric tail index of the 5 year Italy credit default swap series. The horizontal axis lists dates from 01 forward slash 08 forward slash 2007 at the left to 28 forward slash 02 forward slash 2018 at the right, with tick marks roughly every 1 to 2 years. The vertical axis shows degrees of freedom values from 0 to 8 with tick marks at 1 unit intervals. A horizontal reference line runs at value 1. The plotted series starts near value 2 in 2007, dips slightly below 1 in 2008, then rises sharply through 2009 to a local peak near 6. After falling back toward 2, the series fluctuates between about 2 and 5 from 2010 to 2012, with several pronounced swings. Later, the line climbs to peaks above 7 around 2014 and again near 2016, separated by drops toward 2. Toward the end of the sample, values oscillate mostly between about 2 and 5. A legend in the upper right labels the line as the one step ahead filtered estimate of the tail index parameter. All data are approximate.
Fig. 12.	**Alt-Text:** Two line graphs of dynamic tail index estimates for 5 year Italy credit default swaps from 01 slash 08 slash 2007 to 28 slash 02 slash 2018.
Fig. 12.	**Extended Description:** The figure shows two line graphs of tail index estimates for the 5 year Italy credit default swaps over time. The upper line graph has the horizontal axis labeled with calendar dates from 01 slash 08 slash 2007 to 28 slash 02 slash 2018. The vertical axis ranges from 0 to 7 with tick marks at 1 unit intervals. Two lines are plotted, one for the upper tail estimate and one for the lower tail estimate. Both series start near 2 in 2007, dip below 2 around late 2007 and early 2008, and then gradually rise with several abrupt spikes between 2009 and 2012. From 2012 to 2016 the lines fluctuate mostly between 2 and 4, with the upper tail line reaching a peak slightly above 6 around 2013 to 2014 before declining and stabilizing near 2 to 3 toward 2018. A horizontal reference line is drawn at 1. The lower line graph shows the difference between the upper and lower tail index estimates. Its horizontal axis repeats the same date range as the top graph. The vertical axis ranges from negative 2 to 5 with tick marks at 1 unit intervals and a bold horizontal reference line at 0. The difference series oscillates close to 0 from 2007 to 2009, then shows positive spikes up to about 2 around 2010, several alternating positive and negative swings from 2011 to 2013, a prominent positive peak near 4 around 2013 to 2014, and smaller fluctuations around 0 through 2018. All data are approximate.
Fig. 13.	**Alt-Text:** A line graph of Italian 5 year credit default swap returns with 0.5 percent upper and lower quantiles from GARCH and Dynamic Tail DCS models.
Fig. 13.	**Extended Description:** The figure shows a single line graph comparing Italian 5 year credit default swap returns with modeled upper and lower 0.5 percent quantiles. The horizontal axis shows calendar dates from 01/08/2007 on the left to 28/02/2018 on the right, marked at irregular tick dates roughly every 1 to 2 years. The vertical axis shows returns from about -2.0 to 2.5 in unspecified units, with tick marks at 0.5 unit intervals. A thin central line traces daily credit default swap returns, clustering around 0 and forming frequent narrow spikes. Two inner lines mark the upper and lower 0.5 percent quantiles from the generalized autoregressive conditional heteroskedasticity model. Two wider outer lines mark the corresponding quantiles from the asymmetric Dynamic Tail score driven model. Around mid July 2008, there is a prominent positive spike in the returns, where the central series briefly jumps from near 0 to above 2.0 before returning toward 0. Similar but smaller bursts appear intermittently through 2012 and later years. Throughout the sample, the Dynamic Tail bands stay farther from 0 than the generalized autoregressive conditional heteroskedasticity bands. All data are approximate.
Fig. 14.	**Alt-Text:** Four line graphs showing time series of Expected Shortfall ratios for 5 year Italian credit default swap returns at 4 probability levels.

Fig. 14.	**Extended Description:** The figure shows 4 line graphs arranged in a 2 by 2 grid, each comparing time series of Expected Shortfall ratios for 5 year Italian credit default swap rate returns. All graphs share the same horizontal axis, labeled with calendar dates from 01 slash 10 slash 2017 on the left to 28 slash 02 slash 2018 on the right, with intermediate ticks near 07 slash 11 slash 2017, 15 slash 12 slash 2017, and 21 slash 01 slash 2018. Each graph contains 3 lines representing different dynamic conditional score based models. In the top left graph, labeled p equals 0.15 above the plotting area, the vertical axis ranges from 1 to 2.5 with tick marks at 1, 1.5, 2, and 2.5. The 3 lines fluctuate between these values, with several short upward spikes near the end of the period. In the top right graph, labeled p equals 0.10, the vertical axis ranges from 1 to 5 with unit tick marks. The 3 lines move between 1 and about 4, with a few isolated jumps above 4. In the bottom left graph, labeled p equals 0.05, the vertical axis ranges from 0 to 20 with tick marks every 5 units. The lines start near 5 to 15, then gradually decline with intermittent spikes that reach around 15. In the bottom right graph, labeled p equals 0.01, the vertical axis ranges from 0 to 40 with tick marks every 10 units. One line starts close to 40 and declines toward 10, while the other 2 lines remain between about 5 and 15 with scattered peaks. All data are approximate.
Fig. 15.	**Alt-Text:** Four line graphs comparing ratios of Expected Shortfall models for 5 year Italian Credit Default Swap returns from 2017 to 2018.
Fig. 15.	**Extended Description:** The figure shows four line graphs arranged in a 2 by 2 grid, each titled with a different probability level: p equals 0.15 at the top left, p equals 0.10 at the top right, p equals 0.05 at the bottom left, and p equals 0.01 at the bottom right. All graphs share the same horizontal axis, labeled with calendar dates from 01 forward slash 10 forward slash 2017 to 28 forward slash 02 forward slash 2018, representing time for 5 year Italian Credit Default Swap rate returns. Each graph displays three lines: a symmetric score driven Beta t Exponential Generalized Autoregressive Conditional Heteroskedasticity ratio series, an asymmetric score driven Beta t Exponential Generalized Autoregressive Conditional Heteroskedasticity ratio series, and an asymmetric dynamic tail score driven Exponential Generalized Autoregressive Conditional Heteroskedasticity ratio series based on a one step ahead Expected Shortfall from a Generalized Autoregressive Conditional Heteroskedasticity model. Vertical axes show ratios starting slightly above 1 and rising to about 5 in the top left, 10 in the top right, 50 in the bottom left, and 120 in the bottom right. In all graphs the three lines fluctuate over time with occasional sharp spikes. All data are approximate.