UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Testing and Modelling Time Series with Time Varying Tails

Dario
Palumbo

Abstract

The occurrence of extreme observations in a time series depends on the heaviness of the tails of its distribution. The paper proposes a dynamic conditional score model (DCS) for modelling dynamic shape parameters that govern the tail index. The model is based on the Generalised t family of conditional distributions. The framework is extended to allow also for the presence of asymmetric tails and therefore the possibility of specifying different dynamics for the left and right tail indices. The Paper examines through simulations both the convergence properties of the model and the implications of the link functions used. In addition the paper introduces and studies the size and power properties of a new Lagrange Multiplier (LM) test based on fitted scores to detect the presence of dynamics in the tail index parameter. The paper also shows that the novel LM test is more effective than existing testing methodologies. The model is fitted to Equity Indices and Credit Default Swaps returns. It is found that the tail index for equities has dynamics driven mainly by either the upper or lower tail depending if leverage is taken or not into account. In the case of Credit Default Swaps the test identifieses very persistent dynamics for both the tails. Finally the implications of dynamic tail indices for the estimated conditional distribution are assessed in terms of conditional distribution forecasting showing that the novel model predicts more accurately expected shortfalls and value-at-risk than existing models.

Reference Details

CWPE	2111
Published	29 January 2021
Updated	8 March 2021
Key Words	Heavy Tailed Distributions, Extreme Events, Score-Driven Models, Tail Index, Lagrange Multiplier Test, Financial Markets
JEL Codes	C12, C18, C51, C52, C46, C58, G12
Website	www.econ.cam.ac.uk/cwpe

Testing and Modelling Time Series with Time Varying Tails*

Dario Palumbo

Department of Economics, Ca' Foscari University of Venice, Fondamenta S. Giobbe, 873, 30100 Cannaregio,Venezia, Italy
Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom

Abstract

The occurrence of extreme observations in a time series depends on the heaviness of the tails of its distribution. The paper proposes a dynamic conditional score model (DCS) for modelling dynamic shape parameters that govern the tail index. The model is based on the Generalised t family of conditional distributions. The framework is extended to allow also for the presence of asymmetric tails and therefore the possibility of specifying different dynamics for the left and right tail indices. The Paper examines through simulations both the convergence properties of the model and the implications of the link functions used. In addition the paper introduces and studies the size and power properties of a new Lagrange Multiplier (LM) test based on fitted scores to detect the presence of dynamics in the tail index parameter. The paper also shows that the novel LM test is more effective than existing testing methodologies. The model is fitted to Equity Indices and Credit Default Swaps returns. It is found that the tail index for equities has dynamics driven mainly by either the upper or lower tail depending if leverage is taken or not into account. In the case of Credit Default Swaps the test identifieses very persistent dynamics for both the tails. Finally the implications of dynamic tail indices for the estimated conditional distribution are assessed in terms of conditional distribution forecasting showing that the novel model predicts more accurately expected shortfalls and value-at-risk than existing models.

Keywords: Heavy Tailed Distributions, Extreme Events, Score-Driven Models, Tail Index, Lagrange Multiplier Test, Financial Markets

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

*I am grateful for fruitful discussions, comments and suggestions made by Andrew Harvey, Enrico Ghiorzi, Davide delle Monache, Daniele Massacci, Neil Shephard and Andrè Lucas. and the participants of the Cambridge Score Driven Workshop. 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 currently funding my Carlo Giannini Research Fellowship in Econometrics.

Email address: dp470@cam.ac.uk (Dario Palumbo)

1. Introduction

The analysis of time series is focused on identifying the time varying features of the underlying data generating process. It has been empirically shown that 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 is still not sufficient to explain how the occurrence of extreme events can vary over time. It is important to accurately take into account the potential variations of the tails' lengths when forecasting probability distributions of financial returns, particularly if this is done for the purpose of minimizing portfolio risks and monitoring the stability of financial markets.

The occurrence of extreme events in financial data is described by the tail risk. The main contribution of this paper is to show how to accurately identify and capture the dynamic variations over time in the tails of time series distributions, which are distinct from scale variations. Moreover, the paper introduces a new dynamic model which is able to separate the dynamics of the upper tail from that of the lower tail.

Given the difficulties in modelling the tails of a distribution, testing for the presence of dynamics before attempting to model them is necessary in order to avoid spurious results. For this reason the paper also introduces a new formal test to detect the presence of tail dynamics.

The concept of tail risk can be decomposed into two elements, the variation over time in the overall heaviness of the tails of the distribution and the relative difference in size between the upper and lower tails, defined as asymmetry. Figures 1 and 2 show the estimated scale, σ, the estimated degrees of freedom, η, from fitting a static symmetric t distribution to the Dow Jones Index returns, and the estimated left and right tail degrees of freedom parameters, η_1 and η_2, from the static asymmetric t distribution (AST) of Zhu and Galbraith (2010)[1]. Estimates are obtained using moving windows with 500 and 1000 observations respectively. If the degrees of freedom exceeds 40 we assume that they approach infinity and the fitted distribution approximates a normal distribution. As expected, the scale varies over time, which is consistent with the findings on volatility clustering of financial data. At the same time the degrees of freedom seems also to be time varying. Moreover, in the asymmetric case, the relative magnitude and variation of the two degree of freedom parameters tend to differ, with periods where the lower tail is heavier than the upper tail and vice versa. On doubling the window size the magnitude of variation in the degrees of freedom decreases but large movements can still be detected.

These variations in the tail index parameters of the two tails, and in their relative asymmetry, implies time variations of the higher moments of the distribution. Various observation driven models have been proposed to model directly higher moments of the conditional distribution of the data, focusing particularly on skewness to describe asymmetry, as in Harvey and Siddique (1999), and kurtosis for the heaviness of

[1] The degrees of freedom parameter η are 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 $L(y)$ 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 it has only $k < \eta$ finite moments.

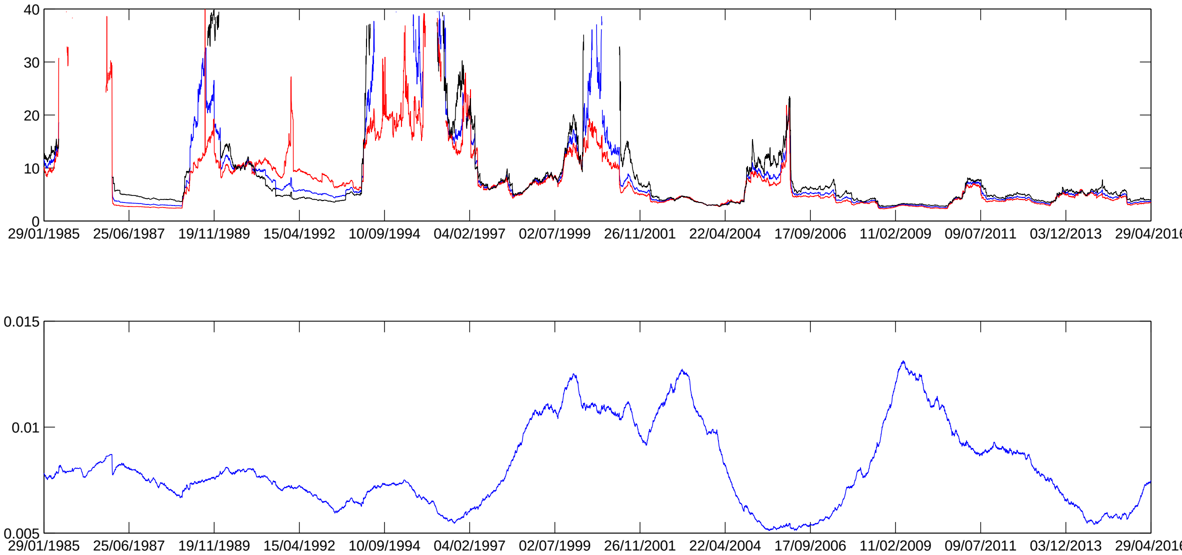

Figure 1: Plot of estimates for static scale (Bottom) and degrees of freedom (Top) with a 500 observations moving window. The top figure shows symmetric degrees of freedom η (Blue), asymmetric left tail degrees of freedom η_1 (Red) and asymmetric right tail degrees of freedom η_2 (Black). The estimated degrees of freedom are only reported if lower than 40.

the tails, as in Brooks et al. (2005). However, as highlighted by Hansen (1994), in order to have valid quasi-Maximum Likelihood properties while modelling conditional moments it is necessary to have tighter restrictions on even higher conditional moments[2]. These conditions can be difficult to be satisfied empirically. Moreover, the moments modelled need always to exist[3]. For these reasons Hansen (1994) suggested that the solution should be to model directly shape parameters of the conditional densities and outlined a general framework to do so using an ARCH type of dynamics.

Another approach for 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. Through this approximation it is possible to focus directly on the distribution of the observations in the tails beyond a given threshold which can be approximated by a Generalised Pareto Distribution or linked to the tail index parameter through a power law. Starting from this theory, Quintos et al. (2001) build 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 bonds futures' returns and U.S. equity returns respectively. In this framework, the occurrence of extreme events can be modelled giving dynamics directly to the tail index parameter, as in Wagner (2005). However, given that the estimation of the parameters of the model depends only on the observations that occur beyond a given threshold, it is necessary to have long time series to describe accurately its dynamics. The problem

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

[3]For example if the variability in the data is too extreme the tail index might be so low up to the point of not being able to guarantee the existence of skewness and kurtosis as well as variance (as for example in the case of a Cauchy distribution).

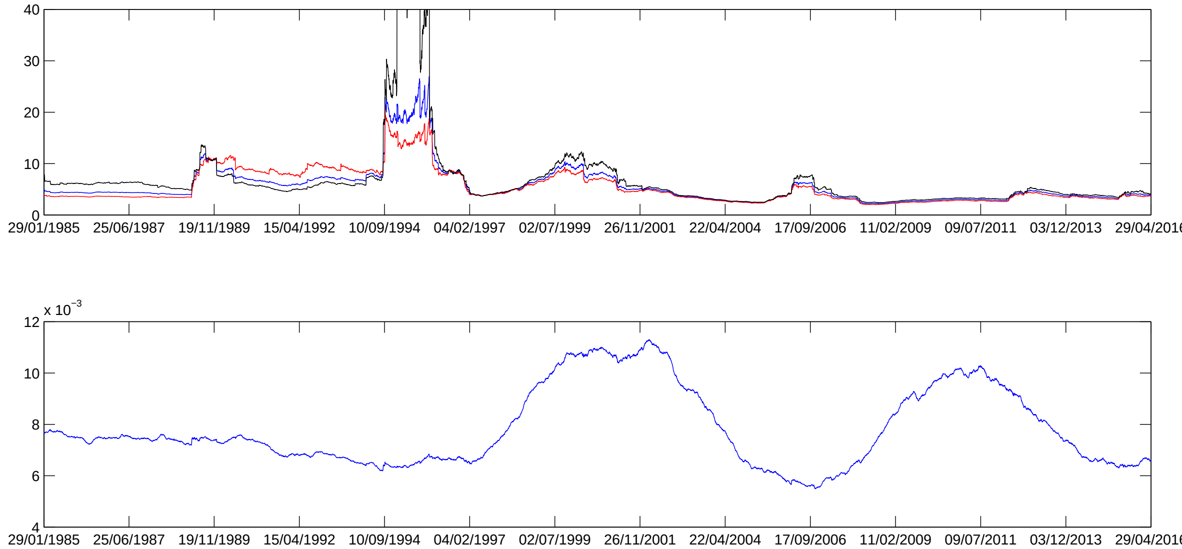

Figure 2: Plot of estimates for static scale (Bottom) and degrees of freedom (Top) with a 500 observations moving window. The top figure shows symmetric degrees of freedom η (Blue), asymmetric left tail degrees of freedom η_1 (Red) and asymmetric right tail degrees of freedom η_2 (Black). The estimated degrees of freedom are only reported if lower than 40.

with this approach is that[4] the parameters governing the dynamics of the tail, as well as other time varying features, might not be stable over such a long time period. To overcome this issue, while looking at the tail risk in equity indexes, Allen et al. (2012), Kelly (2014) and Kelly and Jiang (2014) developed a dynamic power law model which focuses instead on both the time series and the cross-sectional dimensions of the available data exploiting the information from all the stocks traded on an index.

To model the tail index, the present paper suggests instead the use of models from the recent score-driven literature developed by Creal et al. (2013) and Harvey (2013). The motivation comes from the fact that score-driven models, besides allowing for a wider choice of conditional distributions for the data, focus on providing a dynamics directly to the parameters of the conditional distribution rather than to their moments. The score that drives the dynamics is a continuous function of the observations with an adaptive response which gives higher weights to observations at the extreme of the distribution than to the ones close to 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 the tail index assuming a strongly persistent time varying behaviour. Blazsek and Monteros (2017) considered a Dynamic Conditional Score (DCS) model for the degrees of freedom of a t distribution fitted to equity returns.

The main issue with all the aforementioned dynamic tail index models is that, to our knowledge, no simulation study has been made on the effectiveness of these models in picking up the true tail index dynamics as well as on the most effective specification for the score update function[5]. Moreover, no specific

[4] As proven for example for the dynamics of the second moments in GARCH-type of models by Lamoureux and Lastrapes (2002) and Engle and Mustafa (1992).

[5] For example weather to standardise or not the score by the information matrix in the dynamic equation.

formal tool has been introduced in a dynamic setting to assess the actual presence of a dynamic tail index and to justify the use of these models.

Building on this literature, the present work focuses on modelling a dynamic tail index in the DCS framework assuming a distribution of the Generalised t family used by Harvey and Lange (2017).The 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. The paper studies empirically the convergence properties of the odel given different average values of the tail index parameter. In addition, a new test is introduced to detect if the tail index parameter is dynamic. The methodology is based on the Lagrange Multiplier (LM) test, which has been 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 of the data under the null of static dynamics. This test differs from the one of Quintos et al. (2001) since it focuses on the conditional distribution of the data. Other types of LM tests for general parameters instabilities and structural breaks in the DCS framework have been considered by Calvori et al. (2017). However, given that our test takes explicitly into account of the cross-correlation between the scores with respect to the scale and tail index parameters under the alternative of being dynamic, the present study shows that overall our LM test has higher power in detecting dynamics of tail index parameters. We also provide a power and size comparison with a simple version of the LM test based on the Box-Ljung test.

The final contribution of the paper is to extended the Generalised t conditional distribution to its skewed asymmetric version to include a different independent time-varying tail parameter for each of the tails. The reason for this is that in the presence of asymmetric data a symmetric model would incorrectly estimate the quantiles of the conditional distribution somewhere in between the two tails, most likely underestimating the thickness of the heavier tail. On the other hand an asymmetric model would more accurately estimate the thickness of each tail separately and this can be used to describe the time variation in the asymmetry of the distribution. The idea of this dynamic asymmetry in a score-driven framework has only been considered previously in two cases: in a static tails framework by Thiele (2020), which models a dynamic scale in presence of an AST distribution of Zhu and Galbraith (2010), and by Massacci (2017) who, following the extreme value theory approach, proposes a time varying tail index model for modelling directly the tails of the conditional distribution of the data assuming they are conditionally Laplace distributed. To our knowledge, the present study is the first work which introduces an adaptive model for modelling the asymmetry of the full conditional distribution of the data through modelling independently its two tail index parameters.

Finally the paper verifies the empirical relevance of both the symmetric and asymmetric specifications in the modelling of market returns of Equity Index and Credit Default Swap (CDS) rates. The analysis shows that the tail movements in the Equity Index are not particularly persistent and are driven only by the movements either of the lower tail or upper tail depending if leverage is taken into account or not. On the other hand for the CDS both the tails are independently time-varying with very persistent movements. The impact of the tail variations on density forecasts is also assessed on these datasets in terms of fitted quantiles and testing the accuracy of both models in predicting Expected Shortfalls.

The paper is structured as follows: Section 2 introduces the theory behind the statistical framework of the model presented. Section 3 presents the theory behind a formal test for detecting time variability of the tail index parameters, which is then analysed through simulations and compared with other relevant tests in the literature. In Section 4 the statistical framework is extended as to introduce asymmetric tails. Section 5 presents the results from fitting the dynamic tail score-driven models to equity index and CDS returns as well as analysing out-of-sample the quantiles of the forecasted conditional distributions in comparison with standard models.

2. Statistical Framework: DCS Dynamic Tail Index Model

The current study is based on the idea of modelling data series assuming dynamic scale and shape parameters through a DCS model with a conditional distribution from the Generalised t distribution's family, as described in Harvey and Lange (2017). The Generalised t distribution is a location and scale general distribution which is described by the following density

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

Where $B(.,.)$ is a beta function, $\varepsilon_t = (y_t - \mu)/\varphi$ are the residuals, η and υ are both shape parameters and η governs the tail index for $\eta > 0$. The Generalised t distribution is a very flexible distribution which can accommodate many sub distributions as special cases according to different values of η and υ. It can have fat tails for $\upsilon > 1$ and heavy but not fat tails for $0 < \upsilon < 1$. For $\upsilon = 2$ it becomes a t distribution with η degrees of freedom. Then for $\eta \to \infty$ it becomes a $GED(\upsilon)$ distribution which then becomes Laplace for $\upsilon = 1$ and normal for $\upsilon = 2$. Harvey and Lange (2017) shows how to model the scale φ in a DCS framework with an exponential link function $\varphi_{t|t-1} = e^{\lambda_{t|t-1}}$ deriving the score and its information matrix with respect to the dynamic scale parameter $\lambda_{t|t-1}$.

$$\frac{\partial \ln f_t}{\partial \lambda} = (\eta + 1) b_t - 1, \qquad \mathcal{I}_{\lambda\lambda} = \frac{\eta\upsilon}{\upsilon + \eta + 1}, \qquad t = 1, \ldots, T$$

where $b_t = \frac{|\varepsilon_t|^{\upsilon}/\eta}{|\varepsilon_t|^{\upsilon}/\eta + 1}$ is distributed as a $Beta\left(\frac{1}{\upsilon}, \frac{\eta}{\upsilon}\right)$. The dynamics of the scale parameter $\lambda_{t|t-1}$ is then described by $u_t^{\lambda} = \frac{\partial \ln f_t}{\partial \lambda}\mathcal{I}_{\lambda\lambda}^{-1}$. Their paper provides the asymptotic normality results for the estimators. This model is then easily extendible to include a dynamic location parameter[6].

In the present study dynamics for both the conditional scale, $\varphi_{t|t-1}$, and the tail index parameter, $\eta_{t|t-1}$, are assumed. As with the scale parameter, to restrict $\eta_{t|t-1}$ to be strictly positive it is possible to model it using an affine exponential link function of the form $\eta = \eta^{\dagger} + e^{\eta_s \vartheta}$, where $\eta^{\dagger}$ is, as in Lucas and Zhang (2016), a lower-bound for the tail index parameter. This can be used to restrict the parameter to be greater

[6] Harvey (2013) described extensively how to set up a DCS model with dynamic Location and Scale parameters when the conditional distribution allows the two parameters to be independently specified.

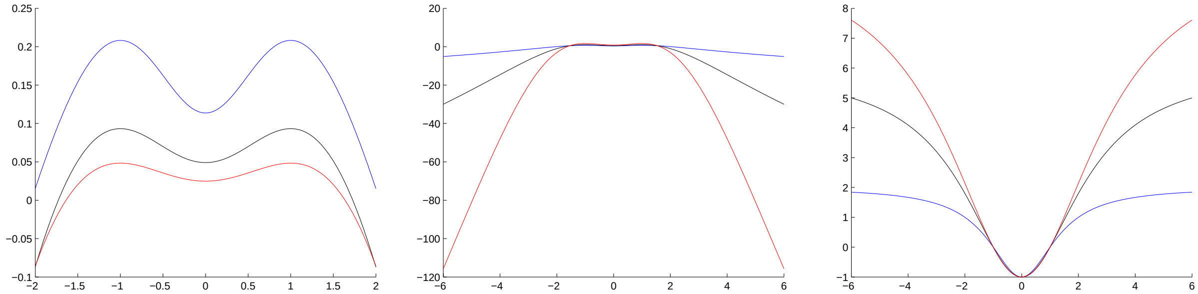

Figure 3: The figure provide the plot of the raw score with respect to ϑ (Left and Middle) and with respect to λ (Right) against different residuals values ε_t, for $\upsilon = 2$ and $\eta = 2$ (Blue), $\eta = 6$ (Black), $\eta = 10$ (Red).

than two for example and guarantee the existence of the variance of the conditional distribution. η_s is a fixed parameter that allows us to either model directly η, if $\eta_s = 1$, or its inverse $\bar{\eta} = 1/\eta$, if $\eta_s = -1$[7], which is often usefull to use in the derivation of the analytical tresults of the Generalised t DCS model[8]. Then the conditional score, as well as its information matrix, with respect to ϑ can be obtained as

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

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

where $\psi\left(x\right)$ and $\psi'\left(x\right)$ are the gamma and digamma functions respectively. It is interesting to notice that the score with respect to λ appears in the last term of the score with respect to ϑ. This highlights the close relation between the scale and the tail index parameter.

From Figure 3 it is possible to see that for values of ε_t close to the median the response of the score with respect to ϑ tends to increase as η falls, while the score with respect to λ remain unchanged. On the other hand for large positive and negative values of ε_t the score with respect to ϑ is unbounded and its response increases in magnitude as η increases, while for the score with respect to λ decreases up to the point of becoming bounded for very low values of η. This makes sense, since, as the degrees of freedom increases, the observations very far from the median are more informative of a variation in the behaviour of the tails and depending on how heavy the fitted distribution is at every point in time these observations would be discounted more. Ultimately given that both ε_t and $\eta_{t|t-1}$ vary over time, it is more helpful to consider the score response at every t as a three dimensional function as showed in Figure 4.

As shown by Harvey and Lange (2017), sometimes it is easier to estimate the tail index parameter by estimating its inverse $\overline{\eta}$. However, in modelling the tail index parameter dynamically with our specification, modelling $\overline{\eta}$ means simply giving dynamics to $-\vartheta_{t|t-1}$; the score then becomes negative but ultimately it

[7]This general set up of the link function nests several specifications. For instance if instead we decide to model the inverse of the degrees of freedom, $\overline{\eta}$, with a logistic function which restrict it to be $0 < \overline{\eta} < 1$, like $\bar{\eta} = \frac{\exp 2\vartheta}{1+\exp 2\vartheta}$, then $\eta = 1 + e^{-2\vartheta}$, which is our specification with $\eta^{\dagger} = 1$ and $\eta_s = -2$.

[8]see Harvey and Lange (2017).

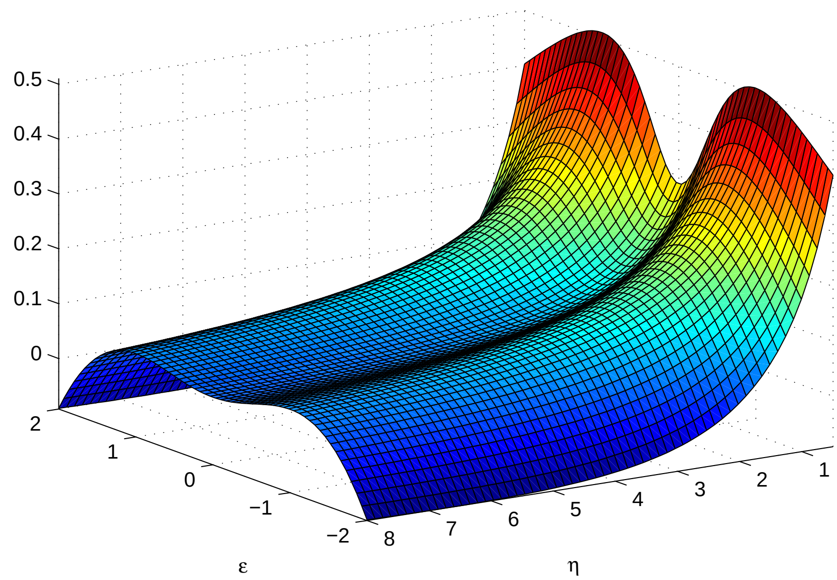

Figure 4: Three dimensional surface of score of the unbounded tail index parameter for $-2 < \varepsilon < 2$, and $1/2 < \eta < 8$.

make no difference in the estimation of the magnitude of the dynamic parameters of $\vartheta_{t|t-1}$. Then the Dynamic Scale-Tail model can be described in the following way,

$$y_t = \mu + \varepsilon_t \exp\left(\lambda_{t|t-1}\right), \qquad \varepsilon_t \mid \mathcal{F}_{t-1} \overset{iid}{\sim} \text{Gen-}t\left(\eta_{t|t-1}; \upsilon\right), \qquad t = 1, \ldots, T \tag{3}$$

A first order DCS model for dynamic scale and tail index can be described by,

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

where $u_t^\vartheta = \frac{\partial \ln f_t}{\partial \vartheta} \mathcal{I}_{\vartheta\vartheta}^{-1}$ and where the information matrix with respect to the static parameters besides μ is,

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

$$\mathcal{I}_{\upsilon\upsilon} = \eta \frac{\upsilon\left[\frac{(1-\eta)\upsilon}{\eta} - \ln\eta + \psi\left(\frac{\eta}{\upsilon}\right) - \psi\left(\frac{1}{\upsilon}\right)\right]^2 + \left[\upsilon + \eta\left(\eta+\upsilon+1\right)\right]\psi'\left(\frac{\eta}{\upsilon}\right) + \left[\eta + \upsilon\left(\eta+1\right)+1\right]\psi'\left(\frac{1}{\upsilon}\right)}{\upsilon^4\left(\upsilon+\eta+1\right)} -$$

$$- \frac{\left(1+\eta\right)^2}{\eta^4\upsilon^4}\psi'\left(\frac{\eta+1}{\upsilon}\right) - \frac{\upsilon^2\left(1+\eta^2\right)+\eta\left(\eta+\upsilon+1\right)}{\eta\upsilon^2\left(\eta+\upsilon+1\right)}$$

$$\mathcal{I}_{\lambda\vartheta} = \eta_s\left(\eta-\eta^\dagger\right)\left[\frac{1}{\left(\eta+\upsilon+1\right)} - \frac{1}{\left(\eta+1\right)}\right]$$

$$\mathcal{I}_{\lambda\upsilon} = \eta\frac{\psi\left(\frac{\eta}{\upsilon}\right) - \psi\left(\frac{1}{\upsilon}\right) - \ln\eta + \frac{\upsilon(\eta+1)}{\eta}}{\upsilon\left(\eta+\upsilon+1\right)}$$

$$\mathcal{I}_{\vartheta\upsilon} = -\eta_s\left(\eta-\eta^\dagger\right)\eta\left[\frac{\psi\left(\frac{\eta}{\upsilon}\right) - \psi\left(\frac{1}{\upsilon}\right) - \frac{\eta(\upsilon+1)+1}{\eta} - \ln\eta}{\upsilon\left(\eta+1\right)\left(\eta+\upsilon+1\right)} + \frac{\eta\psi'\left(\frac{\eta}{\upsilon}\right) - \left(\eta+1\right)\psi'\left(\frac{\eta+1}{\upsilon}\right)}{\upsilon^3}\right]$$

All its elements are independent of λ.

Given how often the tail index dynamics is bounded in the literature, in appendix Appendix A we made an important analysis of the implications of bounding the tail index parameter by $\eta^\dagger$ on the score response. As a result we have identified that bounding the tail index can imply serious distortions to the score response. These distortions can ultimately affect the fit since they makes harder for the dynamic parameter, once next to the bound, to move away from it. Moreover, noting that since the score function naturally tends to push the dynamic parameter away from very low values, the chances of the parameter actually falling below 1 and staying there are much lower when the tail index is unbounded than when is bounded. Therefore, for our modelling purposes we will then assume for the rest of the paper $\eta^\dagger = 0$ and $\eta_s = 1$.

3. Detecting Time varying Dynamics in Tail Index Parameters

3.1. The LM approach

Testing techniques for detecting dynamics in parameters of a DCS model have been presented for dynamic correlation in Harvey and Thiele (2016). Following from their approach, in the case of a single time varying parameter, let's say ϑ, with dynamics as in Equation (4) driven by its unstandardised conditional score[9]

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

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

A test against the presence of dynamics in an otherwise static model can be based on the Portmanteau statistic

$$Q_u(P) = T\sum_{j=1}^{P} r_u^2(j), \tag{5}$$

[9]For simplicity of exposition we limit ourself in the derivation of the LM test statistic in the case of unstandardised scores. However, while working with standardised scores, if the information matrix with respect to the time varying parameters are only dependent on shape parameters like the tail index (which is the case for the t and Generalised t distributions), under the null of static tail index they are fixed scalars. For this reason we could rewrite the "update" part of Equation (4) as $\kappa_\vartheta \hat{u}_t^\vartheta = \kappa_\vartheta \mathcal{I}_{\vartheta\vartheta}^{-1} \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}} = \hat{\kappa}_\vartheta u_t^\vartheta$. In our case this applies to both scale and tail index.

where $r_u(j)$ is the j-th sample autocorrelation of u_t^ϑ. The Box-Ljung modification,

$$Q_u^*(P) = T(T+2)\sum_{j=1}^{P}(T-j)^{-1}r_u^2(j),$$

may also be used. The asymptotic distribution of both statistics under the null hypothesis is χ_P^2.

Remark 1. *Rather than fixing P, it may be selected using a consistent information criterion, as in Escanciano and Lobato (2009). Under the null hypothesis, only the first lag is selected in large samples with probability one. As a result, the asymptotic distribution under the null hypothesis is χ_1^2.*

Since Equation (4) is not identifiable under the null hypothesis $\phi_\vartheta = \kappa_\vartheta = 0$, the Portmanteau test may be derived as a Lagrange Multiplier (LM) test under the null hypothesis that $\kappa_{\vartheta 0} = \kappa_{\vartheta 1} = ... = \kappa_{\vartheta P-1} = 0$, against the alternative $\kappa_{\vartheta i} \neq 0, i = 0, ..., P-1$, in its Q-MA approximate representation

$$\vartheta_{t+1\mid t} = \omega_\vartheta + \kappa_{\vartheta 0}u_t^\vartheta + ... + \kappa_{\vartheta P-1}u_{t+1-P}^\vartheta, \qquad\qquad t = 1, ..., T.$$

defined as

$$LM_u(P) = \frac{1}{T}\begin{bmatrix}\mathbf{0}' & \partial \ln L/\partial \boldsymbol{\kappa}_\vartheta'\end{bmatrix}\begin{bmatrix}\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} & \boldsymbol{\Psi}_{\boldsymbol{\theta\kappa}}\\ \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} & \boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\ \partial \ln L/\partial \boldsymbol{\kappa}_\vartheta\end{bmatrix}, \tag{6}$$

where $\boldsymbol{\kappa} = (\kappa_{\vartheta 0}, \kappa_{\vartheta 1}, ..., \kappa_{\vartheta P-1})$ and $\boldsymbol{\theta}$ is the vector of all the other fixed parameters, which in this case consist only of ω_ϑ. Under this conditions it is shown by Harvey and Thiele (2016) that

$$LM_u(P) = \frac{1}{T}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta'}\boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}^{-1}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta} = T\sum_{j=1}^{P} r_u^2(j) \tag{7}$$

when the process is very persistent, that is when in Equation (4) the dynamic parameter ϕ_ϑ is close to one, larger values of P may yield more powerful tests. Another possibility suggested by Harvey (2013) is to use 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 ends up being based on the same scores as in the Portmanteau test. It can be written as

$$N = \frac{1}{T^2\sigma_{\vartheta u}^2}\sum_{j=1}^{T}\left(\sum_{k=j}^{T} u_k^\vartheta\right)^2.$$

Under the null hypothesis of parameter constancy, N has a Cramer-von Mises distribution. Although the Nyblom test is usually regarded as a test against a random walk alternative, it can also be interpreted as a test against a very persistent, but stationary, alternative[10].

However the LM test statistic simplifies to Equation (7) 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

[10] See, for example, Harvey and Streibel (1998) and Harvey and Thiele (2016).

conditional distribution of the data. When we have fitted a DCS model to the data for a time varying parameter, let's say $\lambda_{t|t-1}$ through a Beta-t-EGARCH model[11], the LM test statistic for detecting dynamics in another parameter, like the tail index $\eta_{t|t-1}$ is[12]

$$LM_u(P) = \frac{1}{T}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}'_\vartheta}\boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta} + \frac{1}{T}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}'_\vartheta}\left[\boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}}\boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}}\left(\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} - \boldsymbol{\Psi}'_{\boldsymbol{\kappa\theta}}\boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}}\boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}}\right)^{-1}\boldsymbol{\Psi}'_{\boldsymbol{\kappa\theta}}\boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa\kappa}}\right]\frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta}, \tag{8}$$

This result leads to the following

Proposition 1. *If the data generating process y_t is*

$$y_t \mid \mathcal{F}_{t-1} \overset{iid}{\sim} \textit{Gen-t}\left(\varphi_{t|t-1}, \eta_{t|t-1}; \upsilon\right), \qquad t = 1, \ldots, T$$

and the dynamic scale $\varphi_{t|t-1}$ is fitted by a Beta-Gen t-EGARCH model[13]*, the Lagrange Multiplier test for the dynamics of $\eta_{t|t-1} = \eta^\dagger + e^{\eta_s \vartheta_{t|t-1}}$ under the null of $\kappa_{\vartheta 0} = \kappa_{\vartheta 1} = ... = \kappa_{\vartheta P-1} = 0$, against the alternative $\kappa_{\vartheta i} \neq 0, i = 0, ..., P-1$, in the dynamic model*

$$\vartheta_{t+1|t} = \omega_\vartheta + \kappa_{\vartheta 0} u_t^\vartheta + ... + \kappa_{\vartheta P-1} u_{t+1-P}^\vartheta, \qquad t = 1, ..., T.,$$

takes the form

$$LM_u(P) = Q_u(P) + T\mathcal{I}^2_{\lambda\vartheta}\mathbf{g}'\left(\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} - \frac{\mathcal{I}^2_{\lambda\vartheta}}{\mathcal{I}^2_{\vartheta\vartheta}}\frac{1-a^{2P}}{1-a^2}\mathbf{g}\mathbf{g}'\right)^{-1}\mathbf{g}\left(\sum_{j=1}^{P} r_{\vartheta u}\left(j\right)a^{j-1}\right)^2 \tag{9}$$

where $Q_u(P)$ is the standard Portmanteau statistic, $r_{\vartheta u}(j)$ are the sample autocorrelations of the fitted scores with respect to ϑ under the null, $\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}}$ is the portion of the dynamic information matrix of the joint model related to the other estimated static parameters $\boldsymbol{\theta} = (\upsilon, \omega_\lambda, \phi_\lambda, \kappa_\lambda, \omega_\vartheta)'$ as described in the appendix. $\mathbf{g} = \left(\kappa_\lambda\left(h_\upsilon - \frac{\kappa_\lambda h_\lambda}{1-a}\mathcal{I}_{\lambda\upsilon}\right), \kappa_\lambda h_\lambda \frac{1-\phi_\lambda}{1-a}, 0, \mathcal{I}_{\lambda\vartheta}, \kappa_\lambda\left(h_\vartheta - \frac{\kappa_\lambda h_\lambda}{1-a}\mathcal{I}_{\lambda\vartheta}\right)\right)'$. $\mathcal{I}_{\lambda\vartheta}$ and $\mathcal{I}_{\vartheta\vartheta}$ are elements of the information matrix of the static model. $a = \phi_\lambda - \kappa_\lambda\mathcal{I}_{\lambda\lambda}$, $h_\upsilon = E\left[u_t^\vartheta\frac{\partial u_t^\lambda}{\partial \upsilon}\right]$, $h_\lambda = E\left[u_t^\vartheta\frac{\partial u_t^\lambda}{\partial \lambda}\right]$ and $h_\vartheta = E\left[u_t^\vartheta\frac{\partial u_t^\lambda}{\partial \vartheta}\right]$. Under the null hypothesis the test is distributed with a Chi-Square asymptotic distribution with P degrees of freedom.

Remark 2. *If the shape parameter υ is not estimated*[14]*, the LM test statistics can be computed in the same way as in Equation (9) just removing from the block matrix $\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}}$ and the vector $\mathbf{g}$ the row and column related to υ.*

In the following sections we will investigate the performance of both the simple $Q_u(P)$ test and the $LM_u(P)$.

[11] which is a DCS model for dynamic scale which assumes t as conditional distribution as described in Harvey (2013).
[12] For details see Harvey and Thiele (2016)
[13] Which is the DCS model for scale that assumes a Generalised t as a conditional distribution, as introduced by Harvey and Lange (2017).
[14] Or fixed to $\upsilon = 2$, as in the case of the t Distribution

3.2. Tests Simulation Study

Here the power and size of the tests are assessed under various parameters assumptions. For this purpose we have designed a simulation study on the same lines as the one used to assess the implications of bounding the score in appendix Appendix A: we have generated $N = 1000$ samples of length $T = 500,\ 1,000,\ 2,000$ assuming that the data generating process is conditionally distributed with a t distribution with dynamic scale, $\varphi_{t|t-1} = \exp(\lambda_{t|t-1})$, and dynamic degrees of freedom, $\eta_{t|t-1} = \exp(\vartheta_{t|t-1})$. The dynamics of the two parameters are modelled using an exponential link function in a DCS framework with dynamics as described in Equation (4) with $\omega_\lambda = -4.7$, $\phi_\lambda = 0.985$, $\kappa_\lambda = 0.03$, and ω_ϑ, ϕ_ϑ and κ_ϑ, adjusted in each simulation to prevent the tail index to explode towards infinity. The exact specifications are described in Figures G.14 and G.15. In particular we have $\omega_\vartheta = \log 2, \log 8, \log 15$ and $\log 30$. For the size of the test we repeat the simulations with the same dynamic parameters just assuming that $\phi_\vartheta = \kappa_\vartheta = 0$.

To perform the test, we first fit the Beta-t-EGARCH model, which is a DCS model for time varying scale that assumes that the data are conditionally t-distributed, therefore a Generalised t with $\upsilon = 2$[15]. From this we have obtained the fitted scale, $\hat{\varphi}_{t|t-1}$, and the fitted residuals, as $x_t = y_t e^{-\hat{\lambda}_{t|t-1}}$. Then we use the estimated $\hat{\eta}$ parameter to compute the scores with respect to ϑ under the null hypothesis of no dynamics, as

$$\hat{u}_t^{\vartheta\dagger} = \frac{\hat{\eta}}{2}\left[\psi\left(\frac{\hat{\eta}+1}{2}\right) - \psi\left(\frac{\hat{\eta}}{2}\right) - \frac{1}{\hat{\eta}} + \ln\left(1-\hat{b}_t\right) + \frac{(1+\hat{\eta})}{\hat{\eta}}\hat{b}_t\right], \quad t = 1, ..., T \tag{10}$$

and then construct the full LM test statistic for various P.

We also compare the results from the simple Portmanteau test on the fitted scores, $Q_u^*(P)$, which is referred to as the simple LM test statistic, here presented in its more robust Box-Ljung version. To do so we instead fit a static t distribution to the x_t and then we use the new estimated degrees of freedom $\hat{\eta}^*$ to compute the $\hat{u}_t^{\vartheta *}$ and then the simple $Q_u^*(P)$.

In Figures G.14 and G.15 we can see that, as expected, the power of the test tends to decrease as the sample size T decreases while the implied size tends to increase slightly. Since ω_ϑ is the unconditional mean of $\vartheta_{t|t-1}$ we can see that, as the true unconditional mean of the time varying degrees of freedom $\eta_0 = \exp(\omega_\vartheta)$ increases, the power of the test tends to decrease while its size tends to increase. The highest power is when η_0 is close to 2. 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 increases beyond 10 it is difficult to estimate the exact values of η which would maximise the likelihood. This can also be noticed in the results in Tables H.3 to H.5 where it is possible to see that as ω_ϑ increases the standard errors of the estimates of the dynamic parameters of $\vartheta_{t|t-1}$ also increase, making the estimates less accurate. Overall we can see that the model is quite reliable in estimating the correct dynamics of $\eta_{t|t-1}$ when $\omega_\vartheta \leq \log 15$.

We can also notice that fitting the dynamic tail also helps the fit of the scale. This can be seen from the results of the Box-Ljung test in the same tables which shows how the residual correlation in the fitted score with respect to the scale parameter, $\hat{u}_t^\lambda$, tends to disappear when fitting the joint model, in particular for

[15] See Harvey (2013).

true values of $\omega_\vartheta \leq \log 8$.

Comparing the two tests, overall the simple Box-Ljung test has slightly more power than the full LM test and a lower size, with the difference becoming more apparent as we decrease the number of observations and increase the lags P. This is because the constant $\mathcal{I}^2_{\lambda\vartheta}\mathbf{g}'\left(\mathbf{\Psi}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \frac{\mathcal{I}^2_{\lambda\vartheta}}{\mathcal{I}^2_{\vartheta\vartheta}}\frac{1-a^{2P}}{1-a^2}\mathbf{g}\mathbf{g}'\right)^{-1}\mathbf{g}$ in Equation (9) is estimated in all the cases as negative, making the full LM test more conservative than the simple Box-Ljung. However the magnitude of the constant is often close to zero, between 10^{-3} and 10^{-5}, falling rapidly as the number of lags increase. In any case, given such a small difference in power, the gap between the results of the two tests is expected to disappear if we were to correct the LM test for the size[16].

Given these results we can assert that the simple Box-Ljung test is as powerful at detecting dynamics in the tail index parameter as the full LM test, and is even more accurate in presence of smaller sample sizes T.

3.3. Test Comparison

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

The empirical power of the GAS-LM test is compared against the power of the full LM and the simple Q^* test assuming the all the tests are performed both with $P = 1$ and P^* chosen by the automatic algorithm of Escanciano and Lobato (2009). Finally we include also the Nyblom test as a benchmark, since is often also seen as a general test for parameter instability.

In order to do so we have performed a series of simulations of the same model used in Section 3.2 with the same dynamic specification for the parameter $\lambda_{t|t-1}$. For the dynamics of the tail index parameter, $\vartheta_{t|t-1}$, we used two values for its unconditional mean $\omega_\vartheta = \log 2, \log 8$. For other dynamic parameters we have used $\kappa_\vartheta = c/(5T)$ for $\omega_\vartheta = \log 8$ and $\kappa_\vartheta = c/(2.5T)$ for $\omega_\vartheta = \log 2$ while $\phi_\vartheta = \sqrt{1-\kappa_\vartheta}$. Then c is left to vary in between the range $[1, \cdots, 21]$. Under this specification we can assess the performance of the tests under various assumptions of persistence for the dynamic tail index parameter while making sure that the simulated parameter doesn't explode to infinity. For each specification we perform $N = 1,000$ simulations under both $T = 500,\ 1,000$.

From the results of the tests in Figure G.18 we can see that, as showed previously, the performances of the full LM and simple Q^* tests are very similar. The GAS-LM(1) test tends to fail to capture the presence of a dynamic tail index parameter in all the cases while, on the other hand the GAS-LM(*) becomes quite competitive in most of the cases. In particular it has the highest power, for both values of ω_ϑ and sample size T, when c is low and therefore the tail index has a very persistent dynamics. The full LM, Q^* and

[16] Moreover the LM is only asymptotically and locally more powerful than other tests.

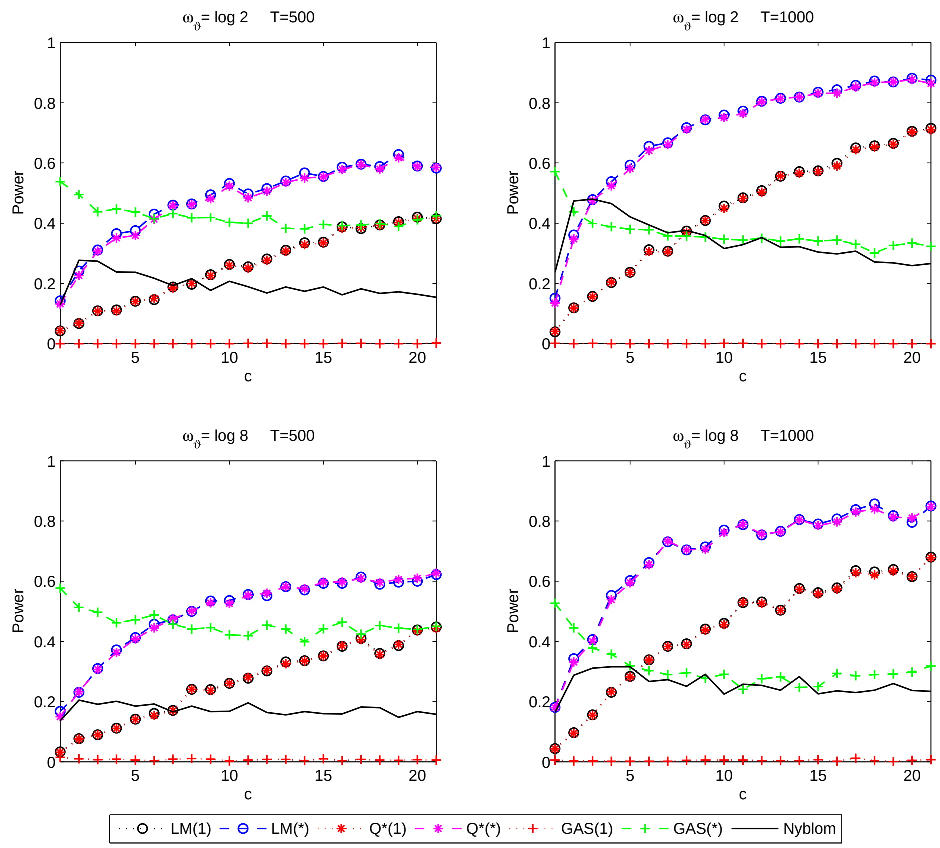

Figure 5: Power comparison under simulation of the full LM test, the simple LM test, Q, the GAS-LM test of Calvori et al. (2017) and the Nyblom for either 1 lag or a number of lags defined by the criterion of Escanciano and Lobato (2009), $(*)$. The comparisons are performed under different assumptions 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 $T = 1,000$.

Nyblom have a relative poor performance for low c but tend to pick up quite rapidly. In particular the power of the full LM(*) and $Q^*(*)$ tends to be in general higher than the power of the GAS-LM(*), particularly with $T = 1,000$. The power of the LM(1) and $Q^*(1)$ is never higher than the power of the GAS-LM(*) for $T = 500$, while for $T = 1,000$ for both the tests is significantly higher except for some small values of c. The performance of the Nyblom test is almost never better than the one of the LM-GAS(*) test. The power of the Nybloom test is much worse when $T = 500$, while for $T = 1,000$ tends to be more or less the same as th one of the LM-GAS(*) test for most of the values of c.

In general we can say that the GAS-LM(*) is a good alternative when the underlying dynamics of the tail index parameter is very persistent and we are in presence of a small sample size. On the other hand, in the majority of the cases the LM(*) and $Q^*(*)$ are better at detecting dynamics in tail index parameters.

4. Extending the Statistical Framework: Asymmetric Tails Modelling

As in Harvey and Lange (2017), given a model as in Section 2, skewness in the Generalised t distribution can be easily introduced by defining negative and positive residuals as follows,

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

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

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

Each η_i and υ_i governs the shape for the left and right side of the distribution. $K_{12} = 1/\left[\alpha/K_1 + (1-\alpha)/K_2\right]$, with $K_i = K(\eta_i, \upsilon_i)$ for $i = 1, 2$. The distribution then is symmetric if $\eta_1 = \eta_2$ as well as $\upsilon_1 = \upsilon_2$. If the distribution is asymmetric the score is more complex and it is different for the left and right tail, as well as the corresponding information matrices. For the dynamic scale parameter

$$\frac{\partial \ln f_t}{\partial \lambda} = \begin{cases} (1+\eta_1)\, b_{1t} - 1 & ; y_t \leq \mu \\ (1+\eta_2)\, b_{2t} - 1 & ; y_t > \mu \end{cases}, \qquad \mathcal{I}_{\lambda\lambda} = \begin{cases} \frac{\eta_1\upsilon_1}{\upsilon_1+\eta_1+1} & ; y_t \leq \mu \\ \frac{\eta_2\upsilon_2}{\upsilon_2+\eta_2+1} & ; y_t > \mu \end{cases}$$

where $b_{1t} = \frac{\left|\varepsilon_t^-\right|^{\upsilon_1}/\eta_{1t|t-1}}{\left|\varepsilon_t^-\right|^{\upsilon_1}/\eta_{1t|t-1}+1}$ and $b_{2t} = \frac{\left|\varepsilon_t^+\right|^{\upsilon_2}/\eta_{2t|t-1}}{\left|\varepsilon_t^+\right|^{\upsilon_2}/\eta_{2t|t-1}+1}$. then we have that

$$u_t^\lambda = \left[\left(1+\eta_{1t|t-1}\right) b_{1t} - 1\right] \frac{\upsilon_1+\eta_{1t|t-1}+1}{\eta_{1t|t-1}\upsilon_1} \mathbf{1}_{(\varepsilon_t \leq 0)} + \left[\left(1+\eta_{2t|t-1}\right) b_{2t} - 1\right] \frac{\upsilon_2+\eta_{2t|t-1}+1}{\eta_{2t|t-1}\upsilon_2} \left(1-\mathbf{1}_{(\varepsilon_t \leq 0)}\right)$$

With only the scale parameter as dynamic we have the model of Harvey and Lange (2017) and with $\upsilon = 2$ we have the AST DCS model of Thiele (2020). Now we can introduce dynamics to the tail index parameters through the conditional scores

$$\frac{\partial \ln f_t}{\partial \vartheta_1} = \begin{cases} \widetilde{\eta}_1 \frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\upsilon_1}\left[\alpha^+\tau_1 + \ln(1-b_{1t}) + \frac{1}{\eta_1}\left(\frac{\partial \ln f_{1t}}{\partial \lambda}+1\right)\right] & ; y_t \leq \mu \\ \widetilde{\eta}_1 \frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\upsilon_1}\alpha^+\tau_1 & ; y_t > \mu \end{cases}, \tag{11}$$

$$\frac{\partial \ln f_t}{\partial \vartheta_2} = \begin{cases} \widetilde{\eta}_2 \frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\upsilon_2}(1-\alpha^+)\,\tau_2 & ; y_t \leq \mu \\ \widetilde{\eta}_2 \frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\upsilon_2}\left[(1-\alpha^+)\,\tau_2 + \ln(1-b_{2t}) + \frac{1}{\eta_2}\left(\frac{\partial \ln f_{2t}}{\partial \lambda}+1\right)\right] & ; y_t > \mu \end{cases}, \tag{12}$$

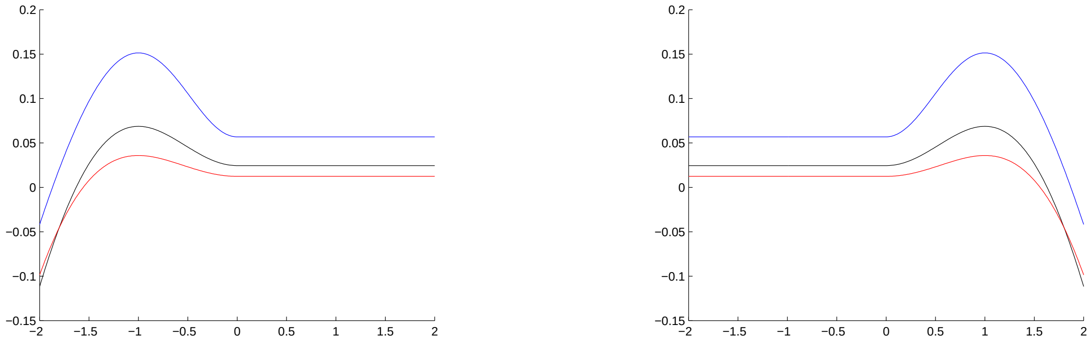

Figure 6: Plot of the score with respect to different residuals values ε_t for $\upsilon_1 = \upsilon_2 = 2$ and $\eta_1 = \eta_2 = 2$ (Blue), $\eta_1 = \eta_2 = 6$ (Black), $\eta_1 = \eta_2 = 10$ (Red). The score with respect to η_1 (Left) and the score with respect to η_2 (Right).

and

$$\mathcal{I}_{1\vartheta\vartheta} = \begin{cases} \widetilde{\eta}_1^2 \frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\upsilon_1}\left(\left[\left(1-\alpha^+\right)\tau_1 + \alpha^+\left(1-\alpha^+\right)\frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\upsilon_1}\tau_1^2 - \frac{\alpha^+}{\widetilde{\eta}_1}\frac{\partial\tau_1}{\partial\vartheta_1}\right] - \right. & \\ \left. -\frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\eta_{1t|t-1}^2}\frac{\eta_{1t|t-1}(\upsilon_1-1)-(\upsilon_1+1)}{\left(\eta_{1t|t-1}+1\right)\left(\eta_{1t|t-1}+1+\upsilon_1\right)}\right) & ; y_t \le \mu \\ \widetilde{\eta}_1^2 \frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\upsilon_1}\left[\alpha^+\left(1-\alpha^+\right)\frac{\left(\eta_{1t|t-1}-\eta_1^\dagger\right)}{\upsilon_1}\tau_1^2 - \frac{\alpha^+}{\widetilde{\eta}_1}\frac{\partial\tau_1}{\partial\vartheta_1} - \alpha^+\tau_1\right] & ; y_t > \mu \end{cases}, \tag{13}$$

$$\mathcal{I}_{2\vartheta\vartheta} = \begin{cases} \widetilde{\eta}_2^2 \frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\upsilon_2}\left[\alpha^+\left(1-\alpha^+\right)\frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\upsilon_2}\tau_2^2 - \frac{\left(1-\alpha^+\right)}{\widetilde{\eta}_2}\frac{\partial\tau_2}{\partial\vartheta_2} - \left(1-\alpha^+\right)\tau_2\right] & ; y_t \le \mu \\ \widetilde{\eta}_2^2 \frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\upsilon_2}\left(\left[\alpha^+\tau_2 + \alpha^+\left(1-\alpha^+\right)\frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\upsilon_2}\tau_2^2 - \frac{\left(1-\alpha^+\right)}{\widetilde{\eta}_2}\frac{\partial\tau_2}{\partial\vartheta_2}\right] - \right. & \\ \left. -\frac{\left(\eta_{2t|t-1}-\eta_2^\dagger\right)}{\eta_{2t|t-1}^2}\frac{\eta_{2t|t-1}(\upsilon_2-1)-(\upsilon_2+1)}{\left(\eta_{2t|t-1}+1\right)\left(\eta_{2t|t-1}+1+\upsilon_2\right)}\right) & ; y_t > \mu \end{cases}, \tag{14}$$

The asymmetry mixing parameter α^+ is defined as

$$\alpha^+ = \frac{\alpha/K_1}{\alpha/K_1 + (1-\alpha)/K_2}$$

which as noted by Harvey and Lange (2017) is the probability of having a negative observation. The parameters τ_i and their derivatives are defined as

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

Then it is possible to model each of the individual tail index parameters through the link function $\eta_{it|t-1} = \eta_i^\dagger + e^{\widetilde{\eta}_i\vartheta_{it|t-1}}$, where $\widetilde{\eta}_i = 1$ if we are modelling the tail index parameter and $\widetilde{\eta}_i = -1$ if we are modelling its inverse, and a dynamic QARMA specification for $\vartheta_{it|t-1}$ of the form

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

where $u_{it}^{\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_i} \mathcal{I}_{i\vartheta\vartheta}^{-1}$, all for $i = 1, 2$. Finally, following Zhu and Galbraith (2010) we can construct the Loglikelihood function as

$$
\begin{aligned}
L(\boldsymbol{\psi}_{\lambda}, \boldsymbol{\psi}_{1\vartheta}, \boldsymbol{\psi}_{2\vartheta}, \alpha\, \upsilon_1, \upsilon_2) = & -\sum_{t=1}^{T} \lambda_{t|t-1} + \sum_{t=1}^{T} \ln K_{12}\left(\eta_{1t|t-1}, \eta_{2t|t-1}, \upsilon_1, \upsilon_2\right) - \\
& -\sum_{t=1}^{T} \frac{\left(\eta_{1t|t-1}+1\right)}{\upsilon_1} \ln\left(1 + \frac{|\varepsilon_t|^{\upsilon_1}}{\eta_{1t|t-1}}\right) \mathbf{1}_{(\varepsilon_t \leq 0)} - \\
& -\sum_{t=1}^{T} \frac{\left(\eta_{2t|t-1}+1\right)}{\upsilon_2} \ln\left(1 + \frac{|\varepsilon_t|^{\upsilon_2}}{\eta_{2t|t-1}}\right) \left(1 - \mathbf{1}_{(\varepsilon_t \leq 0)}\right)
\end{aligned}
$$

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 $\vartheta_{2t|t-1}$.

Existing models in the extreme value theory literature focus only on observations which exceed a pre-determined threshold and are therefore considered as belonging to the "tail" of the distribution. This means that the "non-tail observations" or, particularly in the case of asymmetric tails modelling, the observations that fall in the opposite tail to the one modelled are treated as missing[17]. In the DCS framework, the score with respect to each tail index is still only directly affected by the residuals which appear in its side of the distribution since, as can be seen in Figure 6, its response is flat starting from the median and continuing through for all the residuals values in the opposite side of the distribution. This is because for an observation belonging to the opposite side of the distribution, the last two terms of Equation (11) and Equation (12), which depends on b_{it}, disappear; as would happen for an observation at the median. Therefore, in this case the score generates the same response as if the observation was at the median, instead of treating the observation as missing, and producing a response of 0. Moreover, the score in this case still depends on both α^+ and τ_i which use information from both the tails. Due to this structure also in this case the score remains time varying through its dependence on both $\eta_{1t|t-1}$ and $\eta_{2t|t-1}$. This feature comes directly from the conditional score of the asymmetric distribution, rather than arbitrarily setting a treshold to define which are the "tail observnations". As a consequence, at each point in time the DCS asymmetric tail model uses more information from the observations in both sides of the distribution in fitting the true dynamics of each of the two tail index parameters.

5. Empirical Results

For the reminder of the paper we will be focusing only on the t distribution and its asymmetric counterpart. Therefore we are restricting $\upsilon = \upsilon_1 = \upsilon_2 = 2$ and $\alpha = 1/2$.

In order to investigate the effectiveness of the new dynamic tail model on different types of data series, we have considered returns from Equity Indexes and Credit Default Swaps, which are known for their extreme fluctuations over time.

[17] For example, at time t the residual $\varepsilon_t > 0$ would consider it as a missing observation while modelling the lower tail parameter

	Mean	St. Dev.	Skewness	Kurtosis	Min.	Max.	$Q(20)$
FTSE 100	0,000	0,011	-0,480	12,561	-0,130	0,0934	86,201
CDS 5Y Italy	0,001	0,043	0,288	18,912	-0,437	0,429	25,395

Table 1: Descriptive Statistics

For the equity indexes we have considered the Dow Jones Index daily log returns. The data are collected from Yahoo Finance and are between the 29th of January 1985 to the 29th of April 2016. For the CDS we have considered daily log returns of 5y CDS rates for the Italian sovereign debt. The data are collected from Bloomberg and are from the 1st of March 2007 to the 21st of September 2018. The particular choice of the CDS data series was motivated by the fact that, among the European sovereign CDS, it was the one that exhibits the most extreme behaviour while maintaining a relative high liquidity.

From Table 1 it is possible to see that the two series considered both have a high sample kurtosis, higher for the CDS than for the Equity Index. The CDS series is right skewed while having a sample standard deviation four times higher than the Equity Index, which comes out as left skewed. In all the cases there are signs of residual correlation at lag 20.

In order to estimate the Dynamic Scale-Tail DCS model we have first fitted to both the series a beta-t-EGARCH DCS model, assuming a conditional t distribution. Than, using the fitted residuals we have computed the scores under the null, $\hat{u}_t^{\vartheta\dagger}$, as in Equation (10) and performed the simple Box-Ljung test $Q_u^*(P)$[18]. Then, where appropriate, we have fitted the general Dynamic Scale-Tail DCS model[19] . All the estimations are performed by maximum likelihood[20]. Both Conditional Symmetric and Asymmetric t distributions specifications were considered.

Remark 3. *In modelling the individual tails of the Asymmetric t distribution, the score with respect to the dynamic tail index parameter of each of the tails depends on the observations only if the observation falls in its tail. For this reason, the simple test $Q_u^*(P)$ performed on each individual tail index parameter will effectively use less observations and therefore we expect it to have a lower power compared to a test based on the symmetric tail index parameter.*

In fitting the Beta-t-EGARCH model to the Dow Jones Index returns series we had to assume a two components dynamics for $\lambda_{t|t-1}$, as described in Harvey (2013) pg.91-92, in order to capture the long memory feature of return's volatility and remove all the residual correlation in the fitted scores with respect to λ which could affect, through the scores with respect to the tail index parameter u_t^{ϑ}, the accuracy in the detection and estimation of the tail index parameter dynamics.

[18] Another reason for preferring the simple Box-Ljung version of the test is that, besides from its simplicity and effectiveness, it allows for an immediate comparison with the Box-Ljung test performed on the fitted scores to detect residual correlation after having fitted the Dynamic Tail.

[19] 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 $\mathcal{I}_{\lambda\lambda}$, since this would also be time varying.

[20] Since the estimation of the general Dynamic Scale-Tail model is not trivial, to improve the accuracy of the parameters estimates we have first fitted to the standardised data a Dynamic Tail DCS model, assuming the tail index parameter to be dynamic and the scale constant set to 1. Than we used the estimated parameters in combination with the parameter estimates of the Beta-t-EGARCH DCS model as starting values for the parameters of full Dynamic Scale-Tail DCS model.

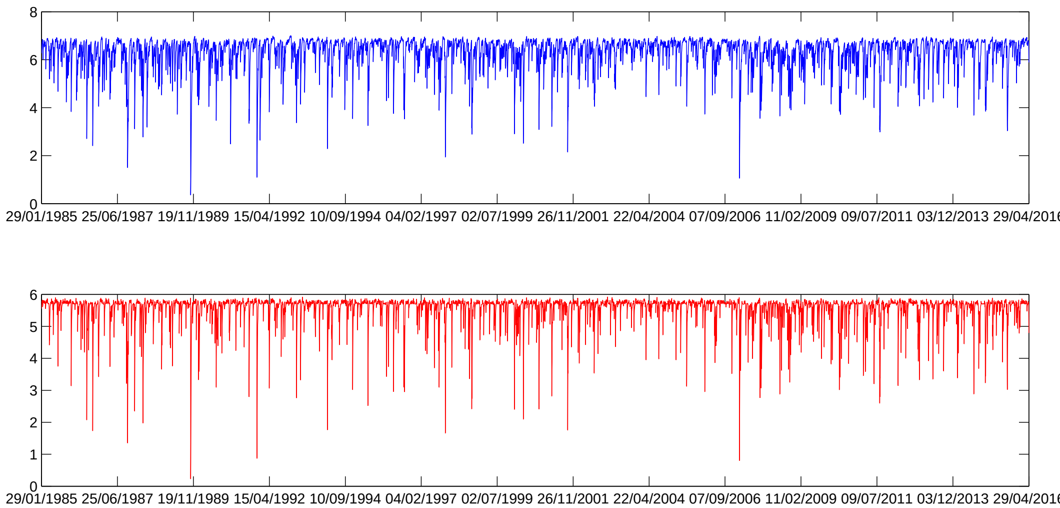

Figure 7: Fitted degrees of freedom of the Dow Jones Index Returns in the symmetric case (Top), $\hat{\eta}_{t|t-1}$, and for the Lower Tail in the asymmetric case (Bottom),$\hat{\eta}_{1t|t-1}$ $\bar{\eta}_{t|t-1}$.

Given the length of the series it raises the question if we should take into account possible leverage effects. The problem in doing so is that we are introducing some sort of asymmetric response to negative returns in the scale dynamics, which could affect the behaviour of our dynamic asymmetric tail model. For this reason we provide in Table H.6 the results without and with leverage effect, which can be added to the dynamics of the components of $\lambda_{t|t-1}$ as

$$\begin{aligned}\lambda_{t|t-1} &= \omega_\lambda + \lambda_{1,t|t-1} + \lambda_{2,t|t-1} \\ \lambda_{i,t+1|t} &= \phi_{i,\lambda}\lambda_{i,t|t-1} + \kappa_{i,\lambda}u_t^\lambda + \kappa_{i,\lambda}^*\mathrm{sgn}(-y_t)(u_t^\lambda + 1) \qquad i = 1, 2\end{aligned}$$

In Table H.7 it is possible to see that in the case of the model without leverage the $Q_u^*(P)$ test rejects the null of static degrees of freedom in the symmetric case, but in the asymmetric case only for the parameter for the lower tail η_1, suggesting a dynamic lower tail and a static upper tail. This result can explain the findings of Mazur and Pipień (2018), who identified the left tail of returns to be more variable and consistently heavier than the right tail. The Dynamic Scale-Tail model is then fitted accordingly. From Table H.6 it is possible to see that the dynamics of the degrees of freedom parameters are not too persistent with the parameter for the lower tail being less persistent than the one for the symmetric tail. The Box-Ljung test results on the fitted scores in Table H.8 suggests that the model fits the dynamic parameter well, removing all the correlation from the $Q_u^*(P)$ test up to lag 50. We have to notice thought that the simple Beta-t-EGARCH with two components, either with symmetric or asymmetric tails, is not capable to remove entirely the residual correlation from the fitted scores with respect to the scale parameter $\lambda_{t|t-1}$. However after letting the tail parameters be dynamic, also all the residual correlation in the dynamic scale parameter $\lambda_{t|t-1}$ is then removed. The improvement in the fit from modelling the data with a dynamic tail is ultimately confirmed

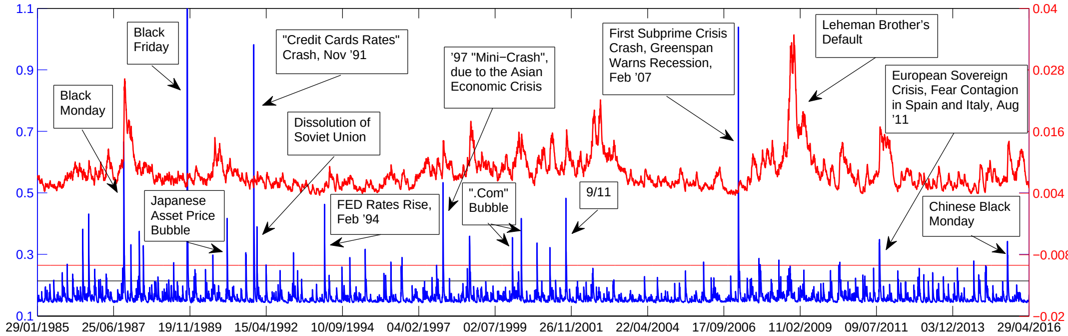

Figure 8: Plot of the fitted scale $\exp\left(\lambda_{t|t-1}\right)$ without leverage, (Red Line), and the inverse of the fitted degrees of freedom for the Lower Tail $\bar{\eta}_{1t|t-1}$ of the Dow Jones Index Returns in the asymmetric case, (Blue Line), with the confidence bounds for one and two standard deviations from the mean.

by the higher likelihood and lower information criteria.

As can be seen in Figure 7, the fitted symmetric parameter $\eta_{t|t-1}$ moves between 1 and 6.5 but mostly staying around 6.5. This somehow is similar to the results found by Blazsek and Monteros (2017); Ayala et al. (2017) on the S&P 500 and Massacci (2017) on the shape parameter behaviour of small firms, which seemed to have a "floor" around a constant number. Ultimately it falls below 1 only in the case of the "Black Friday" market crash of November '89. Figure 7 shows also the plot of the fitted parameter $\eta_{1t|t-1}$ for the lower tail which is on average around 6, slightly heavier than the symmetric one but with almost identical fluctuations just slightly more pronounced.

To identify the effect of the occurrence of notable market events on the lower tail movements, in Figure 8 we have plotted the inverse of the symmetric tail parameter $\bar{\eta}_{t|t-1}$ against the fitted time varying scale $\exp\left(\lambda_{t|t-1}\right)$. From this is possible to see that the heaviness of the lower tail of the returns distribution matches most of the notable market events. However its movements are not necessary linked to volatility. As expected there are cases when volatility is high and the lower tail is also heavier, as for the case of the "Black Monday". However the vast majority of extreme movements in the lower tail happens when the volatility moves the least. From this we can identify the "Black Friday" November '89 market crash which followed the "Black Monday", the November '91 market crash due to congress vote on increasing the credit card rates and the February '07 market crash at the beginning of the subprime crisis when Greenspan suggested the possibility for the US to enter in a recession. All these were unexpected extreme events which moved the market unidirectionally down while the volatility fitted by the Beta-t-EGARCH didn't move much. On the other hand events like the Leheman default are fully taken into account into the volatility of the market leaving the heaviness of the lower tail almost unaltered.

Looking instead at the inverse of the lower tail parameter $\bar{\eta}_{1t|t-1}$ in Figure G.16, we can see that there are less spikes and some of them are less pronounced. However some of the events identified in Figure 8 are still present here, confirming the idea that most of the extreme events tends to occur in presence of negative returns.

Once introduced the asymmetric response in the scale parameter through the leverage component in the

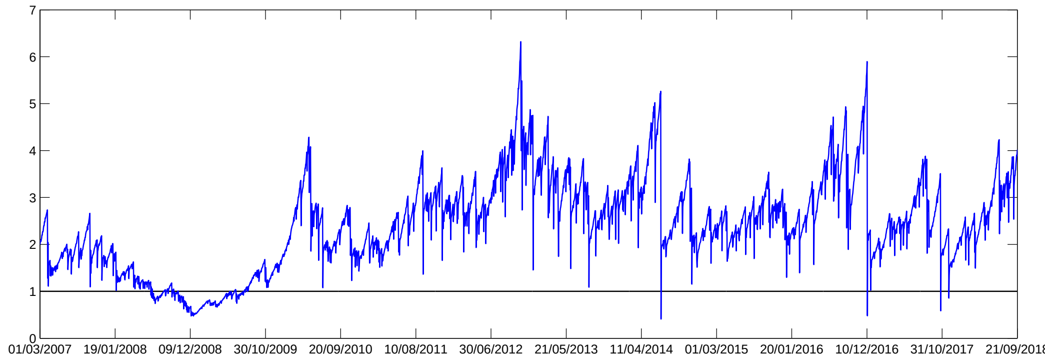

Figure 9: Plot of the the fitted estimated degrees of freedom, $\eta_{t|t-1}$ for the symmetric model.

Beta-t-EGARCH all the remaining residual correlation in the fitted scores with respect to $\lambda_{t|t-1}$ is removed. On the other hand from Table H.9 it is possible to see that in the asymmetric tails case the $Q_u^*(P)$ reveals residual correlation only in the fitted scores with respect to the upper tail parameter, η_2, rather than in the lower tail. This can be explained by the fact that the inclusion of the leverage term allows the scale to capture most of the extreme negative movements neglecting some of the positive which ultimately should be modelled separately. All the estimated $\kappa_{i,\lambda}^*$ are positive and the leverage impact is mostly confined in the less persistent component of $\lambda_{t|t-1}$, confirming the findings of Harvey and Lange (2018). The symmetric dynamic tail model has similar fitted dynamics and paths for the parameter $\eta_{t|t-1}$ to the case without leverage. However, the fitted $\eta_{2t|t-1}$ is quite persistent with a different path from $\eta_{1t|t-1}$ in the case without leverage. The path of its inverse in Figure G.16 reveals much less extreme movements, partly due by its long run average around $\exp(\omega_{\eta 2}) = 10.014$, which occurs at different time periods t than for $\bar{\eta}_{1t|t-1}$. However these are still periods when the volatility is low. Finally, we can see from Table H.6 that there is an overall significant preference in terms of likelihood and information criteria for the asymmetric Scale-Tail model with the leverage term, however this fails to capture some of the residual correlation in the fitted scores with respect to scale at earlier lags.

These results can be explained by the fact that as the tail index parameter of the conditional distribution falls, the score with respect to scale that drives its dynamics becomes more bounded preventing extreme scale movements as long as they are not persistent in the series, see Harvey (2013). This feature, in the score driven literature, it has been explained by the robustness to outliers of the score with respect to scale. However, once allowed for the tails of the conditional distribution to vary, sudden unexpected extreme events, if repeated, rather than moving the scale tend to move the tail, which becomes more heavier and allows allows for more extreme events to occur. The phenomenon is clearer in the asymmetric case where, for example, if the leverage effect on scale is not taken into account the lower tail index moves to capture these rapid non persistent falls of the series neglected by the scale, which are detected in the residual correlation of fitted scores with respect to the tail index parameter. In this way the model can effectively distinguish between scale movements and tail movements, either if they occur occasionally or are more persistent. In the case of Index Returns tail movements seems to be rarely persistent, therefore the effectiveness of the dynamic tail

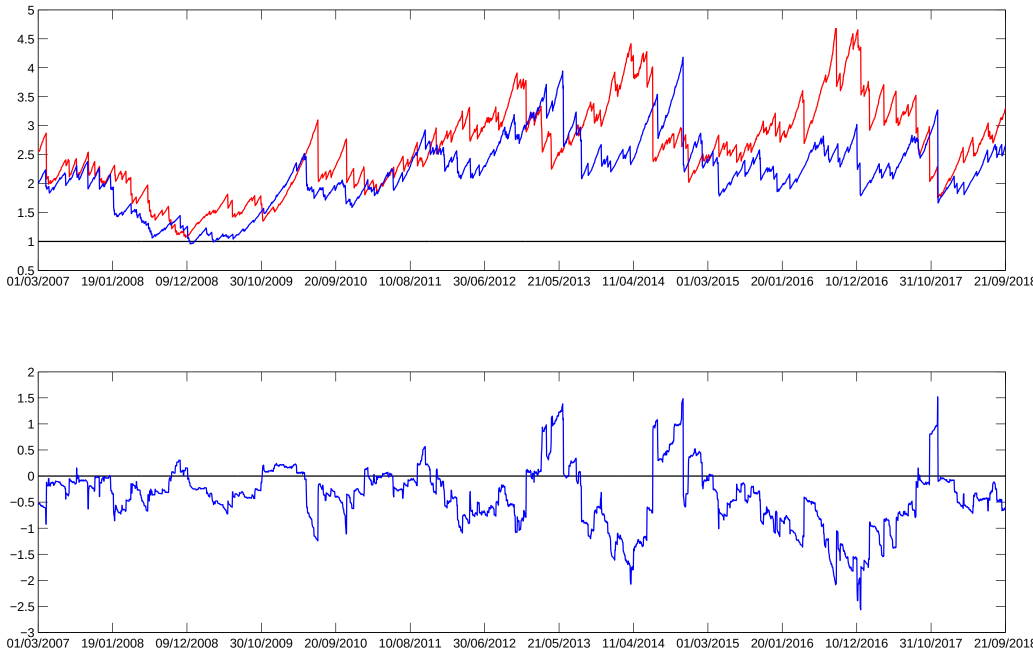

Figure 10: The top figure shows a plot of the the fitted estimated degrees of freedom, $\eta_{2t|t-1}$ and $\eta_{1t|t-1}$, for the upper and lower tail in the asymmetric model, (Black and Red line respectively). On the bottom figure shows the spread between the upper and lower tail dynamic degrees of freedom of the asymmetric distribution, $\eta_{2t|t-1} - \eta_{1t|t-1}$.

model could be better appreciated instead with series which exhibits more extreme and frequent occurrences of extreme events.

In fitting the scale of the Italian CDS series with the Beta-t-EGARCH model a one component dynamics is enough to remove most of the residual correlation in the fitted scores with respect to $\lambda_{t|t-1}$ up to lag 50, a part from lag 1 and 5. This can be noticed from the results of the Box-Ljung test in Table H.12 where is also possible to see that the $Q_u^*(P)$ rejects the null of static degrees of freedom up to lag 50 in both the symmetric and asymmetric cases. To remove fully this residual correlation from the fitted scores with respect to $\vartheta_{t|t-1}$ we have opted in the symmetric case for a QARMA(1,1) specification as

$$\vartheta_{t+1|t} \quad = \quad (1-\phi_\vartheta)\,\omega_\vartheta + \phi_\vartheta \vartheta_{t|t-1} + \kappa_{1\vartheta} u_t^\vartheta + \kappa_{2\vartheta} u_{t-1}^\vartheta$$

while in the asymmetric case we have used the same QAR(1) dynamics as described in Equation (4) for both individual tail parameters.

In Table H.11 is possible to see that all the three fitted dynamic tail index parameters are very persistent, almost $I(1)$. Also in this case the improved fit of the Dynamic Scale-Tail specification is confirmed by a higher Likelihood and lower information criteria. In Figures 9 and 10 is possible to see that the fitted tail index parameters are much more persistent than in the case of Index Returns[21] and tend to move quite

[21] This is because CDS returns exhibits more frequent and extreme movements than Index returns.

closely together between 4 and 1. In both the cases they all stay for long periods under 2 in the early part of the sample, like around the '08 Lehman default, while, as can be seen in Figure 9, the symmetric tail index parameter falls also below 1 in that same period, suggesting that the conditional distribution could at times not have variance[22]. This result shows how dynamic models which focuses on moments of the conditional distribution, like GARCH, can become invalid in this context. However, the tail index parameter only fall below 1 for short periods of time despite not being bounded since, as explained in appendix Appendix A, the score naturally pushes the tail index parameter away from extremely low values unless in presence of a large number of very extreme observations. The tail index parameter of the symmetric distribution is the one that tends to move the most and seems to follow mostly the movements of the lower tail, despite being sometimes higher than either of the two tail index parameters in the asymmetric specification.

In regards to asymmetric distribution, the relative comparison of its tail parameters is presented in form of the spread $\eta_{2t|t-1} - \eta_{1t|t-1}$, which seems informative of periods of financial distress for the country. Indeed the periods when the spread becomes negative[23] coincides with the periods of economic and political turbulence in Italy, when the CDS rates have increased rapidly.

5.1. Conditional Distribution Modelling under Dynamic Tails

The inclusion of dynamic tails has a direct impact on the actual modelling of the conditional distribution of the data which can be better appreciated looking at its quantiles. Figure 11 shows the upper and lower 0.5% quantiles fitted by the GARCH model and the asymmetric Dynamic Tail DCS model on the Italian 5Y CDS returns data series. From this we can see that the returns data series touches quite often the upper and lower quantiles fitted by the GARCH model. This suggests that there have been several occasions across the dataset in which returns exhibits movements that should happen with probability 0.5%. Precisely, across the whole sample 1.36% of the data crosses the GARCH upper quantile and 1.19% the lower quantile, while in the case of the asymmetric Dynamic Tail DCS model only the 0.03% for both the upper and lower tail. This suggest that the GARCH model is far less conservative than the Dynamic Tail Index model underestimating the occurrence of extreme events. This can be clearly seen also in the occasion of the 15th of July 2008, two months before the Leheman bankruptcy, when the 5Y Italy CDS moved from 21.167 to 32.5 in one day. The GARCH model estimated that this event could have occurred with a probability of 0.06%, while the static Beta-t-EGARCH DCS model with a probability of 0.57% and the symmetric and asymmetric Dynamic Tail DCS models with probabilities of 2.22% and 1.18% respectively.

In order to see if these significant differences can also be detected out-of-sample we have made a density forecasting exercise where we have obtained one-step-ahead point and density forecasts on the 5y Italy CDS data for the two years in the sample. A total of 730 observations out-of-sample. The forecasts are obtained re-estimating the models using all the data up to the previous date to one forecasted.

[22] These low values coincides effectively with periods when the CDS is quite illiquid and there are many consecutive zeros which makes the conditional distribution very heavy tailed. However the total number of zeros in the entire sample is less than 5% and are are mainly located in these early parts of the sample.

[23] These are periods in which the lower tail is closer to Gaussianity than the upper tail and therefore implies more extreme returns towards positive values.

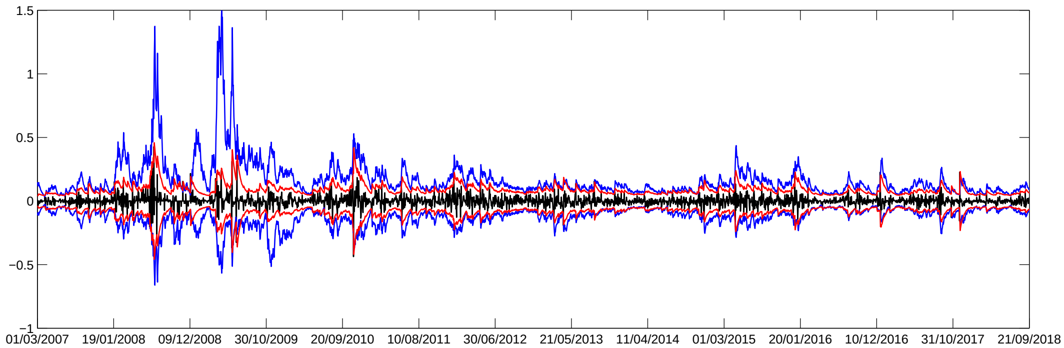

Figure 11: Plot of the 0.5% upper and lower quantiles of the conditional distribution of the returns series for the Italian 5y CDS Rates (Black Line), when fitted by a GARCH model (Red Line) and a DCS with dynamic asymmetric tails (Blue Line).

We define the one-step-ahead p-lower and p-upper Value-at-Risk (VaR) as the quantity

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

which for a symmetric distribution around 0 are respectively $\mathrm{VaR}_{1,1-p}\left(y_{T+1}\right)=F^{-1}_{Y_{T+1|T}}\left(p\right)$ and $\mathrm{VaR}_{2p}\left(y_{T+1}\right)=-F^{-1}_{Y_{T+1|T}}\left(p\right)$, where $F_{Y_{T+1|T}}\left(.\right)$ is the one-step-ahead forecasted conditional CDF of the quantity y_t. In case of an asymmetric distribution centred at 0 with one-step-ahead forecasted conditional CDFs $F_{1Y_{T+1|T}}\left(.\right)$ and $F_{2Y_{T+1|T}}\left(.\right)$ for the distributions describing respectively the left and right tail of the distribution of y_t, we have that $\mathrm{VaR}_{1p}\left(y_{T+1}\right)=F^{-1}_{1Y_{T+1|T}}\left(p\right)$ and $\mathrm{VaR}_{2p}\left(y_{T+1}\right)=-F^{-1}_{2Y_{T+1|T}}\left(p\right)$. On the other hand we define the one-step-ahead lower and upper Expected Shortfall (ES) as

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

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

$$\begin{aligned}h_{1T+1}\left(p\right)=&\mathbf{1}_{\left(y_{T+1}<\mathrm{VaR}_{1p}\left(y_{T+1}\right)\right)}\\ h_{2T+1}\left(p\right)=&\mathbf{1}_{\left(y_{T+1}>\mathrm{VaR}_{2p}\left(y_{T+1}\right)\right)}\\ h_{T+1}\left(2*p\right)=&h_{1T+1}\left(p\right)+2h_{2T+1}\left(p\right)\end{aligned}$$

In this way we can construct both the unconditional coverage and independence likelihood ratio tests of Christoffersen (1998) for the VaR violations for both the upper and the lower tail, individually or jointly. The first test corresponds to the null $H_0:E\left[h_{iT+1}\left(q\right)\right]=\mathbb{P}\left(h_{iT+1}\left(p\right)=1\right)=p$, while the second tests the null hypothesis $H_0:\mathbb{P}\left(h_{iT+1}\left(p\right)=1|h_{iT}\left(p\right)\right)=p$. For both the individual tails the tests are distributed as $\chi^2\left(1\right)$. The tests for the joint violations are described in the paper as tests for the asymmetry of the predictive distribution which can be easily constructed also in our framework. Since they have three possible outcomes, 2 for up violation, 1 for low violation and 0 for no violation, they are distributed with a $\chi^2(2)$ and $\chi^2(4)$ respectively.

To evaluate the ES for both the tails we have used the unconditional backtest of Du and Escanciano (2017). These tests can be constructed from the cumulative violation processes

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

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) show that testing the correct specification of the ES simplifies to test weather the mean of the $H_{iT+1}(p)$ is equal to $p/2$ and can be tested through the t statistics

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

where $\bar{H_i}(p)$ is the sample mean of the $H_{iT+1}(p)$, $v_{ES} = Var(H_{iT+1}(p)) = p(1/3 - p/4)$ and T_f is the number of out-of-sample observations[24].

From the results of the unconditional coverage test in Table H.14 we can see that in the case of fixed tails the quantiles levels are significantly misspecified, in particular in the case of the GARCH and for the lower tails of the DCS models. This can be explained by the fact that, given the assumption of gaussianity of the GARCH, the time variation in the quantiles only depends on the variation in the conditional variance which tends to spike in presence of extreme events. As a consequence the model overestimates the quantiles closer to the median in favour to the one in the tails, as can be seen from the results of the unconditional coverage tests. The results of the unconditional backtests shows that in the lower tail the lower ES are underestimated for the quantiles closer to the median and in the higher tail the upper ES are overestimated for the quantiles further in the tails.

In the case of the symmetric fixed tail DCS model, or beta-t-EGARCH, the estimated degrees of freedom are pushed low by the extreme movements of the upper tail overestimating the quantiles further in the lower tail. This produce an overal good estimate of the upper ESs for the higher tail and underestimates the lower ESs for the lower tail. In the case of the asymmetric fixed tail DCS model, or the asymmetric t (AT) DCS model of Thiele (2020), the upper tail index parameter is estimated smaller than the lower tail index parameter. However not taking into account of the time variation in the tails the quantiles tend to still be overestimated in the case of the lower tail and underestimated in the case of the upper tail, with a more significant problem for the ES of the lower tail.

On the other hand looking at the results for the dynamic tails DCS models, the tails are much better

[24] We have also considered the conditional backtest of Du and Escanciano (2017), however the low number of rejections was not enough to discriminate between models, therefore the results are not reported.

taken into account. However still in the symmetric model the dynamics is mainly driven by the upper tail movements, therefore some of the quantiles further in the lower tail are significantly overestimated in respect to the asymmetric model. In terms of unconditional independence test, we have that the violations of the quantiles tends to be significantly dependent only for some quantiles in upper tails of the the fixed tails models, while the dependence is completely removed in the dynamic tail models.

The overall predictive likelihood of the DCS models is much higher than in the GARCH with comparable sizes across the the various specifications. The only exception is for the symmetric dynamic tail DCS model, which due to its more erratic ARMA specification in the dynamics of the tail index parameter, has a predictive likelihood slightly lower than the other DCS models. For this reason, and the results of the tests, we can assume that the asymmetric dynamic tails DCS model is the most appropriate to model the 5y Italy CDS dataset.

As a further illustration of the results, in Figure G.17 and Figure G.18 we can see the lower and upper out-of-sample ES from the same analysis[25] reported as ratios on the ES forecasted by the GARCH. From these we can see that the GARCH model underestimates in both the cases the length of the tails. The ES forecasted are for the 10% quantiles half the one forecasted with the dynamic tails models and for the 0.1% quantile from 5 times up to in some occasions more than 35 times the ones forecasted with the dynamic tails models. In general the Expected Shortfall ratios from the asymmetris fixed tail DCS model tends to be higher for the upper tail and lower for the lower tail than for the symmetric fixed tail DCS model. On the other hand, the Expected Shortfall ratios for the asymmetric dynamic tail DCS model tends to vary a lot across the sample. For most of the out-of-sample dataset they are lower than the one of the asymmetric fixed tail DCS model in the case of the lower tail, while they move rapidly both above and below the the one of the asymmetric fixed tail DCS model depending on the time periods. As expected, the largest fluctuations across the forecasted sample happen for the 0.1% quantile.

6. Conclusion

The present work studies the time variability of the occurrence of extreme events in time series. This can be described by the fluctuations over time of the tail index of the conditional distribution of the data. The paper introduces a dynamic DCS model for the tail index parameter while assuming that the data are generated by a conditional Generalised t distribution.

An LM test to detect the presence of dynamics in the tail index parameter is also introduced. This is based on the autocorrelation of the score with respect to the tail index parameter under the null of no variability. A closed form solution of the test is derived. The power and size of the full LM test are then compared with a simple Box-Ljung test performed on the fitted scores of the model under the null. The results reveal that the full LM test is a more conservative version of the Box-Ljung with a lower probability

[25] Here the Expected Shortfall from the Symmetric Scale-Tail model has been excluded to better appreciate the difference between the other models given the fact that due to the its ARMA specification in the tail parameter its tail has very large fluctuations.

of rejection. The difference is more pronounced in cases when the tail index is particularly low or there are fewer observations. The comparison is then extended to include also the GAS-LM test of Calvori et al. (2017) and the Nyblom test. The GAS-LM test, with lags automatically set by the algorithm of Escanciano and Lobato (2009), performs better than the LM in presence of an extremely persistent tail index parameter and with a small sample size. However in all the other cases the newly introduced tests, both LM and the simple Box-Ljung, are superior in terms of power than all the other competitors.

The efficiency of the Dynamic Tail DCS model in estimating the dynamic parameters of the tail index is also assessed under various parameter assumptions. As expected, the results show that the estimation accuracy of the model falls as the sample size decrease. However, this happens faster when the true tail index is on average around 30 or larger. On the other hand, the model is particularly effective when the true tail index is on average smaller than 15.

Finally the Generalised t distribution is extended to its asymmetric version in order to give a separate dynamics to the upper and lower tail index parameters.

Further implications of bounding the dynamics of the tail index parameter to guarantee the existence of moments are also analysed in the appendix. The analysis reveal that the bounding can imply serious distortions in the score response and therefore affect the performance of the filter in capturing the true dynamics of the tail index.

Both the models, symmetric and asymmetric, and the tests are then empirically implemented on market returns data from the Dow Jones Equity Index and the 5Y Italy CDS. The results show that, in the case of the Equity Index, the tests detect a dynamics in the symmetric tail index parameter. However if the distribution is believed to be asymmetric the dynamics is detected only in the lower tail index parameter if we do not including a leverage term in the scale dynamics, or in the upper tail index parameter if we include the leverage term. Moreover both the fitted dynamic tail indexes are not too persistent and tend to be bounded from above falling only rarely below 1.

In the case of the CDS returns both the symmetric and the two asymmetric tail index parameters are detected to be dynamic. All three parameters have a very persistent dynamics moving from 4-6 down below 1 occasionally. The analysis of the spread between the upper and the lower tail index parameters in the asymmetric case shows how the relative heaviness of the two tails varies considerably over time. The two parameters tends to move together for most of the data sample, diverging mostly in the last part where, in particular between 2016 and 2017, the upper tail is heavier than the lower tail. This is consistent with the rapid increase of the CDS price during the political crisis in Italy. Finally, an out-of-sample analysis of the forecasted quantiles and Expected Shortfalls have proven that the dynamic Tail DCS models are much less conservative than the GARCH in forecasting the tails length, and therefore forecasting higher probabilities of occurrence of extreme events with significant evidences of asymmetries and time variation in magnitudes depending on the time periods.

This tails behaviours are of high interest for practitioners, therefore the model can have many empirical applications. In particular, in the asymmetric model would be interesting to investigate if there are cases in which an increase in magnitude of one tail can imply an increase in magnitude in the other tail, as shown by

Massacci (2017). Moreover would be interesting to investigate the impact of the inclusion of other variables regarding the real economy as explanatory variables on the fit of the tails for both the Equity and CDS datasets. Finally, in terms of systemic risk, would be interesting to look at these analysis in a multivariate framework also across countries. For example, we could try to assess the possible relation between cross-country or cross-assets tail movements. Finally, this model could give another perspective on the idea of tail association while setting up dynamic copulas.

References

References

Allen, L., T. Bali, and Y. Tang (2012). Does systemic risk in the financial sector predict future economic downturns? *The Review of Financial Studies 25*(10), 3000–36.

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

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

Blazsek, S. and L. Monteros (2017). Dynamic conditional score models of degrees of freedom: filtering with score-driven heavy tails. *Applied Economics 49*(53), 5426–40.

Brooks, C., S. Burke, S. Heravi, and G. Persand (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics 3*(3), 399–421.

Calvori, F., D. Creal, S. Koopman, and A. Lucas (2017). Testing for parameter instability across different modeling frameworks. *Journal of Financial Econometrics 15*(2), 223–46.

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

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

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

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). Modelling extremal events: for insurance and finance.

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

Escanciano, J. and I. Lobato (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics 151*(2), 140–9.

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

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

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

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

Harvey, A. and M. Streibel (1998). Tests for deterministic versus indeterministic cycles. *Journal of Time Series Analysis 19*(5), 505–29.

Harvey, A. and S. Thiele (2016). Testing against changing correlation. *Journal of Empirical Finance 38*, 575–589.

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

Harvey, C. and A. Siddique (1999). Autoregressive conditional skewness. *The Journal of Financial and Quantitative Analysis 34*(4), 465–87.

Kelly, B. (2014). The dynamic power law model. *Extremes 17*(4), 557–83.

Kelly, B. and H. Jiang (2014). Tail risk and asset prices. *Review of Financial Studies 27*(10), 2841–871.

Lamoureux, C. and W. Lastrapes (2002). Persistence in variance, structural change and the garch model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 64*(2), 253–280.

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

Lucas, A. and X. Zhang (2016). Score-driven exponentially weighted moving averages and value-at-risk forecasting. *International Journal of Forecasting 32*(2), 293–302.

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

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

McDonald, J. and R. Michelfelder (2017). Partially adaptive and robust estimation of asset models: Accommodating skewness and kurtosis in returns. *Journal of Mathematical Finance 7*, 219–237.

Muller, U. and P. Petalas (2010). Efficient estimation of the parameter path in unstable time series models. *The review of Economics Studies* (77), 1508–1539.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association 84*(405), 223–30.

Quintos, C., Z. Fan, and P. Phillips (2001). Structural change tests in tail behaviour and the asian crisis. *The Review of Economic Studies 68*(1), 633–663.

Thiele, S. (2020). Modeling the conditional distribution of financial returns with asymmetric tails. *Journal of Applied Econometrics 35*(1), 46–60.

Wagner, N. (2005). Autoregressive conditional tail behavior and results on government bond yield spreads. *International Review of Financial Analysis 14*, 247–61.

Werner, T. and C. Upper (2004). Time variation in the tail behaviour of bund futures returns. *Journal Futures Markets 24*(4), 387–98.

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

Appendix A. Bounding the Tail Index Dynamics

The score with respect to the tail index parameter is a continuous function of the residuals at every time period t. The advantage in modelling the tail index parameter with such a score-driven approach in comparison to the one from extreme value theory is that, rather than focusing only on the observations that fall in the tails, it makes coherent use of all the observations in the time series. It is important to bear in mind that in a DCS-EGARCH model, once η is estimated and fixed for every t, the score with respect to λ is only a function of ε_t. On the other hand, in a DCS dynamic tail model, once the fixed shape parameters are estimated, at every t the score with respect to ϑ can be considered as a three dimensional function of both ε_t and $\eta_{t|t-1}$.

As can be seen also in Figure 4, the score response for observations around the median increases as η_t becomes smaller. The reason for this is that as η decrease the t distribution becomes more heavy tailed, with longer tails, expecting a more frequent occurrence of extreme events. Since at low values of η events around the median should be less frequent than events in the tails, every new non-extreme observation should contain more information on a potential tail index movement towards Gaussianity. This is taken into account by the score which increases the tail index and pushes the distribution more towards normality.

Given a tail index value of η a distribution has only $k < \eta$ finite moments. For example, for $\eta = 1$ the t distribution becomes a Cauchy distribution which doesn't have finite variance. For this reason when modelling a dynamic tail index previous studies have tried to restrict η_t not to fall below either 2 or 1. The easier way to do so is to modify the link function $\eta = \eta^\dagger + e^{\eta_s \vartheta}$ so that $\eta^\dagger = 1, 2$. The problem in doing this is that it creates distortions to the score function which for $\eta^\dagger = 1$ becomes as in Figure A.12.

Under these conditions, counter intuitively, score response towards new realization decreases as the tail index parameter approaches 1. This means that if the tail index parameter is around 1.5, despite the fact that the conditional distribution is quite heavy tailed at that point, the tail index parameter is much less responsive to movements of ε_t, taking much more time to go back to normality even if the majority of the new observations are close to the median of the distribution. As showed in Figure A.12 this effect can be mitigated by standardising the score by the information quantity $\mathcal{I}_{\vartheta\vartheta}$[26], however the issue now is that, as η approaches 1, the response of the score to new realizations is very high and can move the tail index parameter very rapidly towards infinity.

In order to better understand the implications for the score function of bounding the tail index and standardising it by the information quantity, we have made a simulation study. We have generated data from a conditional symmetric t distribution with a dynamic DCS model for scale and degrees of freedom with dynamics given by Equation (4) with $\omega_\lambda = -4.7$, $\phi_\lambda = 0.985$, $\kappa_\lambda = 0.03$, $\omega_\vartheta = (1/\eta_s)\log(2 - \eta^\dagger)$,

[26] The idea of standardising by the information matrix is not new in the score driven literature, Harvey (2013) and Creal et al. (2013) have already proposed this correction to the score on the line of the method of scoring. However, given that in general while modelling location or scale parameters the information matrices with respect to these parameters are only dependent on the shape parameters of the conditional distribution, if these are static then also the information matrix is time invariant. This means that the standardization simply results in scaling the time varying scores by a constant factor having little or no effect on the score response. On the other hand, if when the shape parameters are time varying, like the tail index, it makes a significant difference which can be appreciated in the current study.

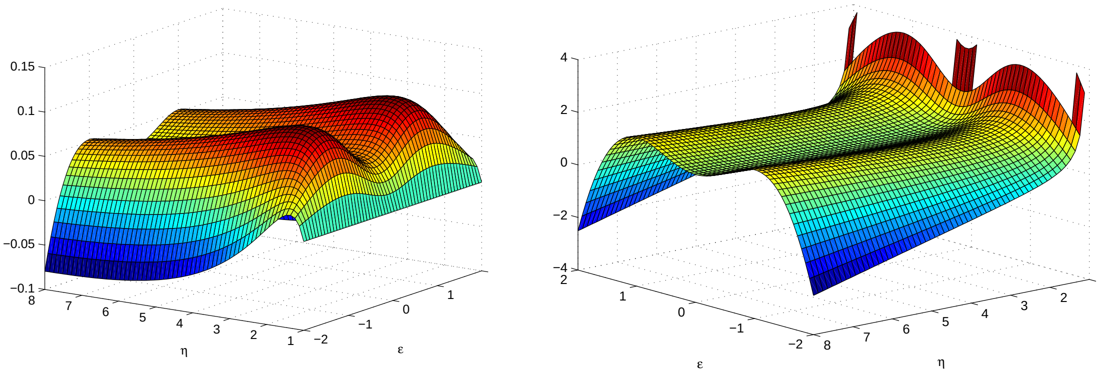

Figure A.12: Three dimensional surface of the score of the bounded tail index parameter with lower-bound $\eta^{\dagger} = 1$ for residuals $-2 < \varepsilon < 2$, and $1 < \eta < 8$, unstandardised (Left) and standardised (Right) by the information quantity.

$\phi_{\vartheta} = 0.99$ and $\kappa_{\vartheta} = 0.025$. In Table A.2 we can see the results of four simulations between bounding or not the tail index with $\eta^{\dagger} = 1$ and standardising the score it by its information quantity[27]. In each of the cases the results reported are the average across $N = 1000$ simulations of length $t = 2000$. From those results it is possible to see that bounding the tail index makes its dynamics even less responsive to variations in ε_t, indeed both the range and the standard deviation of the simulated paths of the tail index decreases. On the other hand standardising the score when the tail index is bounded by 1 makes the score function very responsive to ε_t up to approaching an explosive behaviour which in 20% of the simulations pushes η towards very large positive numbers, approaching infinity. On the other hand, in both the cases of not bounding the tail index less than 1% of the simulated η happen to fall below the bound of 1 and even in these cases the magnitude of the average violation below 1 is around 0.06. This is due to the tendency of the score function to push the η higher when is already low, therefore the number of bound violations is marginal on average and very little in magnitude even without bounding. Looking at the results for Range and at Figure A.13 the unbounded standardised score function is the one most responsive without becoming explosive.

For al these reasons we suggest that bounding the tail index is not advisable. However if it is found to be necessary, one should do it by standardising the score by the information quantity. In any case our preference is to model dynamic tail indexes with an unbounded tail index and a standardised score which appears to be the most flexible and reliable model specification.

Appendix B. Expectations of functions of Beta functions

Given Lemma 1 in Harvey (2013), pg 23, a random variable b distributed with a $beta\,(\alpha,\,\beta)$[28] and $w\,(b)$ is a function of a b with finite expectation,

$$E\left[b^h\,(1-b)^k\,w\,(b)\right] = \frac{B\,(\alpha+h,\,\beta+k)}{B\,(\alpha,\,\beta)} E\,[w\,(b)]\,, \qquad h > -\alpha,\, k > -\beta$$

[27]the results are identical from either setting $\eta_s = 1$ or $\eta_s = -1$
[28]this means that $1-b$ is distributed with a $beta\,(\beta,\,\alpha)$

$\omega_\vartheta = \log 2$	$\eta^\dagger = 1$ Unstand Score	Stand Score	$\eta^\dagger = 0$ Unstand Score	Stand Score
Mean	2,001	2,777	2,006	2,095
Std	0,041	1,371	0,159	0,662
Min	1,882	1,104	1,530	0,905
Max	2,092	11,064	2,351	4,506
Range	0,210	9,960	0,821	3,601
Avg n. per sym $\eta \leq 1$	0,000	0,000	0,001	28,679
% $\eta \leq 1$	0,000	0,000	0,000	0,014
Avg $\left\|\eta_{t\|t-1} - 1\right\| \leq 1$	-	-	0,001	0,067
Avg n. per sym $\eta \geq 100$	0,000	19,334	0,000	0,000
% $\eta \geq 100$	0,000	0,010	0,000	0,000

Table A.2: Results of simulating from a Dynamic Scale-Tail DCS model considering the bounding and standardising the score by the info matrix.

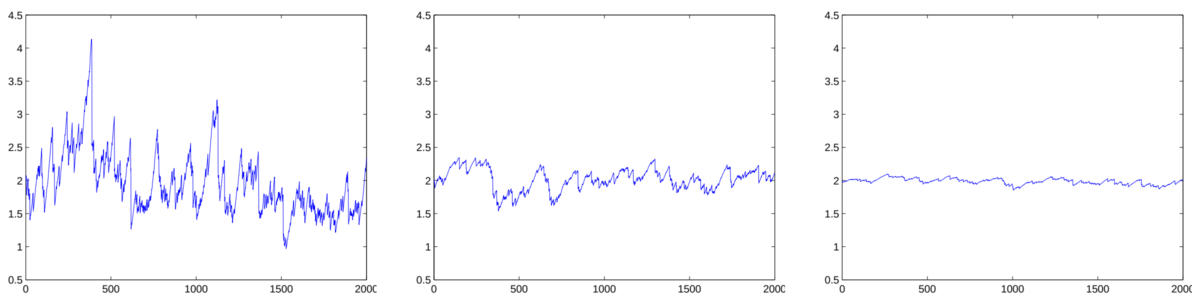

Figure A.13: Simulated dynamic tail index patters with a Scale-Tail DCS model, with unbounded and standardised score (Left), unbounded and unstandardised score (Middle), bounded and unstandardised score (Right).

where $B(\alpha, \beta)$ is a beta function and now the expectation on the right-hand side is now understood to be with respect to a $beta\,(\alpha + h,\, \beta + k)$ distribution. Then

$$\begin{aligned} E\left[\ln^h(b)\ln^k(1-b)\right] &= \int_0^1 \ln^h(x)\ln^k(1-x)\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\,\beta)}dx \\ &= \frac{1}{B(\alpha,\,\beta)}\int_0^1 \frac{\partial^h}{\partial\alpha^h}\frac{\partial^k}{\partial\beta^k}x^{\alpha-1}(1-x)^{\beta-1}\,dx \\ &= \frac{1}{B(\alpha,\,\beta)}\frac{\partial^h}{\partial\alpha^h}\frac{\partial^k}{\partial\beta^k}B(\alpha,\,\beta) \end{aligned}$$

Bearing in mind that $\frac{\partial}{\partial\alpha}B(\alpha,\beta) = B(\alpha,\beta)\left[\psi(\alpha) - \psi(\alpha+\beta)\right]$, and $\frac{\partial^{(l)}}{\partial\alpha^{(l)}}\psi(\alpha) = \psi^{(l)}(\alpha)$ which are the digamma and multigamma functions respectively.

Appendix C. Derivation of the score and the information matrix with respect to the Tail Index Parameter

Given a link function for the tail index of the form $\eta_{t|t-1} = \eta^\dagger + e^{\eta_s \vartheta_{t|t-1}}$, its derivative with respect to ϑ is $\partial\eta/\partial\vartheta = \eta_s\left(\eta - \eta^\dagger\right)$.

Now given the function $K(\eta,\, \upsilon) = \frac{\upsilon}{2\eta^{\frac{1}{\upsilon}}} \frac{1}{B\left(\frac{1}{\upsilon},\, \frac{\eta}{\upsilon}\right)}$, its derivative with respect to ϑ is,

$$\begin{aligned}\frac{\partial K}{\partial \vartheta} &= \frac{\upsilon}{2}\left(-\frac{1}{\eta^{\frac{1}{\upsilon}}} \frac{\left(\psi\left(\frac{\eta}{\upsilon}\right) - \psi\left(\frac{\eta+1}{\upsilon}\right)\right)}{B\left(\frac{1}{\upsilon},\, \frac{\eta}{\upsilon}\right)} - \frac{1}{\upsilon}\frac{1}{\eta^{\frac{1+\upsilon}{\upsilon}}} \frac{\left(\eta - \eta^{\dagger}\right)\eta_s}{B\left(\frac{1}{\upsilon},\, \frac{\eta}{\upsilon}\right)}\right) \\ &= K(\eta,\, \upsilon)\, \frac{\eta_s\left(\eta - \eta^{\dagger}\right)}{\upsilon}\left[\psi\left(\frac{\eta+1}{\upsilon}\right) - \psi\left(\frac{\eta}{\upsilon}\right) - \frac{1}{\eta}\right]\end{aligned}$$

Then the log likelihood function of the Generalised t distribution for a single observation is

$$\ln f(y_t) = -\ln\varphi + \ln K\left(\eta_{t|t-1},\, \upsilon\right) - \frac{\eta_{t|t-1} - 1}{\upsilon}\ln\left(1 + \frac{1}{\eta_{t|t-1}}\left|\frac{y_t - \mu}{\varphi}\right|^{\upsilon}\right)$$

Then its derivative with respect to ϑ is

$$\begin{aligned}\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}} &= \frac{K\left(\eta_{t|t-1},\, \upsilon\right)\frac{\eta_s\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon}\left[\psi\left(\frac{\eta_{t|t-1}+1}{\upsilon}\right) - \psi\left(\frac{\eta_{t|t-1}}{\upsilon}\right) - \frac{1}{\eta_{t|t-1}}\right]}{K\left(\eta_{t|t-1},\, \upsilon\right)} - \\ &- \left[\frac{\eta_s\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon}\ln\left(1 + \frac{1}{\eta_{t|t-1}}\left|\frac{y_t-\mu}{\varphi}\right|^{\upsilon}\right) - \frac{\left(\eta_{t|t-1}+1\right)}{\upsilon}\frac{\eta_s\left(\eta_{t|t-1} - \eta^{\dagger}\right)\left|\frac{y_t-\mu}{\varphi}\right|^{\upsilon}}{\eta_{t|t-1}^2\left(1 + \frac{1}{\eta_{t|t-1}}\left|\frac{y_t-\mu}{\varphi}\right|^{\upsilon}\right)}\right] \\ &= \frac{\eta_s\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon}\left[\psi\left(\frac{\eta_{t|t-1}+1}{\upsilon}\right) - \psi\left(\frac{\eta_{t|t-1}}{\upsilon}\right) + \ln(1-b_t) + \frac{\left(\eta_{t|t-1}+1\right)}{\eta_{t|t-1}}b_t - \frac{1}{\eta_{t|t-1}}\right] \\ &= u_t^{\vartheta}\end{aligned}$$

Where $b_t = \frac{|\varepsilon_t|^{\upsilon}/\eta_{t|t-1}}{1+|\varepsilon_t|^{\upsilon}/\eta_{t|t-1}}$ and $\varepsilon_t = \frac{y_t-\mu}{\varphi}$. If ε_t is distributed Generalised t with shape parameters $\eta_{t|t-1}$ and υ, then b_t is distributed with a $beta\left(1/\upsilon,\, \eta_{t|t-1}/\upsilon\right)$. Then bearing in mind that

$$\frac{\partial\, b_t}{\partial \vartheta_{t|t-1}} = -\frac{|\varepsilon_t|^{\upsilon}}{\eta_{t|t-1}^2}\frac{\eta_s\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\left(1 + \frac{|\varepsilon_t|^{\upsilon}}{\eta_{t|t-1}}\right)^2} = -\eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}}b_t(1-b_t)$$

Then the second derivative of u_t with respect to $\vartheta_{t|t-1}$.

$$\begin{aligned}\frac{\partial^2 \ln f_t}{\partial \vartheta_{t|t-1}^2} &= \frac{\partial u_t^{\vartheta}}{\partial \vartheta_{t|t-1}} \\ &= \eta_s\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}} + \eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon}\left[\left(\psi'\left(\frac{\eta_{t|t-1}+1}{\upsilon}\right) - \psi'\left(\frac{\eta_{t|t-1}}{\upsilon}\right)\right)\eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon} + \eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}^2} + \right. \\ &+ \left.\frac{1}{1-b_t}b_t(1-b_t)\,\eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}} - \frac{b_t}{\eta_{t|t-1}^2}\eta_s\left(\eta_{t|t-1} - \eta^{\dagger}\right) - \left(1 + \frac{1}{\eta_{t|t-1}}\right)\eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}}b_t(1-b_t)\right] \\ &= \eta_s^2\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon}\left[\tau + \frac{1}{\eta_s}\frac{\partial\tau}{\partial\vartheta} + \ln(1-b_t) + \left(1 + \frac{1}{\eta_{t|t-1}} + \frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}} - \frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}^2}\right)b_t - \right. \\ &- \left.\left(1 + \frac{1}{\eta_{t|t-1}}\right)\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\eta_{t|t-1}}b_t(1-b_t)\right]\end{aligned}$$

Where $\tau = \psi\left(\frac{\eta_{t|t-1}+1}{\upsilon}\right) - \psi\left(\frac{\eta_{t|t-1}}{\upsilon}\right) - \frac{1}{\eta_{t|t-1}}$ while $\frac{\partial\tau}{\partial\vartheta} = \eta_s\frac{\left(\eta_{t|t-1} - \eta^{\dagger}\right)}{\upsilon}\left[\psi'\left(\frac{\eta_{t|t-1}+1}{\upsilon}\right) - \psi'\left(\frac{\eta_{t|t-1}}{\upsilon}\right) + \frac{\upsilon}{\eta_{t|t-1}^2}\right]$.

Then, taking the expectation

$$E\left[\frac{\partial^2 \ln f_t}{\partial \vartheta^2_{t|t-1}}\right] = \eta_s^2 \frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\upsilon}\left(\frac{1}{\eta_s}\frac{\partial \tau}{\partial \vartheta}\right) + \eta_s^2 \frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\upsilon}\left\{\tau + E\left[\ln\left(1-b_t\right)\right] + \right\}.$$

$$\left(1+\frac{1}{\eta_{t|t-1}}+\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta_{t|t-1}}-\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta^2_{t|t-1}}\right)E\left[b_t\right]-\left(1+\frac{1}{\eta_{t|t-1}}\right)\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta_{t|t-1}}E\left[b_t\left(1-b_t\right)\right]\Bigg]$$

$$= \eta_s^2 \frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\upsilon}\left[\frac{1}{\eta_s}\frac{\partial \tau}{\partial \vartheta}\right] +$$

$$+ \eta_s^2 \frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\upsilon}\left[\left(1+\frac{1}{\eta_{t|t-1}}+\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta_{t|t-1}}-\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta^2_{t|t-1}}\right)\frac{1}{1+\eta_{t|t-1}}-\frac{1}{\eta_{t|t-1}}-\right.$$

$$\left.-\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta_{t|t-1}\left(\eta_{t|t-1}+1+\upsilon\right)}\right]$$

$$= \eta_s^2 \frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\upsilon}\left[\frac{1}{\eta_s}\frac{\partial \tau}{\partial \vartheta}+\frac{\left(\eta_{t|t-1}-\eta^\dagger\right)}{\eta_{t|t-1}}\frac{\left(\eta_{t|t-1}-1\right)\left(\eta_{t|t-1}+1+\upsilon\right)-\eta_{t|t-1}\left(1+\eta_{t|t-1}\right)}{\eta_{t|t-1}\left(1+\eta_{t|t-1}\right)\left(\eta_{t|t-1}+1+\upsilon\right)}\right]$$

Therefore ultimately we have that $-E\left[\frac{\partial^2 \ln f_t}{\partial \vartheta^2_{t|t-1}}\right] = \mathcal{I}_{\vartheta\vartheta}$ in Equation (2). Now looking at the asymmetric Generalised t distribution, given that for $i=1,2$ we have that

$$\frac{\partial \alpha^+}{\partial \vartheta_{it|t-1}} = (-1)^i \alpha^+\left(1-\alpha^+\right)\frac{\eta_{is}\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\upsilon_i}\left[\psi\left(\frac{\eta_{it|t-1}+1}{\upsilon_i}\right)-\psi\left(\frac{\eta_{it|t-1}}{\upsilon_i}\right)-\frac{1}{\eta_{it|t-1}}\right]$$

and

$$\frac{\partial \ln K_{12}\left(\eta_{it|t-1},\, \eta_{jt|t-1}, \upsilon_i,\, \upsilon_j\right)}{\partial \vartheta_{it|t-1}} = -a_i^+\eta_{is}\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\upsilon_i}\left[\psi\left(\frac{\eta_{it|t-1}+1}{\upsilon_i}\right)-\psi\left(\frac{\eta_{it|t-1}}{\upsilon_i}\right)-\frac{1}{\eta_{it|t-1}}\right]$$

where $a_i^+ = \alpha^+$ if $i=1$ and $a_i^+ = 1-\alpha^+$ for $i=2$, the result in Equations (11) and (12) follows. Now taking the second derivative with respect to $\vartheta_{it|t-1}$ we have that

$$\frac{\partial^2 \ln f_t\left(\left|\varepsilon_t^i\right|\right)}{\partial \vartheta^2_{it|t-1}} = \eta_{is}\frac{\partial \ln f_t\left(\left|\varepsilon_t^i\right|\right)}{\partial \vartheta_{it|t-1}} + \eta_{is}\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\upsilon_i}\left[\frac{\partial a_i^+}{\partial \vartheta_{it|t-1}}\tau_i + a_i^+\frac{\partial \tau_i}{\partial \vartheta_{it|t-1}} + \eta_{is}\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\eta_{it|t-1}}b_t^i-\right.$$

$$\left.-\eta_{is}\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\eta^2_{it|t-1}}b_t^i-\eta_{is}\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\eta_{it|t-1}}\left(1+\frac{1}{\eta_{it|t-1}}\right)b_t^i\left(1-b_t^i\right)\right]$$

$$= \eta_{is}^2\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\upsilon_i}\left[a_i^+\tau_i + a_i^+\frac{1}{\eta_{is}}\frac{\partial \tau_i}{\partial \vartheta_{it|t-1}} - a_i^+\left(1-a_i^+\right)\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\upsilon_i}\tau_i^2\right]+$$

$$+ \eta_{is}^2\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\upsilon_i}\left\{\ln\left(1-b_t^i\right)+\left[\left(1+\frac{1}{\eta_{it|t-1}}\right)+\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\eta_{it|t-1}}\left(1-\frac{1}{\eta_{it|t-1}}\right)\right]b_t^i-\right.$$

$$\left.-\frac{\left(\eta_{it|t-1}-\eta_i^\dagger\right)}{\eta_{it|t-1}}\left(1+\frac{1}{\eta_{it|t-1}}\right)b_t^i\left(1-b_t^i\right)\right\}$$

Than taking the unconditional expectation

$$\left[\frac{\partial^2 \ln f_t\left(\left|\varepsilon_t^i\right|\right)}{\partial \vartheta_{it|t-1}^2}\right] = \eta_{is}^2 \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\left[a_i^{+}\tau_i + a_i^{+}\frac{1}{\eta_{is}}\frac{\partial \tau_i}{\partial \vartheta_{it|t-1}} - a_i^{+}\left(1-a_i^{+}\right)\frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\tau_i^2\right] +$$

$$+ \eta_{is}^2 \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\left[\psi\left(\frac{\eta_{it|t-1}}{\upsilon_i}\right) - \psi\left(\frac{\eta_{it|t-1}+1}{\upsilon_i}\right) + \frac{1}{\eta_{it|t-1}} + \right.$$

$$\left. + \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\eta_{it|t-1}^2}\frac{\left(\eta_{it|t-1}-1\right)}{\left(\eta_{it|t-1}+1\right)} - \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\eta_{it|t-1}\left(\eta_{it|t-1}+1+\upsilon_i\right)}\right]$$

$$= \eta_{is}^2 \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\left[a_i^{+}\frac{1}{\eta_{is}}\frac{\partial \tau_i}{\partial \vartheta_{it|t-1}} - a_i^{+}\left(1-a_i^{+}\right)\frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\tau_i^2 - \left(1-a_i^{+}\right)\tau_i\right] +$$

$$+ \eta_{is}^2 \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\eta_{it|t-1}^2}\frac{\eta_{it|t-1}\left(\upsilon_i-1\right) - \left(\upsilon_i+1\right)}{\left(\eta_{it|t-1}+1\right)\left(\eta_{it|t-1}+1+\upsilon_i\right)}$$

while

$$\frac{\partial^2 \ln f_t\left(\left|\varepsilon_t^j\right|\right)}{\partial \vartheta_{it|t-1}^2} = \eta_{is}^2 \frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\left[a_i^{+}\frac{1}{\eta_{is}}\frac{\partial \tau_i}{\partial \vartheta_{it|t-1}} - a_i^{+}\left(1-a_i^{+}\right)\frac{\left(\eta_{it|t-1} - \eta_i^{\dagger}\right)}{\upsilon_i}\tau_i^2 + a_i^{+}\tau_i\right]$$

where $i = 1,2$ and $j = 1,2$ for $i \neq j$. Ultimately we have that $-E\left[\frac{\partial^2 \ln f_t\left(\left|\varepsilon_t^i\right|\right)}{\partial \vartheta_{it|t-1}^2}\right] = \mathcal{I}_{\vartheta\vartheta}^i$ in Equations (13) and (14).

Appendix D. Derivation of the basic LM test

The derivation below is essentially as in Harvey (2013), sub-section 2.5.1, but stated in terms of ϑ. Let the bold face vector $\boldsymbol{\theta}$ denote other fixed parameters, including ω_ϑ, and let $\boldsymbol{\kappa}_\vartheta' = (\kappa_{\vartheta 0}, \kappa_{\vartheta 1}, ..., \kappa_{\vartheta P-1})$. from the dynamic equation of $\vartheta_{t|t-1}$ The LM test statistic is given by

$$LM_u(P) = \frac{1}{T}\begin{bmatrix}\mathbf{0}' & \partial \ln L/\partial \boldsymbol{\kappa}_\vartheta'\end{bmatrix}\begin{bmatrix}\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} & \boldsymbol{\Psi}_{\boldsymbol{\theta\kappa}} \\ \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} & \boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0} \\ \partial \ln L/\partial \boldsymbol{\kappa}_\vartheta\end{bmatrix}, \tag{D.1}$$

where $\boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}$ denotes the information matrix for $\boldsymbol{\kappa}_\vartheta$ for a single observation, $\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}}$ is the corresponding matrix for $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}_{\boldsymbol{\theta\kappa}}$ is the cross-product matrix. All of these matrices are evaluated at $\boldsymbol{\kappa}_\vartheta = \mathbf{0}$, as is the score vector $\partial \ln L/\partial \boldsymbol{\kappa}_\vartheta$. For the illustration of the simple test we assume that all the other fixed parameters in $\boldsymbol{\theta}$ besides ω_ϑ are calibrated rather than estimated[29]

Now for the t-th observation

$$\frac{\partial \ln f_t}{\partial \boldsymbol{\kappa}_\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}_\vartheta}$$

[29]For example $\upsilon = 2$ which would imply that the the DCS model assumes a Dynamic Conditional t-distribution for y_t.

and so

$$
\begin{aligned}
\boldsymbol{\Psi}_{\kappa\kappa} &= E\left[\frac{\partial \ln f_t}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \ln f_t}{\partial \boldsymbol{\kappa}'_\vartheta}\right]_{\kappa=0} = EE_{t-1}\left[\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'_\vartheta}\right] \\
&= E\left[E_{t-1}\left[\left(\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\right)^2\right]\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'_\vartheta}\right] \\
&= E\left[\left(\frac{\partial \ln f_t}{\partial \vartheta}\right)^2\right]E\left[\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'_\vartheta}\right] \\
&= \mathcal{I}_{\vartheta\vartheta}E\left[\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'}\right].
\end{aligned}
$$

Under the null hypothesis, the conditional expectation of the squared score, $\sigma^2_{\vartheta u}$, is fixed and hence equal to the information quantity in the static model.

We have

$$
\frac{\partial \vartheta_{t+1|t}}{\partial \kappa_{\vartheta j}} = \sum_{i=0}^{P-1}\kappa_{\vartheta i}\frac{\partial u^{\vartheta}_{t-i}}{\partial \kappa_{\vartheta j}} + u^{\vartheta}_{t-j}, \qquad j = 0, ..., P-1,
$$

or

$$
\frac{\partial \vartheta_{t+1|t}}{\partial \boldsymbol{\kappa}_\vartheta} = \boldsymbol{\kappa}'_\vartheta\frac{\partial \mathbf{u}^{\vartheta}_t}{\partial \boldsymbol{\kappa}_\vartheta} + \mathbf{u}^{\vartheta}_t,
$$

where $\mathbf{u}^{\vartheta}_t = (u^{\vartheta}_t, u^{\vartheta}_{t-1}, ..., u^{\vartheta}_{t-P-1})'$, but under the null hypothesis $\boldsymbol{\kappa}_\vartheta = \mathbf{0}$ and so $\partial \vartheta_{t+1|t}/\partial \boldsymbol{\kappa}_\vartheta = \mathbf{u}^{\vartheta}_t$. Hence

$$
E\left(\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'_\vartheta}\right) = \sigma^2_{\vartheta u}\mathbf{I}_P,
$$

where $\mathbf{I}_P$ is a $P \times P$ identity matrix, and $\boldsymbol{\Psi}_{\boldsymbol{\kappa}\boldsymbol{\kappa}} = \sigma^4_{\vartheta u}\mathbf{I}_P = \mathcal{I}^2_{\vartheta\vartheta}\mathbf{I}_P$. Furthermore[30]

$$
\begin{aligned}
E\left[\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}}\frac{\partial \ln f_t}{\partial \boldsymbol{\kappa}'_\vartheta}\right]_{\kappa_\vartheta=0} &= EE_{t-1}\left[\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}}\frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'_\vartheta}\right] \\
&= E\left[\frac{\partial \ln f_t}{\partial \boldsymbol{\theta}}\frac{\partial \ln f_t}{\partial \vartheta}\right]E\left[\frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\kappa}'_\vartheta}\right] = \mathbf{0}
\end{aligned}
$$

Note that because ω_ϑ appears in the dynamic equation

$$
\frac{\partial \ln f_t}{\partial \omega_\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\frac{\partial \vartheta_{t|t-1}}{\partial \omega_\vartheta}
$$

but under the null hypothesis $\partial \vartheta_{t|t-1}/\partial \omega_\vartheta = 1$. Hence $E(\partial \ln f_t/\partial \omega_\vartheta . \partial \ln f_t/\partial \vartheta) = E(u_{\vartheta t})^2 = \sigma^2_{\vartheta u} = \mathcal{I}_{\vartheta\vartheta}$. Thus $\boldsymbol{\Psi}_{\boldsymbol{\theta}\boldsymbol{\kappa}} = \mathbf{0}$ and so

$$
LM_u(P) = \frac{1}{T}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}'_\vartheta}\boldsymbol{\Psi}^{-1}_{\boldsymbol{\kappa}\boldsymbol{\kappa}}\frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta}. \tag{D.2}
$$

[30] Note that because ω_ϑ, which is treated as an element of $\boldsymbol{\vartheta}$, appears directly in the dynamic equation,

$$
\frac{\partial \ln f_t}{\partial \omega_\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}\frac{\partial \vartheta_{t|t-1}}{\partial \omega_\vartheta} = \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}}.1.
$$

On substituting for $\boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}$ and noting that

$$\frac{\partial \ln L}{\partial \kappa_{\vartheta j}} = \sum \frac{\partial \ln f_t}{\partial \vartheta_{t|t-1}} \frac{\partial \vartheta_{t|t-1}}{\partial \kappa_{\vartheta j}} = \sum u_t^{\vartheta} u_{t-1-j}^{\vartheta}, \quad j = 0, 1, ..., P-1,$$

the Q-statistic, Equation (5), is obtained.

Appendix E. Derivation of the full LM test

When some of the other parameters are time-varying or are time-invariant but have to be estimated, the LM test becomes more complicated.

Given the block matrix decomposition in Equation (D.1) we have that.

$$LM_u(P) = \frac{1}{T} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_{\vartheta}'} \boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}^{-1} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_{\vartheta}} + \frac{1}{T} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_{\vartheta}'} \left[\boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} \left(\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}} - \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}}' \boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}} \right)^{-1} \boldsymbol{\Psi}_{\boldsymbol{\kappa\theta}}' \boldsymbol{\Psi}_{\boldsymbol{\kappa\kappa}}^{-1} \right] \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_{\vartheta}}, \qquad \text{(E.1)}$$

In our case, the Generalised-t distribution has also an additional parameter υ to be estimated. For these reasons, assuming also a first order dynamics for $\lambda_{t|t-1}$, our $\boldsymbol{\theta}$ vector is defined as, $\boldsymbol{\theta} = \left(\upsilon, \boldsymbol{\psi}_{\lambda}', \omega_{\vartheta}\right)'$, where $\boldsymbol{\psi}_{\lambda}$ is the vector that contains the parameters that govern the dynamics of $\lambda_{t|t-1}$, $\boldsymbol{\psi}_{\lambda}' = (\omega_{\lambda}, \phi_{\lambda}, \kappa_{\lambda})'$. Starting from deriving the scores with respect to the fixed parameters, we have that

$$\begin{aligned}
\frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \upsilon} &= \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \vartheta_{t|t-1}} \frac{\partial \vartheta_{t|t-1}}{\partial \upsilon} + \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \upsilon} \\
\frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\psi}_{\lambda}} &= \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}_{\lambda}} + \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \vartheta_{t|t-1}} \frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\psi}_{\lambda}} \\
\frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\psi}_{\vartheta}} &= \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}_{\vartheta}} + \frac{\partial \ln f_t\left(y_t \mid Y_{t-1}; \boldsymbol{\theta}\right)}{\partial \vartheta_{t|t-1}} \frac{\partial \vartheta_{t|t-1}}{\partial \boldsymbol{\psi}_{\vartheta}}
\end{aligned}$$

However, under the null hypothesis of $\boldsymbol{\kappa}_{\vartheta} = \mathbf{0}$, η is estimated as fixed and is independent from $\lambda_{t|t-1}$ and υ, therefore we have that $\partial \vartheta_{t|t-1} / \partial \boldsymbol{\psi}_{\lambda} = \partial \vartheta_{t|t-1} / \partial \upsilon = 0$. Moreover if, as in Harvey and Thiele (2016), we assume that the dynamic parameter $\lambda_{t|t-1}$ is previously fitted with a univariate DCS-EGARCH model, $\hat{\lambda}_{t|t-1}$ would also be independent from the parameters governing the dynamics of η and therefore $\partial \lambda_{t|t-1} / \partial \boldsymbol{\kappa}_{\vartheta} = 0$.

The first element of the block matrix in Equation (D.1) is

$$\boldsymbol{\Psi}_{\boldsymbol{\theta\theta}'} = E\left[\frac{\partial \ln L}{\partial \boldsymbol{\theta}} \frac{\partial \ln L}{\partial \boldsymbol{\theta}'}\right] = \begin{pmatrix} \boldsymbol{\Psi}_{\upsilon\upsilon} & \boldsymbol{\Psi}_{\upsilon\boldsymbol{\psi}_{\lambda}'} & \boldsymbol{\Psi}_{\upsilon\omega_{\vartheta}} \\ \boldsymbol{\Psi}_{\boldsymbol{\psi}_{\lambda}\upsilon} & \boldsymbol{\Psi}_{\boldsymbol{\psi}_{\lambda}\boldsymbol{\psi}_{\lambda}'} & \boldsymbol{\Psi}_{\boldsymbol{\psi}_{\lambda}\omega_{\vartheta}} \\ \boldsymbol{\Psi}_{\omega_{\vartheta}\upsilon} & \boldsymbol{\Psi}_{\omega_{\vartheta}\boldsymbol{\psi}_{\lambda}'} & \boldsymbol{\Psi}_{\omega_{\vartheta}\omega_{\vartheta}} \end{pmatrix}$$

The central element $\boldsymbol{\Psi}_{\boldsymbol{\psi}_{\lambda}\boldsymbol{\psi}_{\lambda}'}$ is nothing more than the information matrix with respect to the dynamic parameters of a DCS-EGARCH model for $\lambda_{t|t-1}$ with first-order dynamic.

$$\mathbf{\Psi}_{\boldsymbol{\psi}_\lambda \boldsymbol{\psi}'_\lambda} = E\left[\frac{\partial \ln f_t(y_t \mid Y_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_\lambda}\frac{\partial \ln f_t(y_t \mid Y_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\psi}'_\lambda}\right] = \mathcal{I}_{\lambda\lambda}\mathbf{D}(\boldsymbol{\psi_\lambda}), \tag{E.2}$$

where

$$\mathbf{D}(\boldsymbol{\psi_\lambda}) = \frac{1}{1-b}\begin{bmatrix} A & D & E \\ D & B & F \\ E & F & C \end{bmatrix}, \qquad b < 1,$$

as in Harvey (2013, p 37). The formulae for A to F are

$$\begin{aligned} A &= \mathcal{I}_{\lambda\lambda}, \qquad B = \frac{\kappa_\lambda^2\mathcal{I}_{\lambda\lambda}^2(1+a\phi_\lambda)}{(1-\phi_\lambda^2)(1-a\phi_\lambda)}, \qquad C = \frac{(1-\phi_\lambda)^2(1+a)}{1-a}, \\ D &= \frac{a\kappa_\lambda\mathcal{I}_{\lambda\lambda}^2}{1-a\phi_\lambda}, \qquad E = \frac{c(1-\phi_\lambda)}{1-a} \quad \text{and} \quad F = \frac{ac\kappa_\lambda(1-\phi_\lambda)}{(1-a)(1-a\phi_\lambda)}, \end{aligned}$$

and

$$a = E_{t-1}(x_t) = \phi_\lambda + \kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \lambda_{t|t-1}}\right) = \phi_\lambda + \kappa_\lambda E\left(\frac{\partial u_t^\lambda}{\partial \lambda}\right) = \phi_\lambda - \kappa_\lambda \mathcal{I}_{\lambda\lambda}$$

$$b = E_{t-1}(x_t^2) = \phi_\lambda^2 - 2\phi_\lambda\kappa_\lambda\mathcal{I}_{\lambda\lambda} + \kappa_\lambda^2 E\left(\frac{\partial u_t^\lambda}{\partial \lambda}\right)^2 \geq 0, \tag{E.3}$$

$$c = E_{t-1}(u_t^\lambda x_t) = \kappa_\lambda E\left(u_t^\lambda \frac{\partial u_t^\lambda}{\partial \lambda}\right), \tag{E.4}$$

with

$$x_t = \phi_\lambda + \kappa_\lambda \frac{\partial u_t^\lambda}{\partial \lambda_{t|t-1}}, \quad t = 1, ..., T. \tag{E.5}$$

The unconditional expectations can then replace the conditional ones because of the assumption that they do not depend on $\lambda_{t|t-1}$, as per Condition 2 in Harvey (2013), p. 35.

Then looking now at the expectation of product of the score with respect to $\boldsymbol{\psi}_\lambda$ and υ.

$$\begin{aligned} \mathbf{\Psi}_{\boldsymbol{\psi}_\lambda \upsilon} &= E\left[\frac{\partial \ln f_t(y_t \mid Y_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_\lambda}\frac{\partial \ln f_t(y_t \mid Y_{t-1};\boldsymbol{\theta})}{\partial \upsilon}\right] \\ &= \mathcal{I}_{\lambda\lambda}E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] + \mathcal{I}_{\lambda\upsilon}E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda}\right] \end{aligned}$$

where for $\lambda_{t+1|t}$ starting in the infinite past, and given $|a| < \infty$, $E\left[\partial\lambda_{t+1|t}/\partial\boldsymbol{\psi}_\lambda\right]$ exist and, as defined in Lemma 6 in Harvey (2013), p. 36, can be expressed as

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda}\right] = E\begin{bmatrix} \frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda} \\ \frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda} \\ \frac{\partial \lambda_{t+1|t}}{\partial \kappa_\lambda} \end{bmatrix} = E\begin{bmatrix} x_t\frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda} + 1 - \phi_\lambda \\ x_t\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda} + \lambda_{t|t-1} - \omega_\lambda \\ x_t\frac{\partial \lambda_{t|t-1}}{\partial \kappa_\lambda} + u_t^\lambda \end{bmatrix} = \begin{pmatrix} \frac{1-\phi_\lambda}{1-a} \\ 0 \\ 0 \end{pmatrix} = \mathbf{d}, \quad t = \ldots, 0, 1, ..., T., \tag{E.6}$$

Consider now the derivative of $\lambda_{t+1|t}$ with respect to υ where $u_t(\theta_{t|t-1})$ indicates that $\lambda_{t|t-1}$ is held fixed

$$\begin{aligned}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} &= \phi_\lambda \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda \frac{\partial u_t^\lambda(\lambda_{t|t-1})}{\partial \upsilon} + \kappa_\lambda \frac{\partial u_t^\lambda}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} \\ &= x_t \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda \frac{\partial u_t^\lambda(\lambda_{t|t-1})}{\partial \upsilon}, \qquad t = 1, ..., T,\end{aligned}$$

here x_t is as in Equation (E.5); see Harvey (2013), p. 35. Taking conditional expectations gives

$$E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) = E_{t-1}\left(x_t \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda \frac{\partial u_t^\lambda(\lambda_{t|t-1})}{\partial \upsilon}\right) = a \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda(\lambda_{t|t-1})}{\partial \upsilon}\right)$$

We can take the unconditional expectation of $\partial u_t^\lambda / \partial \upsilon$ when it does not depend on $\lambda_{t|t-1}$.

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = \frac{\kappa_\lambda}{1-a} E\left[\frac{\partial u_t^\lambda(\lambda_{t|t-1})}{\partial \upsilon}\right] = \frac{-\kappa_\lambda}{1-a}\mathcal{I}_{\lambda\upsilon}., \tag{E.7}$$

Furthermore, dropping $(\lambda_{t|t-1})$ from $u_t^\lambda(\lambda_{t|t-1})$, we can now look at the expectations of the product of the partial derivatives

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = E\begin{bmatrix} \frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \\ \frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \\ \frac{\partial \lambda_{t+1|t}}{\partial \kappa_\lambda} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \end{bmatrix}, \quad t = 0, 1, \dots, T., \tag{E.8}$$

Starting form taking the conditional expectations we have

$$\begin{aligned} E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) &= E_{t-1}\left(x_t^2 \frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) + \kappa_\lambda E_{t-1}\left((1-\phi_\lambda)\frac{\partial u_t^\lambda}{\partial \upsilon}\right) + \\ &\quad + (1-\phi_\lambda) E_{t-1}\left(x_t \frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda}\right) + \kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon} \frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda}\right) \\ &= b \frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda (1-\phi_\lambda) E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\right] + a(1-\phi_\lambda)\frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda} \\ &\quad + \kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right)\frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda} \end{aligned}$$

Given Condition 2 of Harvey (2013, pg 35), $E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right)$ its independent from t. Thus

$$E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right) = \phi_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \upsilon}\right) + \kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right) = -\phi_\lambda \mathcal{I}_{\lambda\upsilon} + \kappa_\lambda E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right] \tag{E.9}$$

Taking now the unconditional expectations given $|b| < 1$ we have that

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = \frac{\kappa_\lambda}{1-b}\left((1-\phi_\lambda)E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\right] + a\frac{(1-\phi_\lambda)}{(1-a)}E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\right] + \frac{(1-\phi_\lambda)}{(1-a)}E\left[x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right]\right) \tag{E.10}$$

$$= \frac{\kappa_\lambda}{1-b}\frac{(1-\phi_\lambda)}{(1-a)}(\kappa_\lambda b_{\lambda\upsilon} - \mathcal{I}_{\lambda\upsilon}), \tag{E.11}$$

where $b_{\lambda\upsilon} = E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right]$.

$$\begin{aligned} E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) &= E_{t-1}\left(x_t^2\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda}\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) + \kappa_\lambda E_{t-1}\left(\left(\lambda_{t|t-1}-\omega_\lambda\right)\frac{\partial u_t^\lambda}{\partial \upsilon}\right) + \\ &\quad + E_{t-1}\left(x_t\left(\lambda_{t|t-1}-\omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) + \kappa_\lambda E_{t-1}\left(x_t\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right) \\ &= b\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda}\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\right]\left(\lambda_{t|t-1}-\omega_\lambda\right) + \\ &\quad + a\left(\lambda_{t|t-1}-\omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E_{t-1}\left(x_t\frac{\partial u_t^\lambda}{\partial \upsilon}\right)\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda}, \end{aligned}$$

Taking conditional expectations at $\mathcal{F}_{t-2}$ of the third term and then unconditional expectations given that $E\left[\lambda_{t|t-1}-\omega_\lambda\right] = 0$ we have

$$\begin{aligned} E_{t-2}\left(\left(\lambda_{t|t-1}-\omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) &= E_{t-1}\left(\left(x_{t-1}\frac{\partial \lambda_{t-1|t-2}}{\partial \upsilon} + \kappa_\lambda\frac{\partial u_{t-1}^\lambda}{\partial \upsilon}\right)\left(\phi_\lambda\left(\lambda_{t-1|t-2}-\omega_\lambda\right) + \kappa_\lambda u_{t-1}^\lambda\right)\right) \\ &= a\phi_\lambda\left(\lambda_{t-1|t-2}-\omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + c\kappa_\upsilon\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\upsilon^2 E\left[u_{t-1}^\lambda\frac{\partial u_{t-1}^\lambda}{\partial \upsilon}\right] \\ &\quad + \kappa_\lambda\phi_\lambda E\left[\frac{\partial u_{t-1}^\lambda}{\partial \upsilon}\right]\left(\lambda_{t-1|t-2}-\omega_\lambda\right) \\ E\left[\left(\lambda_{t|t-1}-\omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right] &= \frac{\kappa_\lambda^2}{1-a\phi_\lambda}\left(E\left[u_{t-1}^\lambda\frac{\partial u_{t-1}^\lambda}{\partial \upsilon}\right] - \frac{c}{1-a}\mathcal{I}_{\lambda\upsilon}\right) \end{aligned}$$

Then taking the unconditional expectation of $E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right)$ we have

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = \frac{a}{(1-a\phi_\lambda)}\frac{\kappa_\lambda^2}{(1-b)}\left(c_\upsilon - \frac{c}{1-a}\mathcal{I}_{\lambda\upsilon}\right) \tag{E.12}$$

where $c_\upsilon = E\left[u_t^\lambda\frac{\partial u_t^\lambda}{\partial \upsilon}\right]$. Then

$$\begin{aligned} E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \kappa_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) &= E_{t-1}\left(x_t^2\frac{\partial \lambda_{t|t-1}}{\partial \kappa_\lambda}\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) + \kappa_\lambda E_{t-1}\left(u_t^\lambda\frac{\partial u_t^\lambda}{\partial \upsilon}\right) + \\ &\quad + E_{t-1}\left(x_t u_t^\lambda\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) + \kappa_\lambda E_{t-1}\left(x_t\frac{\partial \lambda_{t|t-1}}{\partial \kappa_\lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right) \\ &= b\frac{\partial \lambda_{t|t-1}}{\partial \kappa_\lambda}\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E\left[u_t^\lambda\frac{\partial u_t^\lambda}{\partial \upsilon}\right] + \\ &\quad + E_{t-1}\left(x_t u_t^\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E_{t-1}\left(x_t\frac{\partial u_t^\lambda}{\partial \upsilon}\right)\frac{\partial \lambda_{t|t-1}}{\partial \kappa_\lambda} \end{aligned}$$

which after taking unconditional expectations becomes

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \kappa_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = \frac{\kappa_\lambda}{1-b}\left(c_\upsilon - \frac{c}{1-a}\mathcal{I}_{\lambda\upsilon}\right)$$

Therefore

$$\begin{aligned}
\mathbf{\Psi}_{\boldsymbol{\psi}_\lambda \upsilon} &= \frac{\kappa_\lambda}{(1-b)} \mathcal{I}_{\lambda\lambda} \begin{pmatrix} \kappa_\lambda \frac{(1-\phi_\lambda)}{(1-a)} b_{\lambda\upsilon} \\ \frac{a\kappa_\lambda}{(1-a\phi_\lambda)} c_\upsilon \\ c_\upsilon \end{pmatrix} - \mathcal{I}_{\lambda\upsilon} \begin{pmatrix} \frac{(1-\phi_\lambda)}{(1-a)} \left(1 - \frac{\kappa_\lambda}{(1-b)} \mathcal{I}_{\lambda\lambda}\right) \\ \frac{ac\kappa_\lambda^2}{(1-a)(1-b)(1-a\phi_\lambda)} \mathcal{I}_{\lambda\lambda} \\ \frac{c\kappa_\lambda}{(1-a)(1-b)} \mathcal{I}_{\lambda\lambda} \end{pmatrix} \\
&= \frac{\kappa_\lambda}{(1-b)} \mathcal{I}_{\lambda\lambda} \mathbf{g}(\upsilon) + \mathcal{I}_{\lambda\upsilon} \mathbf{d}
\end{aligned}$$

Looking now at the expectation of product of the scores with respect to υ.

$$\begin{aligned}
\mathbf{\Psi}_{\upsilon\upsilon} &= E\left[\frac{\partial \ln f_{t+1}(y_{t+1} \mid Y_t; \boldsymbol{\theta})}{\partial \upsilon} \frac{\partial \ln f_{t+1}(y_{t+1} \mid Y_t; \boldsymbol{\theta})}{\partial \upsilon}\right] \\
&= \mathcal{I}_{\upsilon\upsilon} + \mathcal{I}_{\lambda\lambda} E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] + 2\mathcal{I}_{\lambda\upsilon} E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right]
\end{aligned}$$

Starting from the conditional expectation of $\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}$

$$\begin{aligned}
E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) &= E_{t-1}\left(x_t^2 \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) + \kappa_\lambda^2 E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \upsilon} \frac{\partial u_t^\lambda}{\partial \upsilon}\right) + \\
&\quad + 2\kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right) \\
&= b \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} \frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda^2 E\left[\frac{\partial u_t^\lambda}{\partial \upsilon} \frac{\partial u_t^\lambda}{\partial \upsilon}\right] + 2\kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right) \frac{\partial \lambda_{t|t-1}}{\partial \upsilon}
\end{aligned}$$

Noticing that

$$E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right) = \kappa_\lambda E\left[\frac{\partial u_t^\lambda}{\partial \lambda} \frac{\partial u_t^\lambda}{\partial \upsilon}\right] - \phi_\lambda \mathcal{I}_{\lambda\upsilon}$$

we can then take the unconditional expectation of $E_{t-1}\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right]$ to obtain

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} \frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = \frac{\kappa_\lambda}{1-b}\left[\kappa_\lambda^2 b_{\upsilon\upsilon} - \frac{2\kappa_\lambda^2}{(1-a)} \left(\kappa_\lambda b_{\lambda\upsilon} - \phi_\lambda \mathcal{I}_{\lambda\upsilon}\right) \mathcal{I}_{\lambda\upsilon}\right] \tag{E.13}$$

where $b_{\upsilon\upsilon} = E\left(\frac{\partial u_t^\lambda}{\partial \upsilon}\right)^2$. Therefore

$$\mathbf{\Psi}_{\upsilon\upsilon} = \mathcal{I}_{\upsilon\upsilon} + \frac{\kappa_\lambda^2}{(1-b)} \mathcal{I}_{\lambda\lambda} \left[b_{\upsilon\upsilon} - \frac{2\kappa_\lambda}{(1-a)} \mathcal{I}_{\lambda\upsilon} b_{\lambda\upsilon}\right] - \kappa_\lambda \left[\frac{2}{(1-a)} + \frac{\phi_\lambda \kappa_\lambda}{(1-b)} \mathcal{I}_{\lambda\lambda}\right] \mathcal{I}_{\lambda\upsilon}^2$$

Now keeping in mind that $\partial \vartheta_{t+1|t} / \partial \omega_\vartheta = 1$ we can focus on the blocks which include the partial derivative with respect to ω_ϑ

$$\begin{aligned}
\mathbf{\Psi}_{\upsilon\omega_\vartheta} &= E\left[\frac{\partial \ln f_{t+1}(y_{t+1} \mid Y_t;\boldsymbol{\theta})}{\partial \upsilon}\frac{\partial \ln f_{t+1}(y_{t+1} \mid Y_t;\boldsymbol{\theta})}{\partial \omega_\vartheta}\right] \\
&= E\left[E_{t-1}\left(\frac{\partial \ln f_{t+1}}{\partial \lambda_{t+1|t}}\frac{\partial \ln f_{t+1}}{\partial \lambda_{t+1|t}}\right)\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta} + E_{t-1}\left(\frac{\partial \ln f_{t+1}}{\partial \lambda_{t+1|t}}\frac{\partial \ln f_{t+1}}{\partial \vartheta_{t+1|t}}\right)\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} + \right. \\
&\left. + E_{t-1}\left(\frac{\partial \ln f_{t+1}}{\partial \upsilon_{t+1|t}}\frac{\partial \ln f_{t+1}}{\partial \lambda_{t+1|t}}\right)\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta} + E_{t-1}\left(\frac{\partial \ln f_{t+1}}{\partial \upsilon_{t+1|t}}\frac{\partial \ln f_{t+1}}{\partial \vartheta_{t+1|t}}\right)\frac{\partial \vartheta_{t+1|t}}{\partial \omega_\vartheta}\right] \\
&= \mathcal{I}_{\lambda\lambda}E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] + \mathcal{I}_{\lambda\vartheta}E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] + \mathcal{I}_{\lambda\upsilon}E\left[\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] + \mathcal{I}_{\upsilon\vartheta}
\end{aligned}$$

Consider now the derivative of $\lambda_{t+1|t}$ with respect to ω_ϑ

$$\begin{aligned}
\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta} &= \phi_\lambda \frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda \frac{\partial u_t^\lambda}{\partial \omega_\vartheta} + \kappa_\lambda \frac{\partial u_t^\lambda}{\partial \lambda_{t|t-1}}\frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} \\
&= x_t \frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda \frac{\partial u_t^\lambda}{\partial \vartheta}, \qquad t = 1, ..., T,
\end{aligned}$$

Then, for the same reasons as in Equation (E.7) we have that

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] = \frac{-\kappa_\lambda}{1-a}\mathcal{I}_{\lambda\vartheta}, \tag{E.14}$$

Then, starting from the conditional expectation of $\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}$

$$\begin{aligned}
E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right) &= E_{t-1}\left(x_t^2\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \vartheta}\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \\
&+ \kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \upsilon}\right)\frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda^2 E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \upsilon}\frac{\partial u_t^\lambda}{\partial \vartheta}\right)
\end{aligned}$$

Given Condition 2 of Harvey (2013, pg 35), $E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \vartheta}\right)$ its independent from t. Thus

$$E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \vartheta}\right) = \phi_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \vartheta}\right) + \kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \vartheta}\right) = -\phi_\lambda \mathcal{I}_{\lambda\upsilon} + \kappa_\lambda E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \vartheta}\right]$$

Taking now the unconditional expectations given $|b| < 1$ we have that

$$\begin{aligned}
E\left[\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] &= \frac{\kappa_\lambda^2}{(1-b)}\left(E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\frac{\partial u_t^\lambda}{\partial \vartheta}\right] - \frac{\kappa}{(1-a)}\left(\mathcal{I}_{\lambda\upsilon}E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \vartheta}\right] + \mathcal{I}_{\lambda\vartheta}E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right]\right) + \right. \\
&\left. + \frac{2\phi_\lambda}{(1-a)}\mathcal{I}_{\lambda\upsilon}\mathcal{I}_{\lambda\vartheta}\right)
\end{aligned}$$

Therefore

$$\mathbf{\Psi}_{\upsilon\omega_\vartheta} = \frac{\kappa_\lambda^2}{(1-b)}\left[b_{\upsilon\vartheta} - \frac{\kappa_\lambda}{(1-a)}\left(\mathcal{I}_{\lambda\upsilon}b_{\lambda\vartheta} + \mathcal{I}_{\lambda\vartheta}b_{\lambda\upsilon}\right)\right] - \frac{2\kappa_\lambda}{(1-a)}\left(1 - \frac{\phi_\lambda\kappa_\lambda}{(1-b)}\right)\mathcal{I}_{\lambda\upsilon}\mathcal{I}_{\lambda\vartheta} + \mathcal{I}_{\upsilon\upsilon} \tag{E.15}$$

where $b_{\upsilon\vartheta} = E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\frac{\partial u_t^\lambda}{\partial \vartheta}\right]$ and $b_{\lambda\vartheta} = E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \vartheta}\right]$. Then

$$\begin{aligned}
\boldsymbol{\Psi}_{\boldsymbol{\psi}_\lambda \omega_\vartheta} &= E\left[\frac{\partial \ln f_t(y_t \mid Y_{t-1};\boldsymbol{\theta})}{\partial \boldsymbol{\psi}_\lambda}\frac{\partial \ln f_t(y_t \mid Y_{t-1};\boldsymbol{\theta})}{\partial \omega_\vartheta}\right] \\
&= \mathcal{I}_{\lambda\lambda} E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] + \mathcal{I}_{\lambda\vartheta} E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda}\right]
\end{aligned}$$

We can now focus on the expectations of the product of the partial derivatives of the dynamic parameters of $\lambda_{t|t-1}$ and ω_ϑ.

$$E\left[\frac{\partial \lambda_{t+1|t}}{\partial \boldsymbol{\psi}_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] = E\begin{bmatrix} \frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta} \\ \frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta} \\ \frac{\partial \lambda_{t+1|t}}{\partial \kappa_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta} \end{bmatrix}, \quad t = 0, 1, \ldots, T.,$$

Starting form taking the conditional expectations we have

$$\begin{aligned}
E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) &= E_{t-1}\left(x_t^2\right)\frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda}\frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + (1-\phi_\lambda)\kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \omega_\vartheta}\right) + \\
&\quad + (1-\phi_\lambda) E_{t-1}(x_t)\frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda} + \kappa_\lambda E_{t-1}(x_t)\frac{\partial u_t^\lambda}{\partial \vartheta}\frac{\partial \lambda_{t|t-1}}{\partial \omega_\lambda}
\end{aligned}$$

Given Condition 2 of Harvey (2013, pg 35), $E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \vartheta}\right)$ its independent from t. Thus

$$E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \omega_\vartheta}\right) = -\phi_\lambda \mathcal{I}_{\lambda\vartheta} + \kappa_\lambda E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \vartheta}\right]$$

Taking now the unconditional expectations given $|b| < 1$ we have that

$$\begin{aligned}
E\left[\frac{\partial \lambda_{t+1|t}}{\partial \omega_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] &= \frac{\kappa_\lambda}{1-b}\left((1-\phi_\lambda) E\left[\frac{\partial u_t^\lambda}{\partial \vartheta}\right] + a\frac{(1-\phi_\lambda)}{(1-a)} E\left[\frac{\partial u_t^\lambda}{\partial \vartheta}\right] + \frac{(1-\phi_\lambda)}{(1-a)} E\left[x_t \frac{\partial u_t^\lambda}{\partial \vartheta}\right]\right) \\
&= \frac{\kappa_\lambda}{1-b}\frac{(1-\phi_\lambda)}{(1-a)}\left(\kappa_\lambda b_{\lambda\vartheta} - \mathcal{I}_{\lambda\upsilon}\right)
\end{aligned}$$

Then

$$\begin{aligned}
E_{t-1}\left(\frac{\partial \lambda_{t+1|t}}{\partial \phi_\lambda}\frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right) &= E_{t-1}\left(x_t^2\right)\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda}\frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \upsilon}\right)\left(\lambda_{t|t-1} - \omega_\lambda\right) + \\
&\quad + E_{t-1}(x_t)\left(\lambda_{t|t-1} - \omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda E_{t-1}\left(x_t \frac{\partial u_t^\lambda}{\partial \vartheta}\right)\frac{\partial \lambda_{t|t-1}}{\partial \phi_\lambda},
\end{aligned}$$

Taking first the conditional expectations at $\mathcal{F}_{t-2}$ of the third term and then unconditional expectations given that $E\left[\lambda_{t|t-1} - \omega_\lambda\right] = 0$ we have that

$$E\left[\left(\lambda_{t|t-1} - \omega_\lambda\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon}\right] = \frac{\kappa_\lambda^2}{1 - a\phi_\lambda}\left(E\left[u_{t-1}^\lambda \frac{\partial u_{t-1}^\lambda}{\partial \vartheta}\right] - \frac{c}{1-a}\mathcal{I}_{\lambda\vartheta}\right)$$

Then taking the unconditional expectation of $E_{t-1}\left(\frac{\partial\lambda_{t+1|t}}{\partial\phi_\lambda}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right)$ we have

$$E\left[\frac{\partial\lambda_{t+1|t}}{\partial\phi_\lambda}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]=\frac{a}{(1-a\phi_\lambda)}\frac{\kappa_\lambda^2}{(1-b)}\left(c_\vartheta-\frac{c}{1-a}\mathcal{I}_{\lambda\upsilon}\right) \tag{E.16}$$

where $c_\vartheta=E\left[u_t^\lambda\frac{\partial u_t^\lambda}{\partial\vartheta}\right]$. Then

$$\begin{aligned}E_{t-1}\left(\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right) &= E_{t-1}\left(x_t^2\right)\frac{\partial\lambda_{t|t-1}}{\partial\kappa_\lambda}\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}+\kappa_\lambda E_{t-1}\left(u_t^\lambda\frac{\partial u_t^\lambda}{\partial\vartheta}\right)+\\ &\quad +E_{t-1}\left(x_tu_t^\lambda\right)\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}+\kappa_\lambda E_{t-1}\left(x_t\frac{\partial u_t^\lambda}{\partial\vartheta}\right)\frac{\partial\lambda_{t|t-1}}{\partial\kappa_\lambda}\end{aligned}$$

which after taking unconditional expectations becomes

$$E\left[\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]=\frac{\kappa_\lambda}{1-b}\left(c_\vartheta-\frac{c}{1-a}\mathcal{I}_{\lambda\vartheta}\right)$$

Therefore

$$\begin{aligned}\mathbf{\Psi}_{\boldsymbol{\psi}_\lambda\omega_\vartheta} &= \frac{\kappa_\lambda}{(1-b)}\mathcal{I}_{\lambda\lambda}\begin{pmatrix}\kappa_\lambda\frac{(1-\phi_\lambda)}{(1-a)}b_{\lambda\vartheta}\\ \frac{a\kappa_\lambda}{(1-a\phi_\lambda)}c_\vartheta\\ c_\vartheta\end{pmatrix}-\mathcal{I}_{\lambda\vartheta}\begin{pmatrix}\frac{(1-\phi_\lambda)}{(1-a)}\left(1-\frac{\kappa_\lambda}{(1-b)}\mathcal{I}_{\lambda\lambda}\right)\\ \frac{ac\kappa_\lambda^2}{(1-a)(1-b)(1-a\phi_\lambda)}\mathcal{I}_{\lambda\lambda}\\ \frac{c\kappa_\lambda}{(1-a)(1-b)}\mathcal{I}_{\lambda\lambda}\end{pmatrix}\\ &= \frac{\kappa_\lambda}{(1-b)}\mathcal{I}_{\lambda\lambda}\mathbf{g}\left(\vartheta\right)-\mathcal{I}_{\lambda\vartheta}\mathbf{d}\end{aligned}$$

Looking now at the expectation of product of the scores with respect to ω_ϑ.

$$\begin{aligned}\mathbf{\Psi}_{\omega_\vartheta\omega_\vartheta} &= E\left[\frac{\partial\ln f_{t+1}(y_{t+1}\mid Y_t;\boldsymbol{\theta})}{\partial\omega_\vartheta}\frac{\partial\ln f_{t+1}(y_{t+1}\mid Y_t;\boldsymbol{\theta})}{\partial\omega_\vartheta}\right]\\ &= \mathcal{I}_{\vartheta\vartheta}+\mathcal{I}_{\lambda\lambda}E\left[\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]+2\mathcal{I}_{\lambda\vartheta}E\left[\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]\end{aligned}$$

given that the conditional expectation of $\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}$ can be expressed as follows

$$E_{t-1}\left(\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right) = E_{t-1}\left(x_t^2\right)\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}+\kappa_\lambda^2E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial\vartheta}\frac{\partial u_t^\lambda}{\partial\vartheta}\right)+2\kappa_\lambda E_{t-1}\left(x_t\frac{\partial u_t^\lambda}{\partial\vartheta}\right)\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}$$

Noticing that

$$E_{t-1}\left(x_t\frac{\partial u_t^\lambda}{\partial\vartheta}\right)=\kappa_\lambda E\left[\frac{\partial u_t^\lambda}{\partial\lambda}\frac{\partial u_t^\lambda}{\partial\vartheta}\right]-\phi_\lambda\mathcal{I}_{\lambda\vartheta}$$

we can then take the unconditional expectation of $E_{t-1}\left[\frac{\partial\lambda_{t+1|t}}{\partial\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\vartheta}\right]$ to obtain

$$E\left[\frac{\partial\lambda_{t+1|t}}{\partial\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\vartheta}\right]=\frac{\kappa_\lambda}{1-b}\left[\kappa_\lambda^2b_{\vartheta\vartheta}-\frac{2\kappa_\lambda^2}{(1-a)}\left(\kappa_\lambda b_{\lambda\vartheta}-\phi_\lambda\mathcal{I}_{\lambda\vartheta}\right)\mathcal{I}_{\lambda\vartheta}\right] \tag{E.17}$$

where $b_{\vartheta\vartheta} = E\left(\frac{\partial u_t^\lambda}{\partial \vartheta}\right)^2$. Therefore

$$\mathbf{\Psi}_{\omega_\vartheta \omega_\vartheta} = \mathcal{I}_{\vartheta\vartheta} + \frac{\kappa_\lambda^2}{(1-b)}\mathcal{I}_{\lambda\lambda}\left[b_{\vartheta\vartheta} - \frac{2\kappa_\lambda}{(1-a)}\mathcal{I}_{\lambda\vartheta} b_{\lambda\vartheta}\right] - \kappa_\lambda\left[\frac{2}{(1-a)} + \frac{\phi_\lambda \kappa_\lambda}{(1-b)}\mathcal{I}_{\lambda\lambda}\right]\mathcal{I}_{\lambda\vartheta}^2$$

The second element of the block matrix in Equation (D.1) is

$$\mathbf{\Psi}_{\boldsymbol{\kappa}\boldsymbol{\theta}'} = E\left[\frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \ln L}{\partial \boldsymbol{\theta}'}\right] = \left[\mathbf{\Psi}_{\boldsymbol{\kappa}\upsilon}, \mathbf{\Psi}_{\boldsymbol{\kappa}\boldsymbol{\psi}'_\lambda}, \mathbf{\Psi}_{\boldsymbol{\kappa}\omega_\vartheta}\right]$$

its first component can be represented as

$$\begin{aligned}
\mathbf{\Psi}_{\boldsymbol{\kappa}_\vartheta \upsilon} &= E\left[\frac{\partial \ln f_{t+1}(y_{t+1} \mid Y_t; \boldsymbol{\theta})}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \ln f_{t+1}(y_{t+1} \mid Y_t; \boldsymbol{\theta})}{\partial \upsilon}\right] \\
&= E\left[E_{t-1}\left(\frac{\partial \ln f_{t+1}}{\partial \vartheta_{t+1|t}}\frac{\partial \ln f_{t+1}}{\partial \lambda_{t+1|t}}\right)\frac{\partial \vartheta_{t+1|t}}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon} + E_{t-1}\left(\frac{\partial \ln f_{t+1}}{\partial \vartheta_{t+1|t}}\frac{\partial \ln f_{t+1}}{\partial \upsilon}\right)\frac{\partial \vartheta_{t+1|t}}{\partial \boldsymbol{\kappa}_\vartheta}\right] \\
&= \mathcal{I}_{\vartheta\lambda} E\left[\frac{\partial \vartheta_{t+1|t}}{\partial \boldsymbol{\kappa}_\vartheta}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] + \mathcal{I}_{\vartheta\upsilon} E\left[\mathbf{u}_t^\vartheta\right]
\end{aligned}$$

where $E\left[\mathbf{u}_t^\vartheta\right] = \mathbf{0}$. Then starting from taking the conditional expectation of the product of the partial derivatives of $\vartheta_{t|t-1}$ with respect to all the individual $\kappa_{\vartheta i}$, for $1 = 0, 1, ..., P-1$ we have first

$$\begin{aligned}
E_{t-1}\left(\frac{\partial \vartheta_{t+1|t}}{\partial \kappa_{\vartheta 0}}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right) &= E_{t-1}\left(x_t u_t^\vartheta\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E_{t-1}\left(u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \upsilon}\right) \\
&= E_{t-1}\left(\left(\phi_\lambda + \kappa_\lambda \frac{\partial u_t^\lambda}{\partial \lambda}\right) u_t^\vartheta\right)\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + \kappa_\lambda E_{t-1}\left(u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \upsilon}\right) \\
&= \kappa_\lambda\left(E\left[u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \lambda}\right]\frac{\partial \lambda_{t|t-1}}{\partial \upsilon} + E\left[u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \upsilon}\right]\right)
\end{aligned}$$

Then taking unconditional expectations we have

$$E\left[\frac{\partial \vartheta_{t+1|t}}{\partial \kappa_{\vartheta 0}}\frac{\partial \lambda_{t+1|t}}{\partial \upsilon}\right] = \kappa_\lambda\left(h_\upsilon - \frac{\kappa_\lambda h_\lambda}{1-a}\mathcal{I}_{\lambda\upsilon}\right), \tag{E.18}$$

where $h_\lambda = E\left[u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \lambda}\right]$ and $h_\upsilon = E\left[u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \upsilon}\right]$. Then taking the conditional expectation with respect to $\mathcal{F}_{t-j-1}$ of the product of the partial derivative with respect to $\kappa_{\vartheta j}$, by the tower property of conditional expectation we have

$$
\begin{aligned}
E_{t-j-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\upsilon}\right) &= E_{t-j-1}\left(E_{t-1}\left(x_t\right)u_{t-j}^{\vartheta}\frac{\partial\lambda_{t|t-1}}{\partial\upsilon}+\kappa_\lambda E_{t-1}\left(\frac{\partial u_t^{\lambda}}{\partial\upsilon}\right)u_{t-j}^{\vartheta}\right)\\
&= E_{t-j-1}\left(au_{t-j}^{\vartheta}\left(x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\upsilon}+\kappa_\lambda\frac{\partial u_{t-1}^{\lambda}}{\partial\upsilon}\right)-\kappa_\lambda\mathcal{I}_{\vartheta\lambda}u_{t-j}^{\vartheta}\right)\\
&\;\vdots\\
&= a^jE_{t-j-1}\left(x_{t-j}u_{t-j}^{\vartheta}\right)\frac{\partial\lambda_{t-j|t-j-1}}{\partial\upsilon}+a^j\kappa_\lambda E_{t-j-1}\left(u_{t-j}^{\vartheta}\frac{\partial u_{t-j}^{\lambda}}{\partial\upsilon}\right)-\\
&\quad -\kappa_\lambda\mathcal{I}_{\vartheta\lambda}E_{t-j-1}\left(u_{t-j}^{\vartheta}\right)\sum_{i=0}^{j-1}a^i\\
&= a^j\kappa_\lambda E\left[u_{t-j}^{\vartheta}\frac{\partial u_{t-j}^{\lambda}}{\partial\lambda}\right]\frac{\partial\lambda_{t-j|t-j-1}}{\partial\upsilon}+a^j\kappa_\lambda E\left[u_{t-j}^{\vartheta}\frac{\partial u_{t-j}^{\lambda}}{\partial\upsilon}\right]
\end{aligned}
$$

Taking the unconditional expectations we then get

$$
E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\upsilon}\right]=a^j\kappa_\lambda\left(h_\upsilon-\frac{\kappa_\lambda g}{1-a}\mathcal{I}_{\lambda\upsilon}\right)
$$

Therefore

$$
\mathbf{\Psi}_{\boldsymbol{\kappa}_\vartheta\upsilon}=\kappa_\lambda\left(h_\upsilon-\frac{\kappa_\lambda g}{1-a}\mathcal{I}_{\lambda\upsilon}\right)\mathcal{I}_{\vartheta\lambda}\mathbf{a}^{\dagger}
$$

where $\mathbf{a}^{\dagger}$ is the $P\times 1$ vector defined as, $\mathbf{a}^{\dagger}=\left(1,\, a,\, a^2,...,a^{P-2},\, a^{P-1}\right)'$ Then, given that $\mathcal{I}_{\upsilon\lambda}$ is independent from λ we have that.

$$
\begin{aligned}
\mathbf{\Psi}_{\boldsymbol{\psi}_\lambda\boldsymbol{\kappa}_\vartheta'} &= E\left[\frac{\partial\ln f_{t+1}(y_{t+1}\mid Y_t;\boldsymbol{\theta})}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\ln f_{t+1}(y_{t+1}\mid Y_t;\boldsymbol{\theta})}{\partial\boldsymbol{\psi}_\lambda'}\right]\\
&= E\left[E_{t-1}\left(\frac{\partial\ln f_{t+1}}{\partial\vartheta_{t+1|t}}\frac{\partial\ln f_{t+1}}{\partial\lambda_{t+1|t}}\right)\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\boldsymbol{\psi}_\lambda'}\right]\\
&= \mathcal{I}_{\vartheta\lambda}E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\boldsymbol{\psi}_\lambda'}\right]
\end{aligned}
$$

where

$$
E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\boldsymbol{\psi}_\lambda'}\right]=E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\lambda},\,\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\phi_\lambda},\,\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\right],t=\ldots,0,1,....,T., \tag{E.19}
$$

Starting from the conditional expectations of the individual terms we have first,

$$
\begin{aligned}
E_{t-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\lambda}\right) &= E_{t-1}\left(u_t^{\vartheta}\left(x_t\frac{\partial\lambda_{t|t-1}}{\partial\omega_\lambda}+1-\phi_\lambda\right)\right)\\
&= E_{t-1}\left(x_tu_t^{\vartheta}\right)\frac{\partial\lambda_{t|t-1}}{\partial\omega_\lambda}+\left(1-\phi_\lambda\right)E_{t-1}\left(u_t^{\vartheta}\right)\\
&= \kappa_\lambda E\left[u_t^{\vartheta}\frac{\partial u_t^{\lambda}}{\partial\lambda}\right]\frac{\partial\lambda_{t|t-1}}{\partial\omega_\lambda}
\end{aligned}
$$

which after taking the unconditional expectation becomes

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_{\lambda}}\right]=\kappa_{\lambda}h_{\lambda}\frac{1-\phi_{\lambda}}{1-a}$$

Then taking the conditional expectation with respect to $\mathcal{F}_{t-j-1}$ of the product of the derivative with respect to $\kappa_{\vartheta j}$, by the tower property of conditional expectation we get

$$\begin{aligned}
E_{t-j-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_{\lambda}}\right) &= E_{t-j-1}\left(E_{t-1}\left(x_t\right)u^{\vartheta}_{t-j}\frac{\partial\lambda_{t|t-1}}{\partial\omega_{\lambda}}+\left(1-\phi_{\lambda}\right)u^{\vartheta}_{t-j}\right)\\
&= E_{t-j-1}\left(au^{\vartheta}_{t-j}\left(x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\omega_{\lambda}}+\left(1-\phi_{\lambda}\right)\right)+\left(1-\phi_{\lambda}\right)u^{\vartheta}_{t-j}\right)\\
&\vdots\\
&= a^{j}E_{t-j-1}\left(x_{t-j}u^{\vartheta}_{t-j}\right)\frac{\partial\lambda_{t-j|t-j-1}}{\partial\omega_{\lambda}}+E_{t-j-1}\left(u^{\vartheta}_{t-j}\right)\left(1-\phi_{\lambda}\right)\sum_{i=0}^{j}a^{i}\\
&= a^{j}\kappa_{\lambda}E\left[u^{\vartheta}_{t-j}\frac{\partial u^{\lambda}_{t-j}}{\partial\lambda}\right]\frac{\partial\lambda_{t-j|t-j-1}}{\partial\omega_{\lambda}}
\end{aligned}$$

Which after taking the unconditional expectation becomes

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_{\lambda}}\right]=a^{j}\kappa_{\lambda}h_{\lambda}\frac{1-\phi_{\lambda}}{1-a}$$

Therefore

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_{\vartheta}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_{\lambda}}\right]=\kappa_{\lambda}h_{\lambda}\frac{1-\phi_{\lambda}}{1-a}\mathbf{a}^{\dagger}$$

Then

$$\begin{aligned}
E_{t-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\phi_{\lambda}}\right) &= E_{t-1}\left(u^{\vartheta}_{t}\left(x_t\frac{\partial\lambda_{t|t-1}}{\partial\phi_{\lambda}}+\lambda_{t|t-1}-\omega_{\lambda}\right)\right)\\
&= E_{t-1}\left(x_tu^{\vartheta}_{t}\right)\frac{\partial\lambda_{t|t-1}}{\partial\phi_{\lambda}}+\left(\lambda_{t|t-1}-\omega_{\lambda}\right)E_{t-1}\left(u^{\vartheta}_{t}\right)\\
&= \kappa_{\lambda}E\left[u^{\vartheta}_{t}\frac{\partial u^{\lambda}_{t}}{\partial\lambda}\right]\frac{\partial\lambda_{t|t-1}}{\partial\phi_{\lambda}}
\end{aligned}$$

Which after taking unconditional expectation becomes 0. By the tower property of the conditional expectation, the same result applies also to all the products of the partial derivatives with respect to the other $\kappa_{\lambda j}$, therefore $E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_{\vartheta}}\frac{\partial\lambda_{t+1|t}}{\partial\phi_{\lambda}}\right]=\mathbf{0}$. Moreover

$$\begin{aligned}
E_{t-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\kappa_{\lambda}}\right) &= E_{t-1}\left(u^{\vartheta}_{t}\left(x_t\frac{\partial\lambda_{t|t-1}}{\partial\phi_{\lambda}}+u^{\lambda}_{t}\right)\right)\\
&= E_{t-1}\left(x_tu^{\vartheta}_{t}\right)\frac{\partial\lambda_{t|t-1}}{\partial\kappa_{\lambda}}+E_{t-1}\left(u^{\vartheta}_{t}u^{\lambda}_{t}\right)\\
&= \kappa_{\lambda}E\left[u^{\vartheta}_{t}\frac{\partial u^{\lambda}_{t}}{\partial\lambda}\right]\frac{\partial\lambda_{t|t-1}}{\partial\kappa_{\lambda}}+\mathcal{I}_{\vartheta\lambda}
\end{aligned}$$

which after taking unconditional expectations becomes

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\right]=\mathcal{I}_{\vartheta\lambda}$$

Then taking the conditional expectation with respect to $\mathcal{F}_{t-j-1}$ of the product of the derivative with respect to $\kappa_{\vartheta j}$, by the tower property of conditional expectation we get

$$\begin{aligned}
E_{t-j-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\right) &= E_{t-j-1}\left(E_{t-1}\left(x_t\right)u^\vartheta_{t-j}\frac{\partial\lambda_{t|t-1}}{\partial\kappa_\lambda}+E_{t-1}\left(u^\lambda_t\right)u^\vartheta_{t-j}\right)\\
&= E_{t-j-1}\left(au^\vartheta_{t-j}\left(x_{t-1}\frac{\partial\lambda_{t-1|t-2}}{\partial\kappa_\lambda}+u^\lambda_{t-1}\right)\right)\\
&\vdots\\
&= a^jE_{t-j-1}\left(x_{t-j}u^\vartheta_{t-j}\right)\frac{\partial\lambda_{t-j|t-j-1}}{\partial\kappa_\lambda}+a^jE_{t-j-1}\left(u^\lambda_{t-j}u^\vartheta_{t-j}\right)\\
&= a^j\kappa_\lambda E\left[u^\vartheta_{t-j}\frac{\partial u^\lambda_{t-j}}{\partial\lambda}\right]\frac{\partial\lambda_{t-j|t-j-1}}{\partial\kappa_\lambda}+a^j\mathcal{I}_{\lambda\vartheta}
\end{aligned}$$

which after taking unconditional expectations becomes

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\right]=a^j\mathcal{I}_{\vartheta\lambda}$$

Therefore

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_{\vartheta j}}\frac{\partial\lambda_{t+1|t}}{\partial\kappa_\lambda}\right]=\mathcal{I}_{\vartheta\lambda}\mathbf{a}^\dagger$$

The last component can be represented as

$$\begin{aligned}
\boldsymbol{\Psi}_{\boldsymbol{\kappa}_\vartheta\omega_\vartheta} &= E\left[\frac{\partial\ln f_{t+1}(y_{t+1}\mid Y_t;\boldsymbol{\theta})}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\ln f_{t+1}(y_{t+1}\mid Y_t;\boldsymbol{\theta})}{\partial\omega_\vartheta}\right]\\
&= \mathcal{I}_{\vartheta\lambda}E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]+\mathcal{I}_{\vartheta\upsilon}E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\vartheta_{t+1|t}}{\partial\omega_\vartheta}\right]\\
&= \mathcal{I}_{\vartheta\lambda}E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\boldsymbol{\kappa}_\vartheta}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]+\mathcal{I}_{\vartheta\vartheta}E\left[\mathbf{u}^\vartheta_t\right]
\end{aligned}$$

where $E\left[\mathbf{u}^\vartheta_t\right]=\mathbf{0}$. Then starting from taking the conditional expectation of the product of the partial derivatives of $\vartheta_{t|t-1}$ with respect to all the individual $\kappa_{\vartheta i}$, for $1=0,1,...,P-1$ we have first

$$\begin{aligned}
E_{t-1}\left(\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right) &= E_{t-1}\left(x_tu^\vartheta_t\right)\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}+\kappa_\lambda E_{t-1}\left(u^\vartheta_t\frac{\partial u^\lambda_t}{\partial\vartheta}\right)\\
&= \kappa_\lambda\left(E\left[u^\vartheta_t\frac{\partial u^\lambda_t}{\partial\lambda}\right]\frac{\partial\lambda_{t|t-1}}{\partial\omega_\vartheta}+E\left[u^\vartheta_t\frac{\partial u^\lambda_t}{\partial\vartheta}\right]\right)
\end{aligned}$$

Then taking unconditional expectations we have

$$E\left[\frac{\partial\vartheta_{t+1|t}}{\partial\kappa_{\vartheta 0}}\frac{\partial\lambda_{t+1|t}}{\partial\omega_\vartheta}\right]=\kappa_\lambda\left(h_\vartheta-\frac{\kappa_\lambda h_\lambda}{1-a}\mathcal{I}_{\lambda\vartheta}\right), \tag{E.20}$$

where $h_\vartheta = E\left[u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \vartheta}\right]$. Then taking the conditional expectation with respect to $\mathcal{F}_{t-j-1}$ of the product of the partial derivative with respect to $\kappa_{\vartheta j}$, by the tower property of conditional expectation we have

$$\begin{aligned}
E_{t-j-1}\left(\frac{\partial \vartheta_{t+1|t}}{\partial \kappa_{\vartheta j}} \frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right) &= E_{t-j-1}\left(E_{t-1}\left(x_t\right) u_{t-j}^\vartheta \frac{\partial \lambda_{t|t-1}}{\partial \omega_\vartheta} + \kappa_\lambda E_{t-1}\left(\frac{\partial u_t^\lambda}{\partial \vartheta}\right) u_{t-j}^\vartheta\right) \\
&= E_{t-j-1}\left(a u_{t-j}^\vartheta \left(x_{t-1} \frac{\partial \lambda_{t-1|t-2}}{\partial \omega_\vartheta} + \kappa_\lambda \frac{\partial u_{t-1}^\lambda}{\partial \vartheta}\right) - \kappa_\lambda \mathcal{I}_{\vartheta\lambda} u_{t-j}^\vartheta\right) \\
&\vdots \\
&= a^j E_{t-j-1}\left(x_{t-j} u_{t-j}^\vartheta\right) \frac{\partial \lambda_{t-j|t-j-1}}{\partial \omega_\vartheta} + a^j \kappa_\lambda E_{t-j-1}\left(u_{t-j}^\vartheta \frac{\partial u_{t-j}^\lambda}{\partial \vartheta}\right) - \\
&\quad -\kappa_\lambda \mathcal{I}_{\vartheta\lambda} E_{t-j-1}\left(u_{t-j}^\vartheta\right) \sum_{i=0}^{j-1} a^i \\
&= a^j \kappa_\lambda E\left[u_{t-j}^\vartheta \frac{\partial u_{t-j}^\lambda}{\partial \lambda}\right] \frac{\partial \lambda_{t-j|t-j-1}}{\partial \upsilon} + a^j \kappa_\lambda E\left[u_{t-j}^\vartheta \frac{\partial u_{t-j}^\lambda}{\partial \upsilon}\right]
\end{aligned}$$

Taking the unconditional expectations we then get

$$E\left[\frac{\partial \vartheta_{t+1|t}}{\partial \kappa_{\vartheta j}} \frac{\partial \lambda_{t+1|t}}{\partial \omega_\vartheta}\right] = a^j \kappa_\lambda \left(h_\vartheta - \frac{\kappa_\lambda h_\lambda}{1-a} \mathcal{I}_{\lambda\vartheta}\right)$$

Then

$$\mathbf{\Psi}_{\boldsymbol{\kappa}_\vartheta \upsilon} = \kappa_\lambda \left(h_\vartheta - \frac{\kappa_\lambda h_\lambda}{1-a} \mathcal{I}_{\lambda\vartheta}\right) \mathcal{I}_{\vartheta\lambda} \mathbf{a}^\dagger$$

Therefore

$$\mathbf{\Psi}_{\boldsymbol{\kappa\theta}'} = \mathcal{I}_{\lambda\vartheta}\left(\mathbf{a}^\dagger \cdot \left[\kappa_\lambda \left(h_\upsilon - \frac{\kappa_\lambda h_\lambda}{1-a} \mathcal{I}_{\lambda\upsilon}\right),\, \kappa_\lambda h_\lambda \frac{1-\phi_\lambda}{1-a},\, 0,\, \mathcal{I}_{\lambda\vartheta},\, \kappa_\lambda \left(h_\vartheta - \frac{\kappa_\lambda h_\lambda}{1-a} \mathcal{I}_{\lambda\vartheta}\right)\right]\right) = \mathcal{I}_{\lambda\vartheta} \mathbf{a}^\dagger \mathbf{g}'$$

From these results, once evaluated the conditional expectations, the full form of the test which can be expressed as

$$\begin{aligned}
LM_u(P) &= Q_u(P) + \frac{1}{T} \frac{\mathcal{I}_{\lambda\vartheta}^2}{\mathcal{I}_{\vartheta\vartheta}^2} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_{\vartheta\prime}} \mathbf{a}^\dagger \mathbf{g}' \left(\mathbf{\Psi}_{\boldsymbol{\theta\theta}} - \frac{\mathcal{I}_{\lambda\vartheta}^2}{\mathcal{I}_{\vartheta\vartheta}^2} \mathbf{g} \mathbf{a}^{\dagger\prime} \mathbf{a}^\dagger \mathbf{g}'\right)^{-1} \mathbf{g} \mathbf{a}^{\dagger\prime} \frac{\partial \ln L}{\partial \boldsymbol{\kappa}_\vartheta} \\
&= Q_u(P) + \frac{1}{T} \frac{\mathcal{I}_{\lambda\vartheta}^2}{\mathcal{I}_{\vartheta\vartheta}^2} \mathbf{g}' \left(\mathbf{\Psi}_{\boldsymbol{\theta\theta}} - \frac{\mathcal{I}_{\lambda\vartheta}^2}{\mathcal{I}_{\vartheta\vartheta}^2} \left(\sum_{i=0}^{P-1} a^{2i}\right) \mathbf{g}\mathbf{g}'\right)^{-1} \mathbf{g} \left(\sum_{j=0}^{P-1} u_t^\vartheta u_{t-1-j}^\vartheta a^j\right)^2 \\
&= Q_u(P) + T \mathcal{I}_{\lambda\vartheta}^2 \mathbf{g}' \left(\mathbf{\Psi}_{\boldsymbol{\theta\theta}} - \frac{\mathcal{I}_{\lambda\vartheta}^2}{\mathcal{I}_{\vartheta\vartheta}^2} \frac{1-a^{2P}}{1-a^2} \mathbf{g}\mathbf{g}'\right)^{-1} \mathbf{g} \left(\sum_{j=1}^{P} r_{\vartheta u}\left(j\right) a^{j-1}\right)^2
\end{aligned}$$

This methodology can be easily used in any DCS model to construct a test not only for testing the presence of a time varying tail but more in general for testing the presence of a second time varying parameter once a first on it has been already fitted.

Appendix F. Expectations of scores for the Generalised t distribution

The expression for $E\left(\frac{\partial u_t^\lambda}{\partial \lambda}\right)^2$ in b in Equation (E.3) can be obtained as

$$E\left(\frac{\partial u_t^\lambda}{\partial \lambda}\right)^2 = (1+\eta)^2 \upsilon^2 E\left[b_t^2 (1-b_t)^2\right] = \frac{\eta \upsilon^2 (1+\eta)(1+\upsilon)(\upsilon+\eta)}{(\eta+1+3\upsilon)(\eta+1+2\upsilon)(\eta+1+\upsilon)}$$

The expression for $E\left[u_t^\lambda \frac{\partial u_t^\lambda}{\partial \lambda}\right]$ in c in Equation (E.4) can be obtained as

$$\begin{aligned} E\left[u_t^\lambda \frac{\partial u_t^\lambda}{\partial \lambda}\right] &= (1+\eta)\,\upsilon\left(E\left[b_t(1-b_t)\right] - (1+\eta)\,E\left[b_t^2(1-b_t)\right]\right) \\ &= \frac{\upsilon\eta}{(\eta+1+\upsilon)} - \frac{\eta\upsilon^2(1+\eta)(1+\upsilon)}{(\eta+1+2\upsilon)(\eta+1+\upsilon)} \\ &= \frac{\eta\upsilon^2(1-\eta)}{(\eta+1+2\upsilon)(\eta+1+\upsilon)} \end{aligned}$$

The expression for h_λ in Equation (E.18) can be obtained as

$$\begin{aligned} E\left[u_t^\vartheta \frac{\partial u_t^\lambda}{\partial \lambda}\right] &= -\eta_s\left(\eta-\eta^\dagger\right)(1+\eta)\left(\tau E\left[b_t(1-b_t)\right] + E\left[\ln(1-b_t)\,b_t(1-b_t)\right] + \frac{(1+\eta)}{\eta}E\left[b_t^2(1-b_t)\right]\right) \\ &= -\eta_s\left(\eta-\eta^\dagger\right)\left\{\frac{(1+\upsilon)(1+\eta)}{(\eta+1+\upsilon)(\eta+1+2\upsilon)} - \frac{\eta}{(\eta+1+\upsilon)}\left[\psi\left(\frac{\eta+1}{\upsilon}\right) - \psi\left(\frac{\eta+1}{\upsilon}+2\right) - \right.\right. \\ &\quad \left.\left. - \left(\psi\left(\frac{\eta}{\upsilon}\right) - \psi\left(\frac{\eta}{\upsilon}+1\right)\right)\right]\right\} \\ &= -\eta_s\left(\eta-\eta^\dagger\right)\frac{(1+\upsilon)(2+3\upsilon+6\eta)+\eta^2\left[2-\upsilon(4+\upsilon+\eta)\right]}{(1+\eta+2\upsilon)(1+\eta+\upsilon)^2(1+\eta)} \end{aligned}$$

where $\tau = \psi\left(\frac{\eta+1}{\upsilon}\right) - \psi\left(\frac{\eta}{\upsilon}\right) - \frac{1}{\eta}$ Noticing that when ε_t is distributed with Generalised t distribution with shape parameters η and υ then $\ln|\varepsilon_t| = \left[\ln b_t - \ln(1-b_t) + \ln\eta\right]/\upsilon$, where b_t is distributed $beta\,(1/\upsilon,\ \eta/\upsilon)$.

The expression for h_υ in Equation (E.18) can be obtained as

$$E\left[u_t^{\vartheta}\frac{\partial u_t^{\lambda}}{\partial \upsilon}\right] = \eta_s\left(\eta-\eta^{\dagger}\right)\frac{(1+\eta)}{\upsilon}\left(\tau E\left[\ln\left(|\varepsilon_t|\right)b_t\left(1-b_t\right)\right]+E\left[\ln\left(|\varepsilon_t|\right)\ln\left(1-b_t\right)b_t\left(1-b_t\right)\right]+\right.$$

$$\left.+\frac{(1+\eta)}{\eta}E\left[\ln\left(|\varepsilon_t|\right)b_t^2\left(1-b_t\right)\right]\right)$$

$$=\frac{\eta_s\left(\eta-\eta^{\dagger}\right)}{\upsilon^2}\frac{\eta}{(\eta+1+\upsilon)}\left\{\tau\left[\ln\eta+\psi\left(\frac{1}{\upsilon}+1\right)-\psi\left(\frac{\eta}{\upsilon}+1\right)\right]+\right.$$

$$+\frac{(1+\upsilon)(1+\eta)}{\eta(\eta+1+\upsilon)}\left[\ln\eta+\psi\left(\frac{1}{\upsilon}+1\right)-\psi\left(\frac{\eta}{\upsilon}+1\right)+\frac{\upsilon}{1+\upsilon}\right]-$$

$$\left.-\left[\ln\eta+\psi\left(\frac{1}{\upsilon}+1\right)-\psi\left(\frac{\eta}{\upsilon}+1\right)-\ln\eta\right]\left[\tau+\frac{\upsilon}{(\eta+1+\upsilon)+\frac{\upsilon}{1+\eta}}+\frac{1-\upsilon}{\eta}\right]-\psi'\left(\frac{\eta}{\upsilon}+1\right)\right\}$$

$$=\eta_s\left(\eta-\eta^{\dagger}\right)\left[\frac{\eta\ln\eta}{\upsilon^2(\eta+1+\upsilon)}\tau+\right.$$

$$+\frac{(\eta+1+\upsilon)\left[2\eta(1+\upsilon)+\upsilon+\upsilon^2(1+\eta)-\eta^3\right]-\upsilon\left[\upsilon^2(1+\eta)-\upsilon\eta^2+1\right]}{(1+\eta)(\eta+1+\upsilon)^2(\eta+1+2\upsilon)\upsilon^2}\tau+$$

$$\left.+\frac{(1+\eta)}{(\eta+1+\upsilon)(\eta+1+2\upsilon)\upsilon}+\frac{1+(\eta+\upsilon)(\eta-\upsilon)(1+\eta)}{\upsilon^2(1+\eta)(\eta+1+\upsilon)^2}\ln\eta\right]$$

The expression for h_ϑ in Equation (E.20) can be obtained as

$$E\left[u_t^{\vartheta}\frac{\partial u_t^{\lambda}}{\partial \vartheta}\right] = \frac{\eta_s^2\left(\eta-\eta^{\dagger}\right)^2}{\upsilon}\left(\tau E\left[b_t\right]+E\left[b_t\ln\left(1-b_t\right)\right]+\frac{(1+\eta)}{\eta}E\left[b_t^2\right]-\frac{(1+\eta)}{\eta}E\left[b_t\left(1-b_t\right)\ln\left(1-b_t\right)\right]\right.$$

$$\left.-\tau\frac{1+\eta}{\eta}E\left[b_t\left(1-b_t\right)\right]-\frac{(1+\eta)^2}{\eta^2}E\left[b_t^2\left(1-b_t\right)\right]\right)$$

$$=\frac{\eta_s^2\left(\eta-\eta^{\dagger}\right)^2}{\upsilon}\left\{\frac{\eta}{(1+\eta)}\left[\psi\left(\frac{\eta}{\upsilon}\right)-\psi\left(\frac{\eta+1}{\upsilon}\right)+\frac{\upsilon}{\eta}\right]+\frac{(1+\upsilon)}{\eta(1+\eta+\upsilon)}-\right.$$

$$-\frac{(1+\eta)(1+\upsilon)}{\eta(1+\eta+\upsilon)(1+\eta+2\upsilon)}+\frac{1}{(1+\eta)}\left[\psi\left(\frac{\eta}{\upsilon}\right)-\psi\left(\frac{\eta+1}{\upsilon}\right)-\frac{\upsilon}{(1+\eta)}\right]-$$

$$-\frac{1}{(1+\eta+\upsilon)}\left[\psi\left(\frac{\eta}{\upsilon}\right)-\psi\left(\frac{\eta+1}{\upsilon}\right)+\frac{\upsilon}{\eta}-\frac{\upsilon}{(1+\eta+\upsilon)}-\frac{\upsilon}{(1+\eta)}\right]+$$

$$\left.+\psi\left(\frac{\eta}{\upsilon}\right)-\psi\left(\frac{\eta+1}{\upsilon}\right)+\frac{\upsilon}{\eta}\right\}$$

$$=\eta_s^2\left(\eta-\eta^{\dagger}\right)^2\left\{\left[\psi\left(\frac{\eta}{\upsilon}\right)-\psi\left(\frac{\eta+1}{\upsilon}\right)\right]\frac{(1+2\eta+2\upsilon)}{\upsilon(1+\eta+\upsilon)}+\right.$$

$$\left.+\frac{(1+\eta+\upsilon)(1+\eta+2\upsilon)\left[(1+2\upsilon)(1+\upsilon)-\eta\upsilon(2+2\eta+\upsilon)\right]+(1+\eta)^2\left[\eta(1+2\eta+3\upsilon)-(1+\upsilon)\right]}{\eta(1+\eta)^2(1+\eta+\upsilon)^2(1+\eta+2\upsilon)}\right\}$$

Then the expression for c_υ in Equation (E.12) can be obtained as

$$
\begin{aligned}
E\left[u_t^\lambda \frac{\partial u_t^\lambda}{\partial \upsilon}\right] &= \left(E\left[\ln|\varepsilon_t| b_t^2 (1-b_t)\right] - E\left[\ln|\varepsilon_t| b_t (1-b_t)\right]\right)(1+\eta)^2 \\
&= (1+\eta)\left(\frac{(1+\upsilon)\eta}{\upsilon(\eta+1+\upsilon)(\eta+1+2\upsilon)}\left[\ln\eta + \psi\left(\frac{1}{\upsilon}+2\right) - \psi\left(\frac{\eta}{\upsilon}+1\right)\right] - \right. \\
&\left. - \frac{\eta}{\upsilon(\eta+1+\upsilon)}\left[\ln\eta + \psi\left(\frac{1}{\upsilon}+2\right) - \psi\left(\frac{\eta}{\upsilon}+1\right)\right]\right) \\
&= \frac{(1+\eta)\eta}{\upsilon(\eta+1+\upsilon)(\eta+1+2\upsilon)}\left[(\eta+2\upsilon)\left(\psi\left(\frac{\eta}{\upsilon}+1\right) - \ln\eta\right) + \psi\left(\frac{1}{\upsilon}+2\right) - \right. \\
&\left. - (\eta+1+2\upsilon)\psi\left(\frac{1}{\upsilon}+1\right)\right]
\end{aligned}
$$

Then the expression for c_ϑ in Equation (E.16) can be obtained as

$$
\begin{aligned}
E\left[u_t^\lambda \frac{\partial u_t^\lambda}{\partial \vartheta}\right] &= \eta_s\left(\eta - \eta^\dagger\right)\left(E[b_t] + \frac{(1+\eta)}{\eta}E[b_t(1-b_t)] + (1+\eta)E\left[b_t^2\right] - \frac{(1+\eta)^2}{\eta}E\left[b_t^2(1-b_t)\right]\right) \\
&= \eta_s\left(\eta - \eta^\dagger\right)\left(\frac{\upsilon}{\eta(1+\eta+\upsilon)} - \frac{1}{(1+\eta)} + \frac{(1+\upsilon)}{(1+\eta+\upsilon)} - \frac{(1+\eta)(1+\upsilon)}{(1+\eta+\upsilon)(1+\eta+\upsilon)}\right) \\
&= \eta_s\left(\eta - \eta^\dagger\right)\frac{\upsilon^2\left(2+\eta+\eta^2\right) - \eta(1+\eta)^2 + \eta\upsilon}{\eta(1+\eta)(1+\eta+\upsilon)(1+\eta+2\upsilon)}
\end{aligned}
$$

The expression for $b_{\upsilon\upsilon}$ in Equation (E.13) can be obtained as

$$
\begin{aligned}
E\left(\frac{\partial u_t^\lambda}{\partial \upsilon}\right)^2 &= (1+\eta)E\left[\ln^2|\varepsilon_t| b_t^2(1-b_t)^2\right] \\
&= \frac{1+\eta}{\upsilon}\left(\ln^2\eta E\left[b_t^2(1-b_t)^2\right] + 2\ln\eta E\left[\ln(b_t) b_t^2(1-b_t)^2\right] - 2\ln\eta E\left[\ln(1-b_t) b_t^2(1-b_t)^2\right] + \right. \\
&\left. + E\left[\ln^2(b_t) b_t^2(1-b_t)^2\right] - 2E\left[\ln(b_t)\ln(1-b_t) b_t^2(1-b_t)^2\right] + E\left[\ln^2(1-b_t) b_t^2(1-b_t)^2\right]\right) \\
&= \frac{(1+\upsilon)\eta(\eta+\upsilon)}{\upsilon(\eta+1+\upsilon)(\eta+1+2\upsilon)(\eta+1+3\upsilon)}\left\{\left[\ln\eta + \psi\left(\frac{1}{\upsilon}+2\right) - \psi\left(\frac{\eta}{\upsilon}+2\right)\right]^2 + \right. \\
&\left. + \left[\psi'\left(\frac{1}{\upsilon}+2\right) + \psi'\left(\frac{\eta}{\upsilon}+2\right)\right]\right\}
\end{aligned}
$$

Then expression for $b_{\lambda\upsilon}$ in Equation (E.11) can be obtained as

$$
\begin{aligned}
E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \upsilon}\right] &= -E\left[\ln|\varepsilon_t| b_t^2(1-b_t)^2\right](1+\eta)^2 \\
&= -\frac{(1+\eta)^2}{\upsilon}\left(\ln\eta E\left[b_t^2(1-b_t)^2\right] + E\left[\ln(b_t) b_t^2(1-b_t)^2\right] - E\left[\ln(1-b_t) b_t^2(1-b_t)^2\right]\right) \\
&= -\frac{(1+\upsilon)\eta(\eta+\upsilon)(1+\eta)}{\upsilon(\eta+1+\upsilon)(\eta+1+2\upsilon)(\eta+1+3\upsilon)}\left[\ln\eta + \psi\left(\frac{1}{\upsilon}+2\right) - \psi\left(\frac{\eta}{\upsilon}+2\right)\right] \\
&= \frac{(1+\upsilon)\eta(\eta+\upsilon)(1+\eta)}{\upsilon(\eta+1+\upsilon)(\eta+1+2\upsilon)(\eta+1+3\upsilon)}\left\{\psi\left(\frac{\eta}{\upsilon}\right) - \psi\left(\frac{1}{\upsilon}\right) - \ln\eta - \right. \\
&\left. - \upsilon\frac{[(\upsilon+1)(\eta+\upsilon)+\eta](\eta-1)}{(\upsilon+1)(\eta+\upsilon)\eta}\right\}
\end{aligned}
$$

Then expression for $b_{\lambda\vartheta}$ in Equation (E.15) can be obtained as

$$E\left[\frac{\partial u_t^\lambda}{\partial \lambda}\frac{\partial u_t^\lambda}{\partial \vartheta}\right] = -\eta_s\left(\eta-\eta^\dagger\right)\upsilon\left(1+\eta\right)\left(E\left[b_t^2\left(1-b_t\right)\right]-\frac{\left(1+\eta\right)}{\eta}E\left[b_t^2\left(1-b_t\right)^2\right]\right)$$
$$= -\eta_s\left(\eta-\eta^\dagger\right)\frac{\upsilon\left(1+\upsilon\right)\left[\left(1+4\upsilon\right)\left(1+\eta\right)+6\upsilon^2\right]}{\left(1+\eta+\upsilon\right)\left(1+\eta+2\upsilon\right)\left(1+\eta+3\upsilon\right)}$$

Then expression for $b_{\upsilon\vartheta}$ in Equation (E.15) can be obtained as

$$E\left[\frac{\partial u_t^\lambda}{\partial \upsilon}\frac{\partial u_t^\lambda}{\partial \vartheta}\right] = -\eta_s\left(\eta-\eta^\dagger\right)\left(1+\eta\right)\left(E\left[\ln\left|\varepsilon_t\right|b_t^2\left(1-b_t\right)\right]-\frac{\left(1+\eta\right)}{\eta}E\left[\ln\left|\varepsilon_t\right|b_t^2\left(1-b_t\right)^2\right]\right)$$
$$= -\eta_s\left(\eta-\eta^\dagger\right)\frac{\left(1+\upsilon\right)}{\upsilon\left(1+\eta+\upsilon\right)\left(1+\eta+2\upsilon\right)}\left\{\frac{\left(2\eta-1\right)\upsilon}{\left(1+\eta+3\upsilon\right)}\left[\ln\eta+\psi\left(\frac{1}{\upsilon}\right)-\psi\left(\frac{\eta}{\upsilon}\right)\right]+\right.$$
$$\left.+\upsilon\left[\frac{\eta+\left(\eta-1\right)\left(\upsilon+1\right)}{\left(\upsilon+1\right)}-\frac{\left[\left(\upsilon+1\right)\left(\upsilon+\eta\right)+\eta\right]\left(\eta-1\right)}{\eta\left(1+\eta+3\upsilon\right)}\right]\right\}$$

Then expression for $b_{\vartheta\vartheta}$ in Equation (E.17) can be obtained as

$$E\left(\frac{\partial u_t^\lambda}{\partial \vartheta}\right)^2 = \eta_s^2\left(\eta-\eta^\dagger\right)^2 E\left[\left(b_t-\frac{\left(\eta+1\right)}{\eta}b_t\left(1-b_t\right)\right)^2\right]$$
$$= \eta_s^2\left(\eta-\eta^\dagger\right)^2\left(E\left[b_t^2\right]+\frac{\left(\eta+1\right)^2}{\eta^2}E\left[b_t^2\left(1-b_t\right)^2\right]-2\frac{\left(\eta+1\right)}{\eta}E\left[b_t^2\left(1-b_t\right)\right]\right)$$
$$= \eta_s^2\left(\eta-\eta^\dagger\right)^2\frac{\left(1+\upsilon\right)\left[\left(1+\eta+2\upsilon\right)\left(1+\eta+3\upsilon\right)-\left(1+\eta\right)\left(2+\eta+5\upsilon\right)\right]}{\left(1+\eta\right)\left(1+\eta+\upsilon\right)\left(1+\eta+2\upsilon\right)\left(1+\eta+3\upsilon\right)}$$

Appendix G. Figures

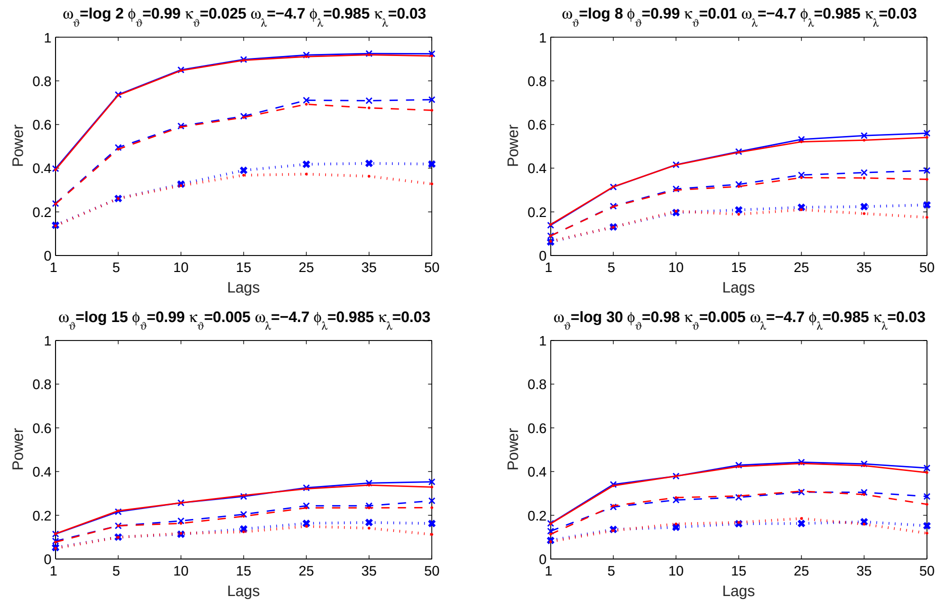

Figure G.14: Plot of the empirical Power of the LM test (Blue Line) and Q^* test (Red Line) for different lags, obtained from $N = 1000$ simulations of the Dynamic Scale-Tail model. The solid lines are for sample size $T = 2000$, the dashed lines for $T = 1000$ and the dotted lines for $T = 500$.

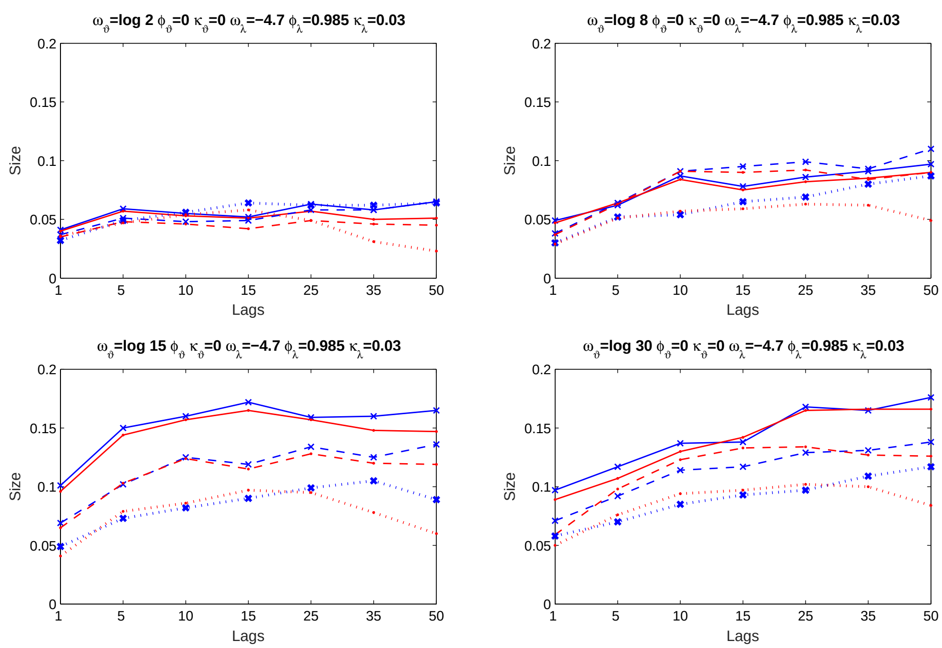

Figure G.15: Plot of the empirical Size of the LM test (Blue Line) and Q^* test (Red Line) for different lags, obtained from $N = 1000$ simulations of the Dynamic Scale-Tail model. The solid lines are for sample size $T = 2000$, the dashed lines for $T = 1000$ and the dotted lines for $T = 500$.

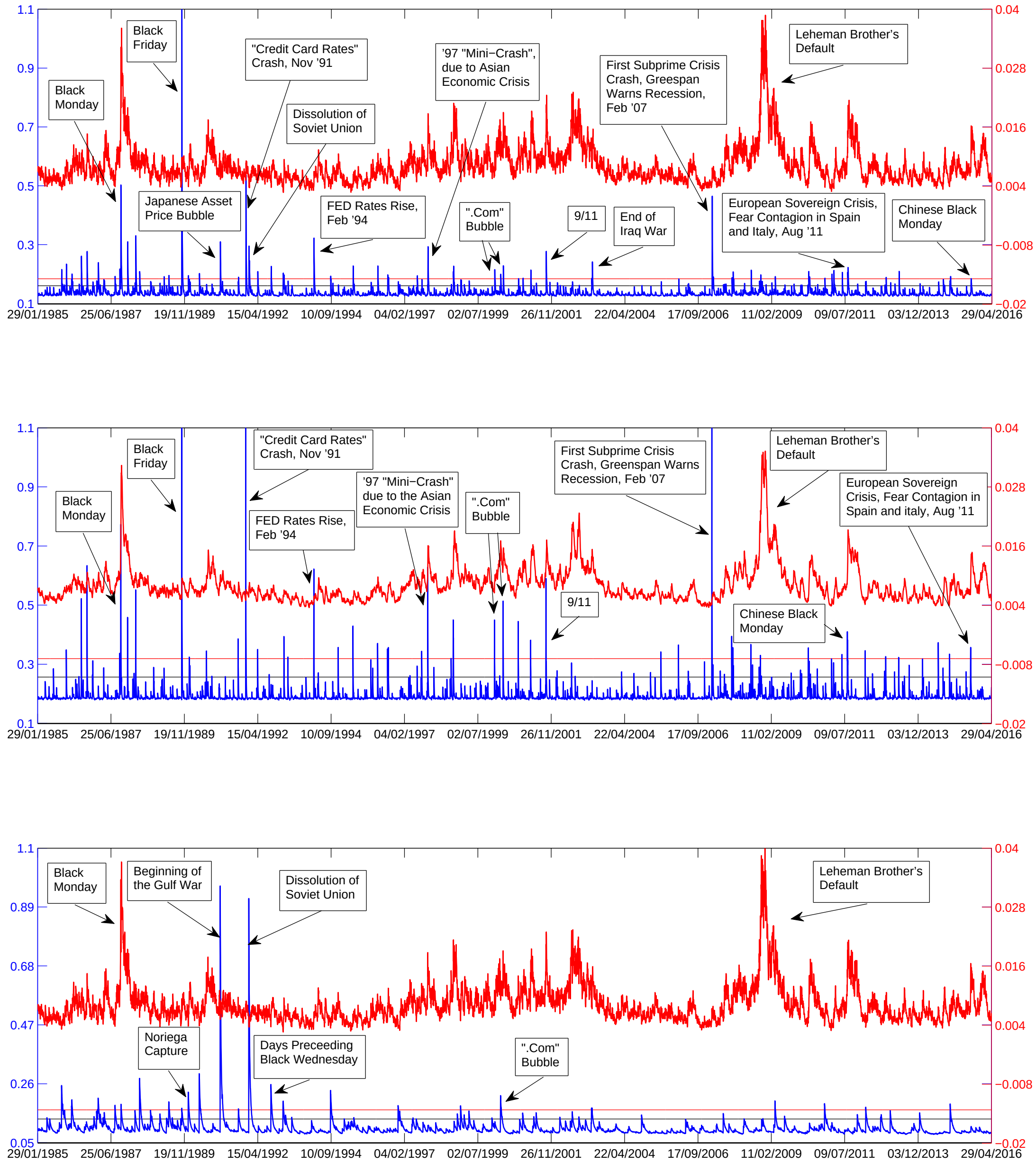

Figure G.16: Plot of the fitted inverse tail index parameters $\bar{\eta}_{t|t-1}$ and scale parameter $\varphi_{t|t-1}$ for the Dow Jones dataset in case of Asymmetric Lower Tail Dynamics without Leverage(Top), Symmetric Tail Dynamics with leverage (Mid) and Asymmetric Upper Tail Dynamics with Leverage (Bottom).

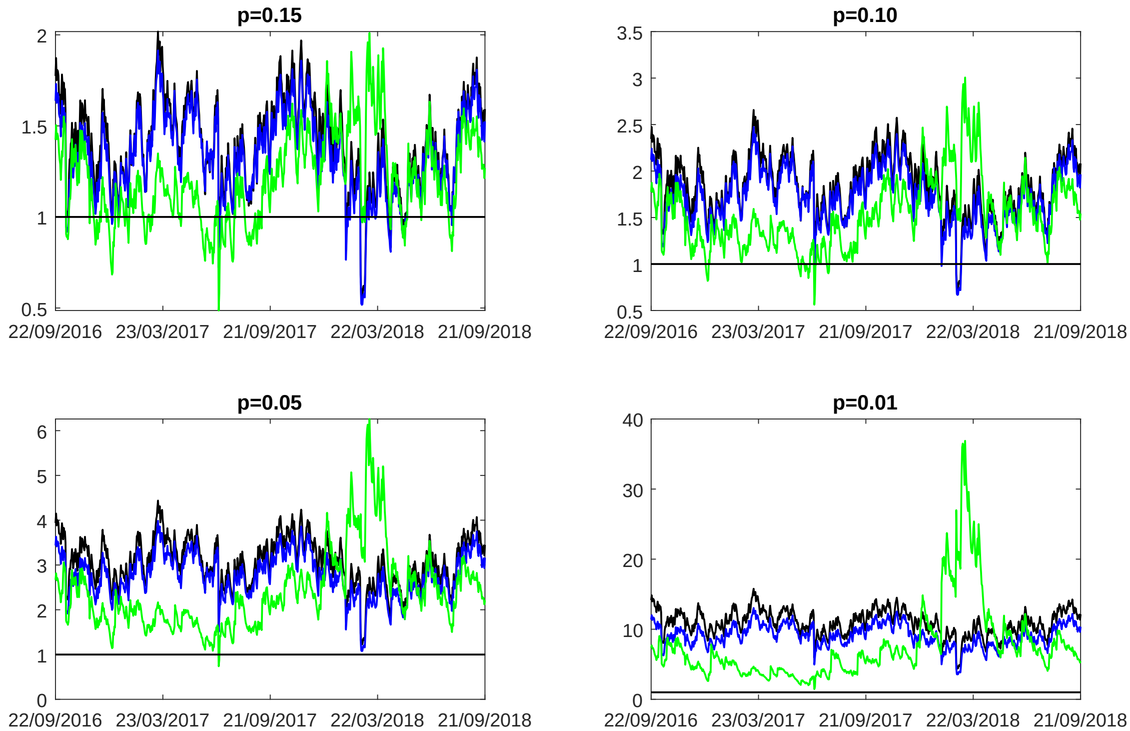

Figure G.17: Plot of the ratios of the Expected Shortfall above the lower 10% (Top Left), 5% (Top Right), 1% (Bottom Left) and 0.5% (Bottom Right) quantiles of the one-step-ahead forecasted conditional distribution of the 5y Italian CDS Rate Returns from fitting a symmetric DCS Beta-t-EGARCH model (Black Line), an asymmetric DCS Beta-t-EGARCH model (Blue Line), an symmetric dynamic tail DCS EGARCH model (Green Line) over the one-step-abead forecasted Expected Shortfall from a GARCH model.

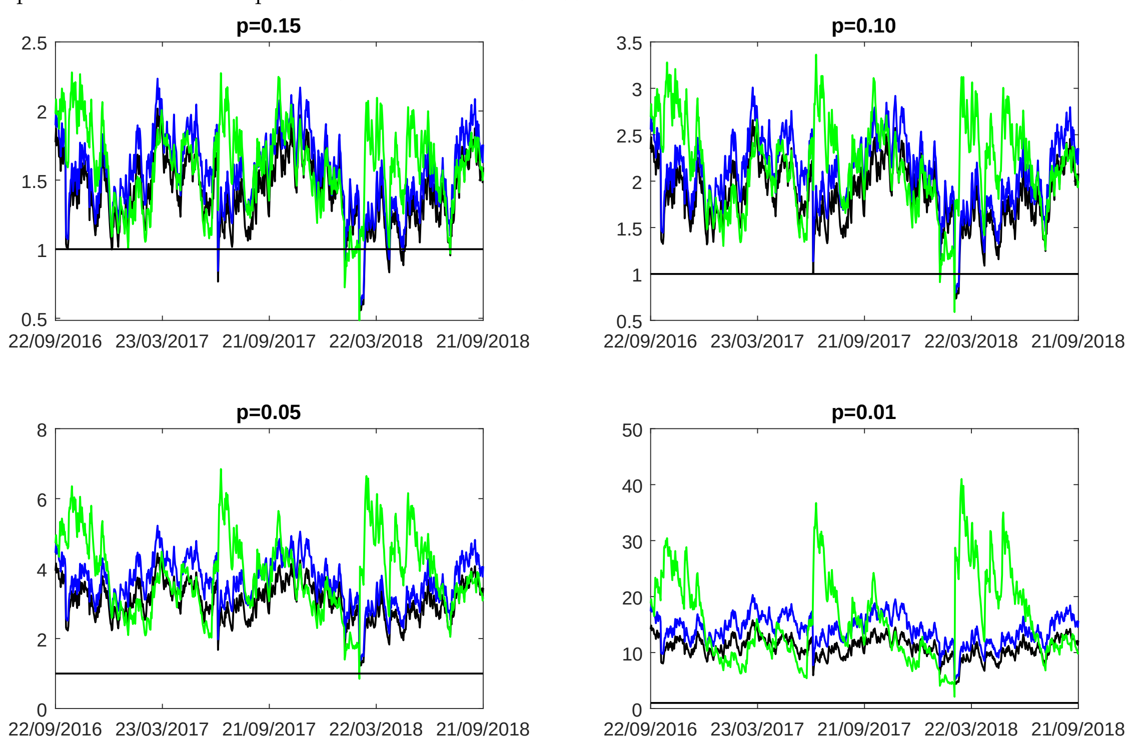

Figure G.18: Plot of the ratios of the Expected Shortfall below the upper 10% (Top Left), 5% (Top Right), 1% (Bottom Left) and 0.5% (Bottom Right) quantiles of the one-step-ahead forecasted conditional distribution of the 5y Italian CDS Rate Returns from fitting a symmetric DCS Beta-t-EGARCH model (Black Line), an asymmetric DCS Beta-t-EGARCH model (Blue Line), an symmetric dynamic tail DCS EGARCH model (Green Line) over the one-step-abead forecasted Expected Shortfall from a GARCH model.

Appendix H. Tables

$T=2000$	$\omega_\vartheta=\log 2,\ \phi_\vartheta=0.99,\ \kappa_\vartheta=0.25$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta=\log 8,\ \phi_\vartheta=0.99,\ \kappa_\vartheta=0.01$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta=\log 15,\ \phi_\vartheta=0.99,\ \kappa_\vartheta=0.005$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta=\log 30,\ \phi_\vartheta=0.98,\ \kappa_\vartheta=0.005$ Beta-t-EGARCH	Dynamic Scale and Tail		
	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% 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 $r_{u\vartheta}(P)$
$Q(1)$	4,633	1,017	0,760	0,021	2,901	3,182	0,972	0,033	7,684	16,476	2,588	0,133	12,895	9,559	2,875	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)	
$Q(5)$	20,204	4,879	3,978	0,031	11,947	7,664	4,054	0,048	21,428	31,076	10,623	0,201	32,203	18,083	12,047	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)	
$Q(10)$	38,057	9,924	8,122	0,027	22,697	12,907	7,896	0,061	29,593	42,149	18,570	0,241	43,636	23,515	22,063	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)	
$Q(15)$	53,361	15,204	12,331	0,029	32,464	18,058	11,934	0,062	36,102	53,139	25,901	0,259	50,814	28,798	31,550	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)	
$Q(25)$	78,734	25,757	20,858	0,028	48,790	28,550	20,291	0,07	48,775	75,217	40,019	0,269	63,117	39,127	48,588	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)	
$Q(35)$	100,182	36,224	29,937	0,025	62,381	39,088	29,502	0,083	60,259	96,736	52,598	0,264	74,225	49,320	65,097	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)	
$Q(50)$	128,365	51,868	43,896	0,033	81,934	55,127	42,569	0,081	77,820	129,303	70,945	0,274	89,406	64,611	87,234	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)	
$\hat{\eta}$	1,903	-			8,795	-			20,756	-			59,181	-		
	(0,286)				(15,198)				(33,762)				(88,537)			
$\exp(\hat{\omega}_\vartheta)$	-	2,048			-	8,828			-	19,858			-	37,361		
		(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 H.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 $T=2000$ generated by a conditional t distribution with dynamic scale and tail with different ω_ϑ assumptions.

$T=1000$	$\omega_\vartheta=\log 2,\ \phi_\vartheta=0.99,\ \kappa_\vartheta=0.25$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta=\log 8,\ \phi_\vartheta=0.99,\ \kappa_\vartheta=0.01$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta=\log 15,\ \phi_\vartheta=0.99,\ \kappa_\vartheta=0.005$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta=\log 30,\ \phi_\vartheta=0.98,\ \kappa_\vartheta=0.005$ Beta-t-EGARCH	Dynamic Scale and Tail		
	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% 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 $r_{u\vartheta}(P)$
$Q(1)$	2,799	1,051	0,755	0,021	2,205	0,983	1,748	0,076	2,516	3,463	2,192	0,15	5,742	4,547	1,329	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)	
$Q(5)$	14,551	5,051	3,828	0,033	9,412	5,150	6,868	0,112	8,541	8,232	7,955	0,222	18,834	12,075	5,378	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)	
$Q(10)$	27,307	10,259	8,099	0,038	18,034	10,445	12,705	0,14	14,743	13,141	13,693	0,241	27,362	20,471	9,725	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)	
$Q(15)$	38,637	15,447	12,313	0,036	25,247	15,582	18,159	0,148	20,468	18,316	19,861	0,264	33,299	28,426	13,964	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)	
$Q(25)$	58,617	25,626	21,154	0,037	39,688	25,834	28,506	0,162	31,982	28,596	30,165	0,266	43,907	44,002	22,605	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)	
$Q(35)$	75,253	36,019	30,291	0,039	52,428	36,327	38,080	0,172	43,278	38,884	39,999	0,272	54,043	59,672	31,011	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)	
$Q(50)$	97,089	51,714	43,777	0,05	69,943	51,655	52,148	0,159	59,242	54,354	54,245	0,254	68,553	83,185	43,558	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)	
$\hat{\eta}$	1,938	-			10,323	-			35,463	-			94,880	-		
	(0,495)				(20,535)				(70,487)				(126,017)			
$\exp(\hat{\omega}_\vartheta)$	-	2,120			-	10,421			-	24,091			-	43,993		
		(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 H.4: 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.

$T = 500$	$\omega_\vartheta = \log 2$, $\phi_\vartheta = 0.99$, $\kappa_\vartheta = 0.25$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta = \log 8$, $\phi_\vartheta = 0.99$, $\kappa_\vartheta = 0.01$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta = \log 15$, $\phi_\vartheta = 0.99$, $\kappa_\vartheta = 0.005$ Beta-t-EGARCH	Dynamic Scale and Tail			$\omega_\vartheta = \log 30$, $\phi_\vartheta = 0.98$, $\kappa_\vartheta = 0.005$ Beta-t-EGARCH	Dynamic Scale and Tail		
	$\hat{u}_\lambda$	$\hat{u}_\lambda$	$\hat{u}_\vartheta$	% 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 $r_{u\vartheta}(P)$
$Q(1)$	1,699 (0,192)	1,097 (0,295)	0,833 (0,361)	0,02	1,149 (0,284)	0,846 (0,358)	3,484 (0,062)	0,109	1,541 (0,214)	2,096 (0,148)	7,235 (0,007)	0,17	2,472 (0,116)	3,817 (0,051)	2,568 (0,109)	0,12
$Q(5)$	9,610 (0,889)	5,080 (0,954)	4,223 (0,975)	0,051	5,850 (0,950)	4,955 (0,974)	12,588 (0,626)	0,155	5,791 (0,908)	7,518 (0,836)	28,958 (0,204)	0,23	9,400 (0,781)	12,876 (0,576)	9,861 (0,766)	0,18
$Q(10)$	18,330 (0,050)	10,190 (0,424)	8,492 (0,581)	0,061	12,315 (0,265)	10,393 (0,407)	21,432 (0,018)	0,179	10,786 (0,374)	14,327 (0,159)	49,986 (0,000)	0,241	15,491 (0,115)	24,211 (0,007)	17,460 (0,065)	0,18
$Q(15)$	26,038 (0,038)	15,292 (0,431)	12,739 (0,622)	0,056	17,779 (0,274)	15,637 (0,407)	28,435 (0,019)	0,195	16,022 (0,381)	21,417 (0,124)	67,902 (0,000)	0,249	20,140 (0,167)	34,947 (0,003)	24,059 (0,064)	0,173
$Q(25)$	40,342 (0,027)	25,508 (0,434)	21,927 (0,640)	0,067	28,976 (0,265)	26,176 (0,398)	40,786 (0,024)	0,193	26,207 (0,397)	34,763 (0,093)	95,830 (0,000)	0,241	30,107 (0,220)	55,668 (0,000)	37,284 (0,054)	0,185
$Q(35)$	53,253 (0,025)	35,617 (0,439)	30,991 (0,662)	0,064	39,467 (0,277)	36,782 (0,386)	51,242 (0,038)	0,178	36,229 (0,411)	47,508 (0,077)	117,822 (0,000)	0,237	39,457 (0,277)	75,138 (0,000)	48,140 (0,069)	0,168
$Q(50)$	70,635 (0,029)	51,067 (0,432)	44,514 (0,692)	0,066	54,049 (0,322)	52,250 (0,387)	65,948 (0,065)	0,169	50,419 (0,457)	65,881 (0,065)	144,210 (0,000)	0,225	53,312 (0,348)	101,293 (0,000)	63,758 (0,091)	0,164
$\hat{\eta}$	2,077 (0,746)	-			25,047 (64,449)	-			63,319 (106,959)	-			124,676 (143,219)	-		
$\exp(\hat{\omega}_\vartheta)$	-	2,275 (0,464)			-	14,335 (1,027)			-	31,182 (1,077)			-	49,877 (0,944)		
$\hat{\phi}_\vartheta$	-	0,947 (0,161)			-	0,928 (0,224)			-	0,916 (0,258)			-	0,907 (0,211)		
$\hat{\kappa}_\vartheta$	-	0,018 (0,031)			-	0,000 (0,022)			-	-0,002 (0,016)			-	0,000 (0,016)		

Table H.5: 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.

	Mean	Static Tail Index			Dynamic Tail Index									Dynamic Scale							Fit		
	μ	η_1	η	η_2	$\omega_{\vartheta 1}$	$\phi_{\vartheta 1}$	$\kappa_{\vartheta 1}$	ω_ϑ	ϕ_ϑ	κ_ϑ	$\omega_{\vartheta 2}$	$\phi_{\vartheta 2}$	$\kappa_{\vartheta 2}$	ω_λ	$\phi_{1,\lambda}$	$\kappa_{1,\lambda}$	$\kappa^*_{1,\lambda}$	$\phi_{2,\lambda}$	$\kappa_{2,\lambda}$	$\kappa^*_{2,\lambda}$	Logl	AIC	BIC
	0,001		6,588					-	-	-				-4,884	0,996	0,027	-	0,958	0,041	-	25.974,76	- 51.935,51	- 51.886,71
	(0,000)		(0,072)					-	-	-				(0,081)	(0,002)	(0,009)	-	(0,015)	(0,009)	-			
	0,001		-					1,829	0,792	0,020				-4,900	0,987	0,062	-	0,257	-0,058	-	25.993,16	- 51.968,32	- 51.905,57
	(0,000)		-					(0,071)	(0,087)	(0,005)				(0,056)	(0,003)	(0,006)	-	(0,158)	(0,013)	-			
	0,001	5,460		9,625	-	-	-				-	-	-	-4,869	0,996	0,028	-	0,958	0,040	-	25.987,40	- 51.958,79	- 51.903,02
	(0,000)	(0,076)		(0,129)	-	-	-				-	-	-	(0,082)	(0,002)	(0,009)	-	(0,015)	(0,009)	-			
	0,001	-		8,718	1,651	0,643	0,022				-	-	-	-4,884	0,986	0,068	-	0,182	-0,048	-	25.996,86	- 51.973,71	- 51.904,00
	(0,000)	-		(0,114)	(0,073)	(0,206)	(0,008)				-	-	-	(0,056)	(0,003)	(0,006)	-	(0,185)	(0,013)	-			
Dow Jones																							
	0,000		6,900					-	-	-				-4,919	0,984	0,061	0,024	0,656	-0,049	0,057	26.094,38	- 52.170,76	- 52.108,02
	(0,000)		(0,072)					-	-	-				(0,050)	(0,003)	(0,007)	(0,004)	(0,073)	(0,014)	(0,007)			
	0,000		-					2,000	0,789	0,012				-4,919	0,984	0,061	0,024	0,656	-0,049	0,057	26.103,70	- 52.185,41	- 52.108,72
	(0,000)		-					(0,076)	(0,166)	(0,004)				(0,050)	(0,003)	(0,007)	(0,004)	(0,072)	(0,014)	(0,007)			
	0,001	5,741		10,379	-	-	-				-	-	-	-4,889	0,985	0,062	0,023	0,690	-0,041	0,053	26.106,53	- 52.193,06	- 52.123,34
	(0,000)	(0,138)		(0,076)	-	-	-				-	-	-	(0,053)	(0,003)	(0,007)	(0,005)	(0,081)	(0,016)	(0,006)			
	0,001	5,644		-	-	-	-				2,304	0,932	0,007	-4,893	0,986	0,058	0,024	0,718	-0,041	0,058	26.115,57	- 52.207,13	- 52.123,47
	(0,000)	(0,074)		-	-	-	-				(0,136)	(0,066)	(0,003)	(0,053)	(0,004)	(0,008)	(0,006)	(0,097)	(0,020)	(0,007)			

Table H.6: Parameter Estimates for the Beta-t-EGARCH Model and dynamic Scale-Tail model without and with leverage term.

	Res Corr Scale		Tail $Q_u^*(P)$ Test		
	Sym	Asym	η_1	η	η_2
$Q(1)$	**8,548**	**8,500**	**4,950**	**14,302**	0,727
	(0,003)	**(0,004)**	**(0,026)**	**(0,000)**	(0,394)
$Q(5)$	**11,240**	**11,366**	**33,731**	**47,049**	3,492
	(0,047)	**(0,045)**	**(0,000)**	**(0,000)**	(0,625)
$Q(10)$	**20,647**	**20,418**	**43,360**	**58,038**	7,150
	(0,024)	**(0,026)**	**(0,000)**	**(0,000)**	(0,711)
$Q(15)$	24,014	24,240	**45,655**	**61,841**	11,375
	(0,065)	(0,061)	**(0,000)**	**(0,000)**	(0,726)
$Q(25)$	35,783	35,313	**51,467**	**70,176**	14,336
	(0,075)	(0,083)	**(0,001)**	**(0,000)**	(0,956)
$Q(35)$	**53,120**	**51,983**	**61,708**	**82,916**	20,714
	(0,025)	**(0,032)**	**(0,004)**	**(0,000)**	(0,974)
$Q(50)$	66,658	64,915	**70,645**	**94,310**	31,717
	(0,058)	(0,076)	**(0,029)**	**(0,000)**	(0,980)

Table H.7: Box-Ljung test on fitted scores with respect to scale $\hat{u}_t^\lambda$ and Simple LM Dynamic Tail test after fitting the Beta-t-EGARCH model without leverage. Symmetric and Asymmetric case.

	Res Corr Scale		Res Corr Tail		
	Sym	Asym	η_1	η	η_2
$Q(1)$	0,328	0,051	0,880	0,772	-
	(0,567)	(0,822)	(0,348)	(0,380)	-
$Q(5)$	1,542	1,329	5,104	3,270	-
	(0,908)	(0,932)	(0,403)	(0,658)	-
$Q(10)$	7,732	6,090	7,744	9,250	-
	(0,655)	(0,808)	(0,654)	(0,509)	-
$Q(15)$	11,935	11,594	10,043	12,869	-
	(0,684)	(0,709)	(0,817)	(0,612)	-
$Q(25)$	27,855	27,761	17,909	26,831	-
	(0,315)	(0,319)	(0,846)	(0,364)	-
$Q(35)$	42,576	42,450	27,576	41,042	-
	(0,177)	(0,181)	(0,810)	(0,223)	-
$Q(50)$	55,279	54,382	37,417	53,944	-
	(0,282)	(0,311)	(0,906)	(0,326)	-

Table H.8: Box-Ljung test on fitted scores with respect to scale $\hat{u}_t^\lambda$ and with respect to the dynamic tail index parameter $\hat{u}_t^\vartheta$ after fitting the dynamic Scale-Tail DCS Model without leverage. Symmetric and Asymmetric case.

	Res Corr Scale		Tail $Q_u^*(P)$ Test		
	Sym	Asym	η_1	η	η_2
$Q(1)$	2,010	2,630	0,161	**7,415**	**6,853**
	(0,156)	(0,105)	(0,688)	**(0,006)**	**(0,009)**
$Q(5)$	4,261	5,007	7,494	**17,652**	**39,557**
	(0,512)	(0,415)	(0,186)	**(0,003)**	**(0,000)**
$Q(10)$	8,648	10,279	15,332	**24,899**	**44,202**
	(0,566)	(0,416)	(0,120)	**(0,006)**	**(0,000)**
$Q(15)$	13,580	15,123	17,811	**30,765**	**47,115**
	(0,558)	(0,443)	(0,273)	**(0,009)**	**(0,000)**
$Q(25)$	28,016	29,337	24,349	**40,167**	**53,347**
	(0,307)	(0,250)	(0,499)	**(0,028)**	**(0,001)**
$Q(35)$	44,776	45,720	33,544	**52,373**	**58,706**
	(0,125)	(0,106)	(0,538)	**(0,030)**	**(0,007)**
$Q(50)$	60,699	60,564	45,243	**68,478**	65,242
	(0,143)	(0,146)	(0,664)	**(0,042)**	(0,073)

Table H.9: Box-Ljung test on fitted scores with respect to scale $\hat{u}_t^\lambda$ and Simple LM Dynamic Tail test after fitting the Beta-t-EGARCH model with leverage. Symmetric and Asymmetric case.

	Res Corr Scale		Res Corr Tail		
	Sym	Asym	η_1	η	η_2
$Q(1)$	**5,759**	**4,091**	-	0,010	0,462
	(0,016)	**(0,043)**	-	(0,920)	(0,497)
$Q(5)$	6,354	6,850	-	0,545	0,705
	(0,273)	(0,232)	-	(0,990)	(0,983)
$Q(10)$	10,741	12,297	-	6,685	1,836
	(0,378)	(0,266)	-	(0,755)	(0,997)
$Q(15)$	15,373	17,416	-	11,352	2,462
	(0,425)	(0,295)	-	(0,727)	(1,000)
$Q(25)$	28,406	31,453	-	24,444	3,295
	(0,290)	(0,174)	-	(0,494)	(1,000)
$Q(35)$	44,028	46,797	-	34,282	8,213
	(0,141)	(0,088)	-	(0,503)	(1,000)
$Q(50)$	59,805	61,075	-	45,775	12,197
	(0,161)	(0,136)	-	(0,644)	(1,000)

Table H.10: Box-Ljung test on fitted scores with respect to scale $\hat{u}_t^\lambda$ and with respect to the dynamic tail index parameter $\hat{u}_t^\vartheta$ after fitting the dynamic Scale-Tail DCS Model with leverage. Symmetric and Asymmetric case.

	Mean	Shape			Dynamic Tail Index										Dynamic Scale			Fit		
	μ	η_1	η	η_2	$\omega_{\vartheta 1}$	$\phi_{\vartheta 1}$	$\kappa_{\vartheta 1}$	ω_{ϑ}	ϕ_{ϑ}	$\kappa_{1\vartheta}$	$\kappa_{2\vartheta}$	$\omega_{\vartheta 2}$	$\phi_{\vartheta 2}$	$\kappa_{\vartheta 2}$	ω_{λ}	ϕ_{λ}	κ_{λ}	Logl	AIC	BIC
5Y CDS Italy	-0,001		2,491					-	-	-	-				-3,862	0,957	0,105	6.102,32	- 12.194,64	- 12.164,58
	(0,000)		(0,063)					-	-	-	-				(0,067)	(0,011)	(0,013)			
	0,000		-					0,693	1,000	-0,034	0,056				-3,872	0,974	0,116	6.155,18	- 12.294,36	- 12.246,27
	(0,000)		-					(0,990)	(0,001)	(0,014)	(0,013)				(0,083)	(0,007)	(0,019)			
	-0,001	2,687		2,330	-	-	-					-	-	-	-3,856	0,957	0,104	6.103,80	- 12.195,60	- 12.159,54
	(0,000)	(0,078)		(0,072)	-	-	-					-	-	-	(0,067)	(0,011)	(0,013)			
	-0,001	-		-	0,933	0,997	0,014					0,693	0,998	0,014	-3,853	0,969	0,100	6.129,02	- 12.238,05	- 12.177,94
	(0,000)	-		-	(0,235)	(0,002)	(0,004)					(0,415)	(0,002)	(0,003)	(0,090)	(0,009)	(0,013)			

Table H.11: Parameter Estimates for the Beta-t-EGARCH Model and dynamic Scale-Tail model

	Res Corr Scale		Tail $Q_u^*(P)$ Test		
	Sym	Asym	η_1	η	η_2
$Q(1)$	**7,914**	**8,204**	**6,692**	**38,982**	**7,613**
	(0,005)	**(0,004)**	**(0,010)**	**(0,000)**	**(0,006)**
$Q(5)$	**15,141**	**15,234**	**11,898**	**53,813**	**20,050**
	(0,010)	**(0,009)**	**(0,036)**	**(0,000)**	**(0,001)**
$Q(10)$	16,703	16,780	16,693	**57,667**	**27,637**
	(0,081)	(0,079)	(0,081)	**(0,000)**	**(0,002)**
$Q(15)$	19,310	19,196	**25,445**	**63,566**	**27,723**
	(0,200)	(0,205)	**(0,044)**	**(0,000)**	**(0,023)**
$Q(25)$	24,325	24,389	**62,917**	**83,861**	**47,803**
	(0,501)	(0,497)	**(0,000)**	**(0,000)**	**(0,004)**
$Q(35)$	41,287	40,884	**66,260**	**101,423**	**63,028**
	(0,215)	(0,228)	**(0,001)**	**(0,000)**	**(0,003)**
$Q(50)$	60,458	60,138	**74,762**	**111,105**	**72,515**
	(0,148)	(0,154)	**(0,013)**	**(0,000)**	**(0,020)**

Table H.12: Box-Ljung test on fitted scores with respect to scale $\hat{u}_t^{\lambda}$ and Simple LM Dynamic Tail test after fitting the Beta-t-EGARCH model. Symmetric and Asymmetric case.

	Res Corr Scale		Res Corr Tail		
	Sym	Asym	η_1	η	η_2
$Q(1)$	0,671	**5,731**	1,808	0,180	2,656
	(0,413)	**(0,017)**	(0,179)	(0,671)	(0,103)
$Q(5)$	5,531	**13,130**	3,653	1,589	4,317
	(0,355)	**(0,022)**	(0,600)	(0,903)	(0,505)
$Q(10)$	7,421	14,335	5,535	6,390	8,827
	(0,685)	(0,158)	(0,853)	(0,782)	(0,549)
$Q(15)$	10,433	18,315	10,005	11,071	11,896
	(0,792)	(0,246)	(0,819)	(0,748)	(0,687)
$Q(25)$	18,073	24,792	22,558	19,460	16,635
	(0,839)	(0,474)	(0,603)	(0,775)	(0,895)
$Q(35)$	30,594	39,927	25,979	27,431	24,432
	(0,681)	(0,260)	(0,866)	(0,815)	(0,909)
$Q(50)$	45,491	58,419	34,687	39,593	33,604
	(0,655)	(0,194)	(0,951)	(0,854)	(0,964)

Table H.13: Box-Ljung test on fitted scores with respect to scale $\hat{u}_t^{\lambda}$ and with respect to the dynamic tail index parameter $\hat{u}_t^{\vartheta}$ after fitting the dynamic Scale-Tail DCS Model. Symmetric and Asymmetric case.

			GARCH				DCS Sym Tails				DCS Asym Tails			
		p	0.15	0.1	0.05	0.01	0.15	0.1	0.05	0.01	0.15	0.1	0.05	0.01
Fixed Tail	Independence	L&U	6.031	8.550*	6.913	3.314	2.910	7.844	2.747	7.139	2.137	8.988	4.109	7.980*
		L	0.015	0.542	0.763	0.250	1.888	0.247	1.701	0.016	0.180	0.573	1.701	0.016
		U	2.320	2.262	3.331**	0.306	1.060	3.018*	0.162	0.116	1.361	4.191**	0.642	0.055
	Coverage	L&U	77.468***	38.194***	19.410***	1.290	1.467	2.621	5.234*	5.728*	5.180*	6.207**	7.137**	7.311**
		L	41.078***	25.146***	15.209***	0.372	0.067	1.589	5.100**	5.460**	0.068	1.931	5.100**	5.460**
		U	25.516***	9.800***	3.519*	0.904	1.464	0.768	0.066	0.249	4.773**	3.658*	1.740	1.802
	Backtesting	L	−5.185***	−4.126***	−1.910*	1.380	−2.271**	−2.505**	−2.378**	-1.262	−2.160**	−2.357**	−2.194**	-1.212
		U	−1.945*	0.988	9.426***	109.527***	-1.037	-0.841	-1.082	-0.688	-1.072	-0.772	-0.752	0.996
	Pred Lik		1,573.58				1,709.59				1,709.12			
Dynamic Tail	Independence	L&U					4.353	1.524	3.215	0.278	3.860	6.912	3.194	4.993
		L					0.839	0.230	0.104	0.016	1.864	0.003	2.485	0.200
		U					0.620	0.276	0.131	0.200	1.804	4.004	0.264	0.055
	Coverage	L&U					0.708	0.420	1.763	5.516*	2.772	2.621	2.189	1.862
		L					0.593	0.060	1.740	5.460**	0.321	0.768	1.740	0.066
		U					0.220	0.389	0.007	0.066	2.690	1.589	0.365	1.802
	Backtesting	L					-0.136	-1.082	-1.489	-1.336	-0.436	-0.871	-0.650	-0.092
		U					-0.398	-0.338	0.512	-0.058	-0.887	-0.378	-0.165	0.868
	Pred Lik						1,686.39				1,709.24			

Table H.14: Results of the unconditional coverage and independence likelihood ratio tests of Christoffersen (1998) for the upper tail, lower tails one-step-ahead quantiles and joint interval violations, 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.