Discussion Paper 4/2022
Guatemalan Econometric Study Group
Universidad Francisco Marroquín
March 31, 2022

Optimal Choice of the Scaling Parameters in Score-Driven Filters

Szabolcs Blazsek, Astrid Loretta Ayala & Adrian Licht
School of Business, Universidad Francisco Marroquín, Guatemala City, Guatemala

ABSTRACT
In the literature on score-driven models, alternative choices of the scaling parameters of the conditional score terms are used, but the optimal choice of those parameters is an open question. For the score-driven models with time-invariant scaling parameters, the choice of those parameters is irrelevant because the filters are identical for all scaling parameters. However, there are relevant score-driven models with time-varying scaling parameters, for which score-driven scale filters appear in the information matrix and the choice of the scaling parameters is relevant. We study this question for the quasi-autoregressive (QAR) plus Beta-t-EGARCH (exponential generalized autoregressive conditional heteroskedasticity) score-driven model. That model includes two score-driven filters (i.e., the location and scale filters), for which each updating term is the product of a scaling parameter and a conditional score. For the QAR plus Beta-t-EGARCH model, we use all alternative scaling parameters of the literature. We show for different scaling parameters in the location filter that both the location and scale filters significantly differ and a ranking of the statistical performances of the alternative specifications can be created. We show that the best-performing scaling parameter for the score-driven location filter is the conditional inverse information matrix.

KEYWORDS Dynamic conditional score (DCS); Generalized autoregressive score (GAS); Scaling parameters of the conditional score function; Quasi-autoregressive (QAR) model; Beta-t-EGARCH (exponential generalized autoregressive conditional heteroskedasticity) model

CONTACT E-mail: Szabolcs Blazsek, sblazsek@ufm.edu; Postal address: School of Business, Universidad Francisco Marroquín, Calle Manuel F. Ayau, Guatemala City, 01010, Guatemala

1. Introduction

In the literature on score-driven models, started by the works of Creal, Koopman, and Lucas (2008) and Harvey and Chakravarty (2008), the optimal choice of the scaling parameter of the updating term in the score-driven filter is an open question. Different choices of the scaling parameter provide different score-driven models (Creal, Koopman, and Lucas 2013). There are no results on the optimal choice of the scaling parameters. For the potential users of these models this ambiguity is problematic, because they would have to choose a certain scaling parameter. This paper is motivated by this issue.

In the work of Harvey (2013), in which the score-driven models are named dynamic conditional score (DCS) models, the use of the following scaling parameters are suggested: (i) For the score-driven location models, the inverse of the unconditional variance of the score $I^{-1} \equiv [\text{Var}(\nabla_{\mu,t})]^{-1}$ is suggested, where $\nabla_{\mu,t} = \partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)/\partial\mu_t$, μ_t is the score-driven location filter, $\ln f(y_t|\mathcal{F}_{t-1}, \Theta)$ is the log conditional density of the dependent variable y_t, $\mathcal{F}_{t-1} = \sigma(y_1, \ldots, y_{t-1}, \mu_1)$ is the σ-algebra, and Θ is the vector of constant parameters. This choice of scaling parameter corresponds to the inverse information matrix of maximum likelihood (ML) estimator. (ii) For the score-driven scale models, the use of the unit scaling parameter is suggested. For the DCS models (i) and (ii) are time-invariant.

In the work of Creal, Koopman, and Lucas (2013), in which the score-driven models are named generalized autoregressive score (GAS) models, the use of the following scaling parameters are suggested for all score-driven models: (i) conditional variance of the score $I_t^{-1} \equiv [\text{Var}(\nabla_{\mu,t}|\mathcal{F}_{t-1}, \Theta)]^{-1}$ scaling parameter (I_t^{-1} refers to the conditional inverse information matrix); (ii) J_t scaling parameter which is defined by the decomposition $I_t^{-1} = J_t' J_t$; (iii) unit scaling parameter. For (i) and (ii) the scaling parameters are time-varying, which provides an additional flexibility for the GAS models compared to the DCS models. We note that this additional flexibility can only be exploited for those score-driven models in which at least one of the score-driven filters appear in I_t^{-1}. If there are no score-driven filters in the information matrix, then the scaling of DCS and GAS models will coincide: $I_t^{-1} = I^{-1}$.

In the empirical analysis of the present paper, we use data on daily Vanguard Standard & Poor's (S&P 500) exchange traded fund (ETF) return y_t for the period of September 10, 2010 to February 9, 2022, for which we use what are perhaps the most popular score-driven location and score-driven scale models. According to the time-invariant scaling parameter approach of Harvey (2013) and the time-varying scaling parameter approach of Creal, Koopman, and Lucas (2013), we classify the score-driven models into two groups, to compare the performances of alternative scaling parameters:

First, we start with the following two score-driven models for which the information matrix does not depend on score-driven filters: (i) dynamic Student's t location model (Harvey 2013), i.e., quasi-autoregressive (QAR) model; (ii) Beta-t-EGARCH (exponential generalized autoregressive conditional heteroskedasticity) model of conditional volatility (Harvey and Chakravarty 2008). The QAR model provides a robust estimate of the conditional location μ_t of the dependent variable, and the Beta-t-EGARCH model provides a robust estimate of the log conditional scale λ_t of the dependent variable. For robustness results, we refer to the works of Creal, Koopman, and Lucas (2011, 2013), Harvey (2013), and Blasques, Koopman, and Lucas (2015). The updating term of μ_t in QAR includes the product of the scaling parameter $S_{\mu,t}$ and the conditional score function $\nabla_{\mu,t}$. Similarly, the updating term of λ_t in Beta-t-EGARCH includes the product of the scaling parameter $S_{\lambda,t}$ and the conditional score function $\nabla_{\lambda,t}$. For each model, we consider all possible choices of $S_{\mu,t}$ and $S_{\lambda,t}$ from the literature. We find that the choices of scaling parameters $S_{\mu,t}$ and $S_{\lambda,t}$ from the alternatives are not important, because for all alternatives we get the same estimates of the filters. These results generalize to all score-driven models, for which the scaling parameters are time-invariant.

Second, we use the QAR plus Beta-t-EGARCH model (Harvey 2013; Blazsek and Mendoza 2016), for which $S_{\mu,t}$ depends on the score-driven filter λ_t. For the QAR plus Beta-t-EGARCH model, the score-driven filters μ_t and λ_t are jointly estimated. We consider all possible choices of $S_{\mu,t}$ and $S_{\lambda,t}$ from the literature. We rank the models for alternative scaling parameters according to the log-likelihood (LL). We show that for the QAR plus Beta-t-EGARCH model the choice of the scaling parameters matters, and estimates of μ_t significantly differ for the alternative scaling parameter choices (likewise for λ_t). In particular, we find that for different scaling parameters in the location filter, both the location and scale filters significantly differ for the QAR plus Beta-t-EGARCH model. We find that the choice of different scaling parameters in the scale filter is irrelevant, because all alternative scaling parameters in the score-driven scale filter provide the same LL for a given scaling parameter in the score-driven location filter. For the score-driven location filter, the best choice is the use of I_t^{-1}.

The remainder of this letter is organized as follows: Section 2 presents the specifications of the score-driven models for which the scaling parameters are time-invariant or time-varying. Section 3 describes the data and summarizes the empirical results on scaling parameters. Section 4 concludes.

2. Score-driven models

Score-driven models are observation-driven models (Cox 1981), for which the dynamic parameters are updated by the partial derivatives of the log conditional density of the dependent variables with respect to dynamic parameters. Some of the advantages of the score-driven models over the classical observation-driven models are the following: (i) Score-driven models are robust to outliers and missing data (Harvey 2013). (ii) In many cases, score-driven models are generalizations of classical observation-driven models (Creal, Koopman, and Lucas 2013; Harvey 2013). (iii) A score-driven update locally reduces the Kullback–Leibler divergence in expectation at every step, and only the score-driven updates have this asymptotic property (Blasques, Koopman, and Lucas 2015). In the following, we classify score-driven models of the present paper into two groups: (i) score-driven models with time-invariant scaling parameters, and (ii) score-driven models with time-varying scaling parameters.

2.1. *Time-invariant scaling parameters*

In this section, we formulate score-driven models for which the scaling parameters are constant over time. This involves either the use of a unit scaling parameter or the inverse information matrix. We show that for these models the choice of the scaling parameters $S_{\mu,t}$ and $S_{\lambda,t}$ is irrelevant.

First, the dynamic Student's t location model (Harvey 2013), i.e., the QAR(1) model, is given by

$$y_t = \mu_t + \exp(\lambda)\epsilon_t \tag{1}$$

$$\mu_t = c + \phi\mu_{t-1} + \kappa S_{\mu,t-1}\nabla_{\mu,t-1} \tag{2}$$

$$\epsilon_t \sim t(\nu) \text{ i.i.d.} \tag{3}$$

where μ_t is the conditional location of y_t, $\exp(\lambda)$ is the constant scale parameter of the t-distribution, $\nabla_{\mu,t}$ is the conditional score with respect to μ_t, and $S_{\mu,t}$ is the scaling parameter of the conditional score. The parameters c, ϕ, κ, and λ are $\in \mathbb{R}$, and $\nu > 0$. We assume that $|\phi| < 1$ (i.e., μ_t is asymptotically covariance stationary), and we initialize μ_t by using its unconditional mean: $E(\mu_t) = c/(1-\phi)$.

The log conditional density of $y_t|(y_1,\ldots,y_{t-1},\mu_1) = y_t|\mathcal{F}_{t-1}$ is

$$\ln f(y_t|\mathcal{F}_{t-1}) = \ln\Gamma\left(\frac{\nu+1}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda - \frac{\nu+1}{2}\ln\left[1 + \frac{1}{\nu}\times\frac{(y_t-\mu_t)^2}{\exp(2\lambda)}\right] \tag{4}$$

As a function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda)$, the conditional score with respect to μ_t is:

$$\nabla_{\mu,t} = \frac{\partial \ln f(y_t|\mathcal{F}_{t-1})}{\partial \mu_t} = \frac{(\nu+1)\epsilon_t}{\exp(\lambda)(\nu+\epsilon_t^2)} \tag{5}$$

In Figure 1(a) we present $\nabla_{\mu,t}$ as a function of ϵ_t, to show the nonlinear transformation of the error term. We consider the following four alternatives for the scaling parameter $S_{\mu,t}$ of the conditional score:

$$S_{\mu,t} = (I^{-1})^0 = 1 \tag{6}$$

$$S_{\mu,t} = (I^{-1})^1(\text{DCS}) = \frac{\nu}{\nu+1}\exp(2\lambda) \tag{7}$$

$$S_{\mu,t} = (I^{-1})^1(\text{GAS}) = \frac{\nu+3}{\nu+1}\exp(2\lambda) \tag{8}$$

$$S_{\mu,t} = (I^{-1})^{1/2}(\text{GAS}) = \left[\frac{\nu+3}{\nu+1}\exp(2\lambda)\right]^{1/2} \tag{9}$$

where for all scaling parameters I^{-1} refers to the inverse of the information quantity with respect to μ_t (Harvey 2013). For Equation (7) we refer to Harvey (2013, p. 60), for Equations (8) and (9) we refer to Creal, Koopman, and Lucas (2013, p. 780). Equations (6) to (9) indicate that the information matrix does not depend on μ_t for these score-driven specifications (see also Harvey 2013, Chapter 3).

Second, the Beta-t-EGARCH(1,1) model (Harvey and Chakravarty 2008) is given by

$$y_t = \mu + \exp(\lambda_t)\epsilon_t \tag{10}$$

$$\lambda_t = \omega + \beta\lambda_{t-1} + \alpha S_{\lambda,t-1}\nabla_{\lambda,t-1} \tag{11}$$

$$\epsilon_t \sim t(\nu) \text{ i.i.d.} \tag{12}$$

where $\exp(\lambda_t)$ is the score-driven scale parameter of the t-distribution, $\nabla_{\lambda,t}$ is the conditional score with respect to λ_t, and $S_{\lambda,t}$ is the scaling parameter of the conditional score. We initialize λ_t by using parameter λ_0. The parameters μ, ω, β, α, and λ_0 are $\in \mathbb{R}$, where $|\beta| < 1$ (i.e., λ_t is asymptotically covariance stationary). Moreover, $\nu > 2$, hence the conditional volatility is $\sigma_t = \exp(\lambda_t)[\nu/(\nu-2)]^{1/2}$.

The log conditional density of $y_t|(y_1, \ldots, y_{t-1}, \lambda_1) = y_t|\mathcal{F}_{t-1}$ is

$$\ln f(y_t|\mathcal{F}_{t-1}) = \ln\Gamma\left(\frac{\nu+1}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda_t - \frac{\nu+1}{2}\ln\left[1 + \frac{1}{\nu} \times \frac{(y_t-\mu)^2}{\exp(2\lambda_t)}\right] \tag{13}$$

As a function of $\epsilon_t = (y_t - \mu)\exp(-\lambda_t)$, the conditional score with respect to λ_t is:

$$\nabla_{\lambda,t} = \frac{\partial \ln f(y_t|\mathcal{F}_{t-1})}{\partial \lambda_t} = \frac{(\nu+1)\epsilon_t^2}{\nu + \epsilon_t^2} - 1 \tag{14}$$

In Figure 1(b) we present $\nabla_{\lambda,t}$ as a function of ϵ_t, to show the nonlinear transformation of the error term. We consider the following three alternatives for the scaling parameter $S_{\lambda,t}$:

$$S_{\lambda,t} = (I^{-1})^0(\text{DCS}) = 1 \tag{15}$$

$$S_{\lambda,t} = (I^{-1})^1(\text{GAS}) = \frac{\nu+3}{2\nu} \tag{16}$$

$$S_{\lambda,t} = (I^{-1})^{1/2}(\text{GAS}) = \left[\frac{\nu+3}{2\nu}\right]^{1/2} \tag{17}$$

where I^{-1} refers to the inverse of the information quantity with respect to λ_t. For Equation (15) we refer to Harvey (2013, p. 99), for Equations (16) and (17) we refer to Creal, Koopman, and Lucas (2013, p. 780). The information matrix does not depend on λ_t for these score-driven specifications (Harvey 2013, Chapter 4). Both models are estimated by using the ML method (Blasques et al. 2022).

[APPROXIMATE LOCATION OF FIGURE 1]

2.2. ***Time-varying scaling parameters***

In this section, we formulate score-driven models for which some scaling parameters are time-varying. This involves a transformation of the conditional variance of the score of the LL, where the variance is conditional on $\mathcal{F}_{t-1}$. We formulate the QAR(1) plus Beta-t-EGARCH(1,1) model as follows:

$$y_t = \mu_t + \exp(\lambda_t)\epsilon_t \tag{18}$$

$$\mu_t = c + \phi\mu_{t-1} + \kappa S_{\mu,t-1}\nabla_{\mu,t-1} \tag{19}$$

$$\lambda_t = \omega + \beta\lambda_{t-1} + \alpha S_{\lambda,t-1}\nabla_{\lambda,t-1} \tag{20}$$

$$\epsilon_t \sim t(\nu) \text{ i.i.d.} \tag{21}$$

where μ_t is the conditional location and $\exp(\lambda_t)$ is the score-driven scale of y_t for the t-distribution. Moreover, $\nabla_{\mu,t}$ is the conditional score with respect to μ_t and $S_{\mu,t}$ is its scaling parameter, and $\nabla_{\lambda,t}$ is the conditional score with respect to λ_t and $S_{\lambda,t}$ is its scaling parameter. We initialize μ_t by using its unconditional mean $E(\mu_t) = c/(1-\phi)$, and we initialize λ_t by using parameter λ_0. The parameters c,

ϕ, κ, ω, β, α, and λ_0 are $\in \mathbb{R}$, and $\nu > 2$. Hence the conditional volatility is $\sigma_t = \exp(\lambda_t)[\nu/(\nu-2)]^{1/2}$. We assume that $|\phi| < 1$ and $|\beta| < 1$ (i.e., μ_t and λ_t are asymptotically covariance stationary).

The log conditional density of $y_t|(y_1, \ldots, y_{t-1}, \mu_1) = y_t|\mathcal{F}_{t-1}$ is

$$\ln f(y_t|\mathcal{F}_{t-1}) = \ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda_t - \frac{\nu+1}{2}\ln\left[1 + \frac{1}{\nu} \times \frac{(y_t - \mu_t)^2}{\exp(2\lambda_t)}\right] \quad (22)$$

As a function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda_t)$, the conditional score with respect to μ_t is:

$$\nabla_{\mu,t} = \frac{\partial \ln f(y_t|\mathcal{F}_{t-1})}{\partial \mu_t} = \frac{(\nu+1)\epsilon_t}{\exp(\lambda_t)(\nu + \epsilon_t^2)} \quad (23)$$

We consider the following four alternatives for the scaling parameter $S_{\mu,t}$ of the conditional score:

$$S_{\mu,t} = (I_t^{-1})^0 = 1 \quad (24)$$

$$S_{\mu,t} = (I_t^{-1})^1(\text{DCS}) = \frac{\nu}{\nu+1}\exp(2\lambda_t) \quad (25)$$

$$S_{\mu,t} = (I_t^{-1})^1(\text{GAS}) = \frac{\nu+3}{\nu+1}\exp(2\lambda_t) \quad (26)$$

$$S_{\mu,t} = (I_t^{-1})^{1/2}(\text{GAS}) = \left[\frac{\nu+3}{\nu+1}\exp(2\lambda_t)\right]^{1/2} \quad (27)$$

where I_t^{-1} refers to the conditional variance of the score of the LL with respect to μ_t, and where the variance is conditional on $\mathcal{F}_{t-1}$ (see the work of Creal, Koopman, and Lucas 2013). As a consequence, the scaling parameters depend on the score-driven filter λ_t in Equations (25) to (27).

As a function of $\epsilon_t = (y_t - \mu_t)\exp(-\lambda_t)$, the conditional score with respect to λ_t is:

$$\nabla_{\lambda,t} = \frac{\partial \ln f(y_t|\mathcal{F}_{t-1})}{\partial \lambda_t} = \frac{(\nu+1)\epsilon_t^2}{\nu + \epsilon_t^2} - 1 \quad (28)$$

We consider the following three alternatives for the scaling parameter $S_{\lambda,t}$:

$$S_{\lambda,t} = (I_t^{-1})^0(\text{DCS}) = 1 \quad (29)$$

$$S_{\lambda,t} = (I_t^{-1})^1(\text{GAS}) = \frac{\nu+3}{2\nu} \quad (30)$$

$$S_{\lambda,t} = (I_t^{-1})^{1/2}(\text{GAS}) = \left[\frac{\nu+3}{2\nu}\right]^{1/2} \quad (31)$$

where I_t^{-1} refers to the conditional variance of the score of the LL with respect to λ_t, and where the variance is conditional on $\mathcal{F}_{t-1}$ (see Creal, Koopman, and Lucas 2013). We note that $S_{\lambda,t}$ does not depend on any of the score-driven filters. Nevertheless, this does not imply that λ_t is identical for all alternatives in Equations (29) to (31), because the filters μ_t and λ_t are jointly estimated.

The model is estimated by using the ML method (Blazsek, Escribano, and Licht 2022).

3. Empirical analysis

We model the daily return of a financial asset defined as the log percentage change in the opening price as follows: $y_t = \ln(p_t/p_{t-1})$ for $t = 1, \ldots, T$ days. We assume that p_0 is from pre-sample data. We use data on daily Vanguard S&P 500 ETF return y_t for the period of September 10, 2010 to February 9, 2022 (data source: Yahoo Finance; ticker: VOO). Variable y_t is the log percentage change in the opening price of VOO. Descriptive statistics are presented in Table 1(a). The parameter estimates for the AR(p)-normal-GARCH(1,1) specifications with $p = 0, 1, 2$, which are presented in Table 1(b), motivate the lag-order selection for the QAR(1) and Beta-t-EGARCH(1,1) models for the score-driven specifications. In Figure 2, we present the VOO opening price and return for the sample period.

For the score-driven models for which the scaling parameters are time-invariant (Section 2.1), the ML estimates of QAR(1) and Beta-t-EGARCH(1,1) are presented in Tables 2(a) and 2(b), respectively. The results show that the LL estimates are identical for all alternative choices of $S_{\mu,t}$ or $S_{\lambda,t}$. We also find that the filtered signal estimates, i.e., μ_t for QAR(1) and λ_t for Beta-t-EGARCH(1,1) coincide for all choices of $S_{\mu,t}$ or $S_{\lambda,t}$ (Figure 3). Different choices of $S_{\mu,t}$ and $S_{\lambda,t}$ result in different estimates of κ and α, respectively, which ensure exactly the same LL and the same score-driven filters. The robust standard error-based z-ratios of parameters, which correspond to alternative choices of $S_{\mu,t}$ and $S_{\lambda,t}$, are very similar for Table 2. Hence, different choices of the scaling parameter provide similarly precise filtered signal estimates. The conclusions for QAR and Beta-t-EGARCH can be extended to the score-driven models with time-invariant scaling parameters (see Theorem 1 and its generalizations in Harvey 2013). For those models, the choice of the scaling parameters $S_{\mu,t}$ and $S_{\lambda,t}$ is irrelevant.

For the score-driven models with time-varying scaling parameters (Section 2.2), the ML estimates of QAR(1) plus Beta-t-EGARCH(1,1) are presented in Table 3. The results show that the LL estimates differ for alternative choices of $S_{\mu,t}$ and $S_{\lambda,t}$. In the notes of Table 3, we present the ranking of alternative specifications with respect to the LL estimates. The best-performing scaling parameter uses $S_{\mu,t} = (I_t^{-1})^1$ and the choice of $S_{\lambda,t}$ is irrelevant. It is noteworthy that the two alternatives $S_{\mu,t} = (I_t^{-1})^1$ (DCS) and $S_{\mu,t} = (I_t^{-1})^1$ (GAS) have identical performance. With respect to $S_{\mu,t} = (I_t^{-1})^1$, we note that this is the choice of Harvey (2013) for the score-driven location models, and Creal, Koopman, and Lucas (2013) present that a score-driven model with I_t^{-1} scaling parameter encompasses a classical observation driven model; in this case, that is the AR moving average (ARMA) model. The $S_{\mu,t} = (I_t^{-1})^1$ choice is followed by $S_{\mu,t} = (I_t^{-1})^{0.5}$ and the choice of $S_{\lambda,t}$ is also irrelevant. With respect to $S_{\mu,t} = (I_t^{-1})^{0.5}$, we note that Creal, Koopman, and Lucas (2013) present that for this choice $S_{\mu,t}\nabla_{\mu,t}$ has a constant unit variance and the statistical properties of μ_t are more tractable. The worst performing choice from the scaling parameters is $S_{\mu,t} = (I_t^{-1})^0$ and the choice of $S_{\lambda,t}$ is irrelevant, as before. The latter choice of the scaling parameter for score-driven location filters in not common in the literature, which may be explained by our scaling parameter selection results for QAR(1) plus Beta-t-EGARCH(1,1).

In addition, for the score-driven models with time-varying scaling parameters (Section 2.2), we also find that the estimates of μ_t and λ_t differ for QAR(1) plus Beta-t-EGARCH(1,1) for alternative choices of $S_{\mu,t}$ and $S_{\lambda,t}$, respectively (Figure 4). The best-performing scaling parameter $S_{\mu,t} = (I_t^{-1})^1$ corresponds to the most volatile filtered estimate of μ_t (i.e., the black line in Figure 4(a)). In the notes of

Figure 4, we report robust statistical test results on the differences between two alternative μ_t estimates for alternative pairs of the scaling parameters. In the same notes we report robust statistical test results on the differences between two alternative λ_t estimates for alternative pairs of the scaling parameters. To use a robust statistical test, we regress the differences on a constant using the heteroskedasticity and autocorrelation consistent (HAC) ordinary least squares (OLS) estimator (Newey and West 1987). We note that we prefer the comparison of the alternative μ_t and λ_t estimates to the comparison of the LL estimates for alternative models, because μ_t and λ_t are relevant in financial applications.

In summary, for the QAR plus Beta-t-EGARCH model we find that for the score-driven location filter the statistical performance of the scaling parameter $S_{\mu,t} = (I_t^{-1})^1$ is superior to the statistical performances of $S_{\mu,t} = (I_t^{-1})^0$ and $S_{\mu,t} = (I_t^{-1})^{0.5}$. For each of those scaling parameters of the location filter, the choice of the scaling parameter of the scale filter is irrelevant. These results validate the main question of the present paper on the choice of the scaling parameters in score-driven filters.

[APPROXIMATE LOCATION OF TABLES 1 TO 3 AND FIGURES 2 TO 4]

4. Conclusions

We have compared the performances of score-driven models using alternative choices of the scaling parameters in score-driven filters from the literature. We have obtained two main results:

(i) For score-driven models with time-invariant scaling parameters, we have shown for all alternative scaling parameters of the literature that the choice of the scaling parameters is irrelevant. We have shown this result for separately estimated QAR(1) and Beta-t-EGARCH(1,1) models by using daily VOO return data for the period of September 10, 2010 to February 9, 2022.

(ii) For score-driven models with time-varying scaling parameters, we have also shown for all alternative scaling parameters of the literature that the choice of the scaling parameters is relevant. We have shown this result for the QAR(1) plus Beta-t-EGARCH(1,1) model, for which the scaling parameters are time-varying, by using the same VOO dataset. For different scaling parameters in the location filter, we have shown that both the location and scale filters differ significantly and we have ranked the statistical performances of alternative QAR plus Beta-t-EGARCH specifications. For the score-driven location filter, we have found that the best choice is the use of the conditional inverse information matrix for the scaling parameter. Our results on the scaling parameters can be extended to more general score-driven models in which the conditional information matrix depends on more than one score-driven filters, motivating future work.

Acknowledgements

The authors gratefully appreciate the comments of Matthew Copley. All remaining errors are our own. Funding from the School of Business of Universidad Francisco Marroquín is gratefully acknowledged. No potential conflict of interest was reported by the authors. Data source is reported in the paper, and data are available from the authors upon request. Computer codes are available from the authors upon request. The authors express their consent for publication.

References

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas. 2022. Maximum likelihood estimation for score-driven models. *Journal of Econometrics* 227 (2):325–346. doi:10.1016/j.jeconom.2021.06.003.

Blasques, F., S. J. Koopman, and A. Lucas. 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* 102 (2):325–343. doi:10.1093/biomet/asu076.

Blazsek, S., A. Escribano, and A. Licht. 2022. Score-driven location plus scale models: asymptotic theory and an application to forecasting Dow Jones volatility. *Studies in Nonlinear Dynamics & Econometrics*. doi:10.1515/snde-2021-0083.

Blazsek, S., and V. Mendoza. 2016. QARMA-Beta-t-EGARCH versus ARMA-GARCH: an application to S&P 500. *Applied Economics* 48 (12):1119–1129. doi:10.1080/00036846.2015.1093086.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (3):307–327. doi:10.1016/0304-4076(86)90063-1.

Box, G. E. P., and G. M. Jenkins. 1970. *Time series analysis, forecasting and control*. San Francisco: Holden-Day.

Cox, D. R. 1981. Statistical analysis of time series: some recent developments (with discussion and reply). *Scandinavian Journal of Statistics* 8 (2):93–115. https://www.jstor.org/stable/4615819.

Creal, D., S. J. Koopman, and A. Lucas. 2008. A general framework for observation driven time-varying parameter models. Tinbergen Institute Discussion Paper 08-108/4. https://papers.tinbergen.nl/08108.pdf.

Creal, D., S. J. Koopman, and A. Lucas. 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* 29 (4):552–563. doi:10.1198/jbes.2011.10070.

Creal, D., S. J. Koopman, and A. Lucas. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28 (5):777–795. doi:10.1002/jae.1279.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4):987–1007. doi:10.2307/1912773.

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

Harvey, A. C., and T. Chakravarty. 2008. Beta-t-(E)GARCH. Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge. https://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe0840.pdf.

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (3):703–708. doi:10.2307/1913610.

Table 1. Descriptive statistics.

(a). Statistics	VOO return		
Start date	September 10, 2010		
End date	February 9, 2022		
Sample size	$2,875$		
Minimum	-0.0884		
Maximum	0.0612		
Average	0.0005		
Standard deviation	0.0100		
Skewness	-0.9409		
Excess kurtosis	8.6957		
(b). GARCH for VOO	constant plus normal-GARCH(1,1)	AR(1)-normal-GARCH(1,1)	AR(2)-normal-GARCH(1,1)
c	$0.0007^{***}(0.0007)$	$0.0007^{***}(0.0001)$	$0.0007^{***}(0.0001)$
ϕ_1	NA	$-0.0522^{**}(0.0209)$	$-0.0523^{**}(0.0210)$
ϕ_2	NA	NA	$-0.0176(0.0216)$
ω	$0.0000^{***}(0.0000)$	$0.0000^{***}(0.0000)$	$0.0000^{***}(0.0000)$
α	$0.1717^{***}(0.0327)$	$0.1722^{***}(0.0330)$	$0.1724^{***}(0.0331)$
β	$0.8040^{***}(0.0328)$	$0.8029^{***}(0.0332)$	$0.8030^{***}(0.0333)$

Notes: Autoregressive (AR); generalized AR conditional heteroskedasticity (GARCH); not available (NA). The most general AR(2)-normal-GARCH(1,1) model in Panel (b) is: $y_{\text{VOO}} = \mu_t + v_t = \mu_t + \sigma_t \epsilon_t$; $\epsilon_t \sim N(0,1)$ i.i.d.; $\mu_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2}$; $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha v_{t-1}^2$ (Box and Jenkins 1970; Engle 1982; Bollerlev 1986). Robust standard errors, by using the Huber sandwich estimator, are reported in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

Table 2. Parameter estimates for score-driven models with time-invariant scaling parameters.

(a). QAR(1)	$S_{\mu,t} = (I^{-1})^0$	$S_{\mu,t} = (I^{-1})^1$ (DCS)	$S_{\mu,t} = (I^{-1})^1$ (GAS)	$S_{\mu,t} = (I^{-1})^{1/2}$ (GAS)
c	$0.0012^{***}(0.0003)$	$0.0012^{***}(0.0003)$	$0.0012^{***}(0.0003)$	$0.0012^{***}(0.0003)$
ϕ	$-0.1648(0.3050)$	$-0.1648(0.3049)$	$-0.1648(0.3048)$	$-0.1648(0.3044)$
κ	$\mathbf{-0.0000^{**}(0.0000)}$	$\mathbf{-0.0947^{**}(0.0461)}$	$\mathbf{-0.0460^{**}(0.0224)}$	$\mathbf{-0.0003^{**}(0.0002)}$
λ	$-5.1101^{***}(0.0267)$	$-5.1101^{***}(0.0267)$	$-5.1101^{***}(0.0267)$	$-5.1101^{***}(0.0267)$
ν	$2.8397^{***}(0.1695)$	$2.8396^{***}(0.1693)$	$2.8396^{***}(0.1693)$	$2.8397^{***}(0.1693)$
LL	**3.3157**	**3.3157**	**3.3157**	**3.3157**
AIC	-6.6280	-6.6280	-6.6280	-6.6280
BIC	-6.6176	-6.6176	-6.6176	-6.6176
HQC	-6.6243	-6.6243	-6.6243	-6.6243
(b). Beta-t-EGARCH(1,1)	$S_{\lambda,t} = (I^{-1})^0$ (DCS)	$S_{\lambda,t} = (I^{-1})^1$ (GAS)		$S_{\lambda,t} = (I^{-1})^{1/2}$ (GAS)
μ	$0.0009^{***}(0.0001)$	$0.0009^{***}(0.0001)$		$0.0009^{***}(0.0001)$
ω	$-0.1660^{***}(0.0417)$	$-0.1660^{***}(0.0417)$		$-0.1660^{***}(0.0417)$
α	$\mathbf{0.0897^{***}(0.0105)}$	$\mathbf{0.1199^{***}(0.0134)}$		$\mathbf{0.1037^{***}(0.0117)}$
β	$0.9668^{***}(0.0083)$	$0.9668^{***}(0.0083)$		$0.9668^{***}(0.0083)$
ν	$6.0506^{***}(0.6700)$	$6.0507^{***}(0.6697)$		$6.0507^{***}(0.6707)$
λ_0	$-4.8936^{***}(0.2390)$	$-4.8937^{***}(0.2390)$		$-4.8933^{***}(0.2389)$
LL	**3.4189**	**3.4189**		**3.4189**
AIC	-6.8337	-6.8337		-6.8337
BIC	-6.8212	-6.8212		-6.8212
HQC	-6.8292	-6.8292		-6.8292

Notes: Dynamic conditional score (DCS); generalized autoregressive score (GAS); log-likelihood (LL); Akaike information criterion (AIC); Bayesian information criterion (BIC); Hannan–Quinn criterion (HQC). The most relevant estimates are highlighted by bold numbers. Robust standard errors, by using the Huber sandwich estimator, are in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

Table 3. Parameter estimates for score-driven models with time-varying scaling parameters.

(a). QAR(1) plus Beta-t-EGARCH(1,1)	$S_{\mu,t} = (I^{-1})^0$ $S_{\lambda,t} = (I^{-1})^0$ (DCS)	$S_{\mu,t} = (I^{-1})^1$ (DCS) $S_{\lambda,t} = (I^{-1})^0$ (DCS)	$S_{\mu,t} = (I^{-1})^1$ (GAS) $S_{\lambda,t} = (I^{-1})^0$ (DCS)	$S_{\mu,t} = (I^{-1})^{1/2}$ (GAS) $S_{\lambda,t} = (I^{-1})^0$ (DCS)
c	$0.0001^{*}(0.0001)$	$0.0001^{***}(0.0000)$	$0.0001^{***}(0.0000)$	$0.0001^{**}(0.0000)$
ϕ	$0.8962^{***}(0.0606)$	$0.9115^{***}(0.0343)$	$0.9115^{***}(0.0343)$	$0.9055^{***}(0.0399)$
κ	$0.0000^{***}(0.0000)$	$-0.0445^{***}(0.0138)$	$-0.0302^{***}(0.0093)$	$-0.0002^{***}(0.0001)$
ω	$-0.1670^{***}(0.0338)$	$-0.1595^{***}(0.0331)$	$-0.1595^{***}(0.0331)$	$-0.1625^{***}(0.0334)$
α	$0.0886^{***}(0.0079)$	$0.0884^{***}(0.0078)$	$0.0884^{***}(0.0078)$	$0.0891^{***}(0.0079)$
β	$0.9666^{***}(0.0067)$	$0.9680^{***}(0.0066)$	$0.968^{***}(0.0066)$	$0.9675^{***}(0.0066)$
ν	$6.0920^{***}(0.5888)$	$6.3166^{***}(0.6113)$	$6.3166^{***}(0.6113)$	$6.1848^{***}(0.6025)$
λ_0	$-0.1642^{***}(0.0411)$	$-0.1569^{***}(0.0399)$	$-0.1569^{***}(0.04)$	$-0.1599^{***}(0.0405)$
LL	3.4205	**3.4215**	**3.4215**	*3.4214*
AIC	−6.8355	−6.8375	−6.8375	−6.8372
BIC	−6.8189	−6.8209	−6.8209	−6.8206
HQC	−6.8295	−6.8315	−6.8315	−6.8312
(b). QAR(1) plus Beta-t-EGARCH(1,1)	$S_{\mu,t} = (I^{-1})^0$ $S_{\lambda,t} = (I^{-1})^1$ (GAS)	$S_{\mu,t} = (I^{-1})^1$ (DCS) $S_{\lambda,t} = (I^{-1})^1$ (GAS)	$S_{\mu,t} = (I^{-1})^1$ (GAS) $S_{\lambda,t} = (I^{-1})^1$ (GAS)	$S_{\mu,t} = (I^{-1})^{1/2}$ (GAS) $S_{\lambda,t} = (I^{-1})^1$ (GAS)
c	$0.0001^{*}(0.0001)$	$0.0001^{***}(0.0000)$	$0.0001^{***}(0.0000)$	$0.0001^{**}(0.0000)$
ϕ	$0.8966^{***}(0.0602)$	$0.9115^{***}(0.0343)$	$0.9115^{***}(0.0343)$	$0.9055^{***}(0.0399)$
κ	$0.0000^{***}(0.0000)$	$-0.0445^{***}(0.0138)$	$-0.0302^{***}(0.0093)$	$-0.0002^{***}(0.0001)$
ω	$-0.1671^{***}(0.0338)$	$-0.1595^{***}(0.0331)$	$-0.1595^{***}(0.0331)$	$-0.1625^{***}(0.0334)$
α	$0.1188^{***}(0.0100)$	$0.1199^{***}(0.0100)$	$0.1199^{***}(0.01)$	$0.1200^{***}(0.0101)$
β	$0.9666^{***}(0.0067)$	$0.9680^{***}(0.0066)$	$0.968^{***}(0.0066)$	$0.9675^{***}(0.0066)$
ν	$6.0905^{***}(0.5884)$	$6.3167^{***}(0.6113)$	$6.3166^{***}(0.6113)$	$6.1848^{***}(0.6025)$
λ_0	$-0.1643^{***}(0.0411)$	$-0.1569^{***}(0.0400)$	$-0.1569^{***}(0.04)$	$-0.1599^{***}(0.0405)$
LL	3.4205	**3.4215**	**3.4215**	*3.4214*
AIC	−6.8355	−6.8375	−6.8375	−6.8372
BIC	−6.8189	−6.8209	−6.8209	−6.8206
HQC	−6.8295	−6.8315	−6.8315	−6.8312
(c). QAR(1) plus Beta-t-EGARCH(1,1)	$S_{\mu,t} = (I^{-1})^0$ $S_{\lambda,t} = (I^{-1})^{1/2}$ (GAS)	$S_{\mu,t} = (I^{-1})^1$ (DCS) $S_{\lambda,t} = (I^{-1})^{1/2}$ (GAS)	$S_{\mu,t} = (I^{-1})^1$ (GAS) $S_{\lambda,t} = (I^{-1})^{1/2}$ (GAS)	$S_{\mu,t} = (I^{-1})^{1/2}$ (GAS) $S_{\lambda,t} = (I^{-1})^{1/2}$ (GAS)
c	$0.0001^{*}(0.0001)$	$0.0001^{***}(0.0000)$	$0.0001^{***}(0.0000)$	$0.0001^{**}(0.0000)$
ϕ	$0.8962^{***}(0.0606)$	$0.9115^{***}(0.0343)$	$0.9115^{***}(0.0343)$	$0.9055^{***}(0.0399)$
κ	$0.0000^{***}(0.0000)$	$-0.0445^{***}(0.0138)$	$-0.0302^{***}(0.0093)$	$-0.0002^{***}(0.0001)$
ω	$-0.1670^{***}(0.0338)$	$-0.1595^{***}(0.0331)$	$-0.1595^{***}(0.0331)$	$-0.1625^{***}(0.0334)$
α	$0.1026^{***}(0.0087)$	$0.1029^{***}(0.0087)$	$0.1029^{***}(0.0087)$	$0.1034^{***}(0.0088)$
β	$0.9666^{***}(0.0067)$	$0.9680^{***}(0.0066)$	$0.968^{***}(0.0066)$	$0.9675^{***}(0.0066)$
ν	$6.0921^{***}(0.5888)$	$6.3166^{***}(0.6113)$	$6.3166^{***}(0.6113)$	$6.1849^{***}(0.6025)$
λ_0	$-0.1642^{***}(0.0411)$	$-0.1569^{***}(0.0400)$	$-0.1569^{***}(0.04)$	$-0.1599^{***}(0.0405)$
LL	3.4205	**3.4215**	**3.4215**	*3.4214*
AIC	−6.8355	−6.8375	−6.8375	−6.8372
BIC	−6.8189	−6.8209	−6.8209	−6.8206
HQC	−6.8295	−6.8315	−6.8315	−6.8312

Notes: Dynamic conditional score (DCS); generalized autoregressive score (GAS); log-likelihood (LL); Akaike information criterion (AIC); Bayesian information criterion (BIC); Hannan–Quinn criterion (HQC). Ranking of LL: 1-bold, 2-italic, and 3-normal. Robust standard errors (Huber) are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

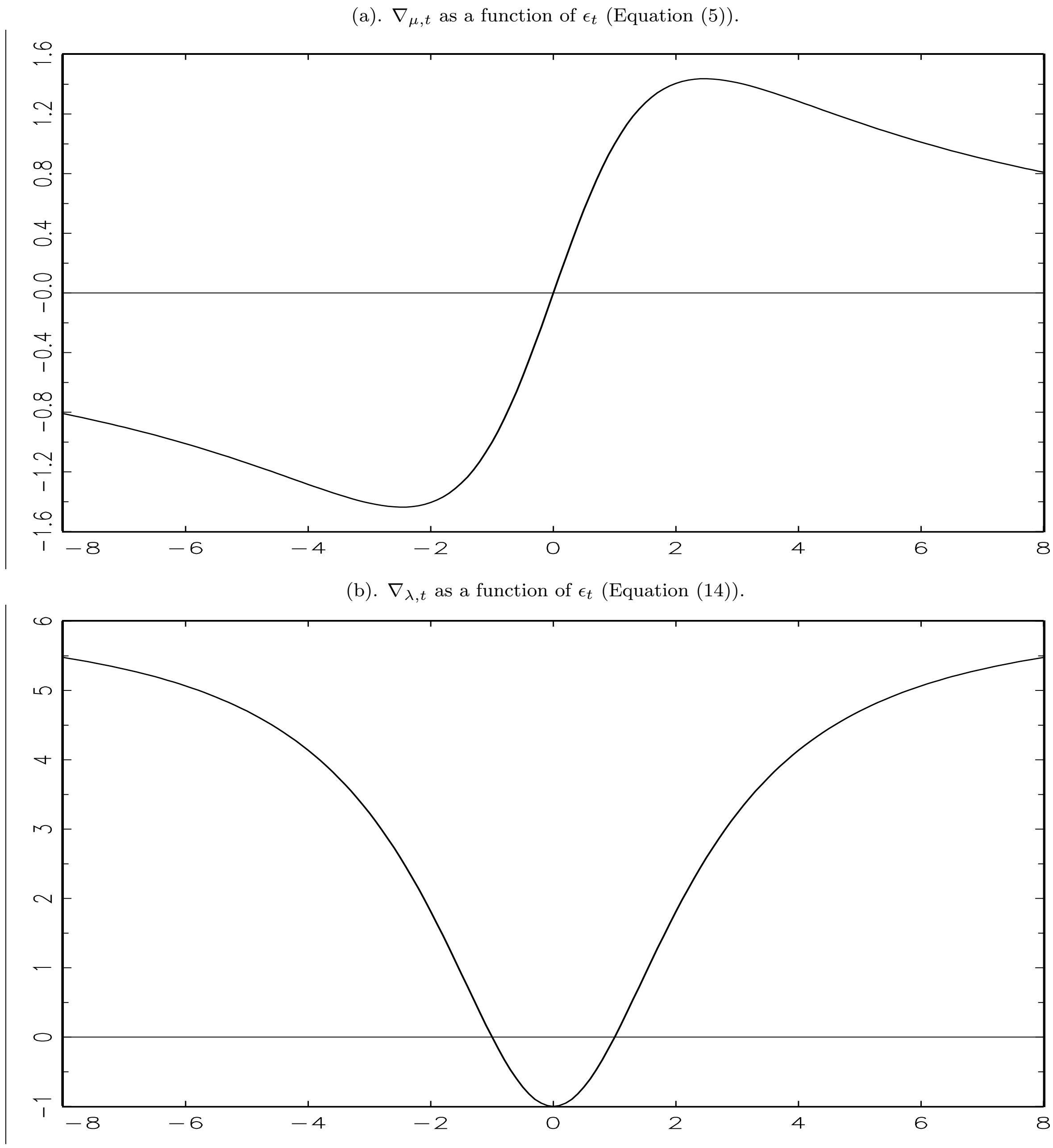


Figure 1. Conditional scores with respect to μ_t and λ_t. Notes: We use the estimates of ν, and we assume that $\lambda = 0$ for Panel (a).

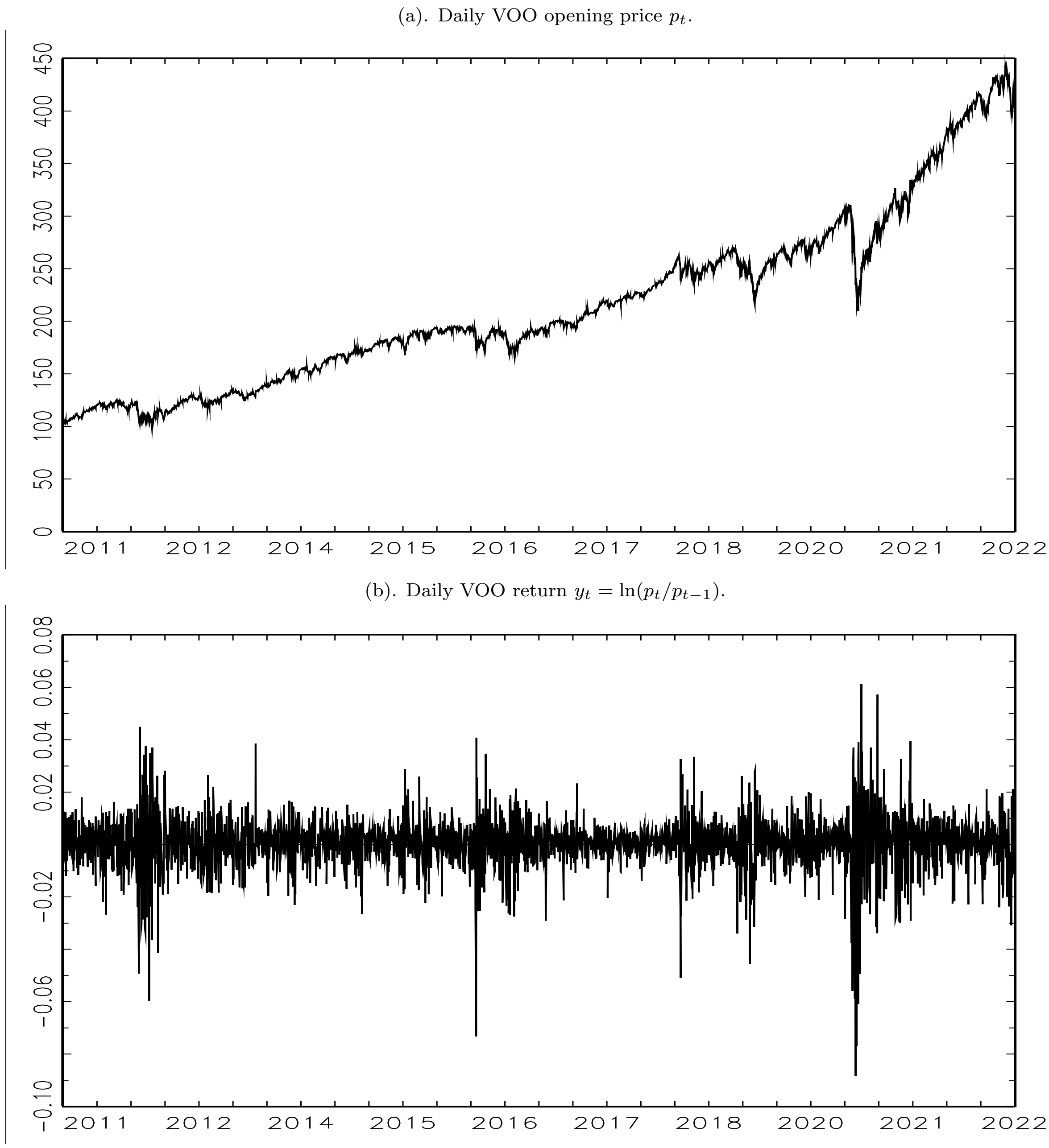
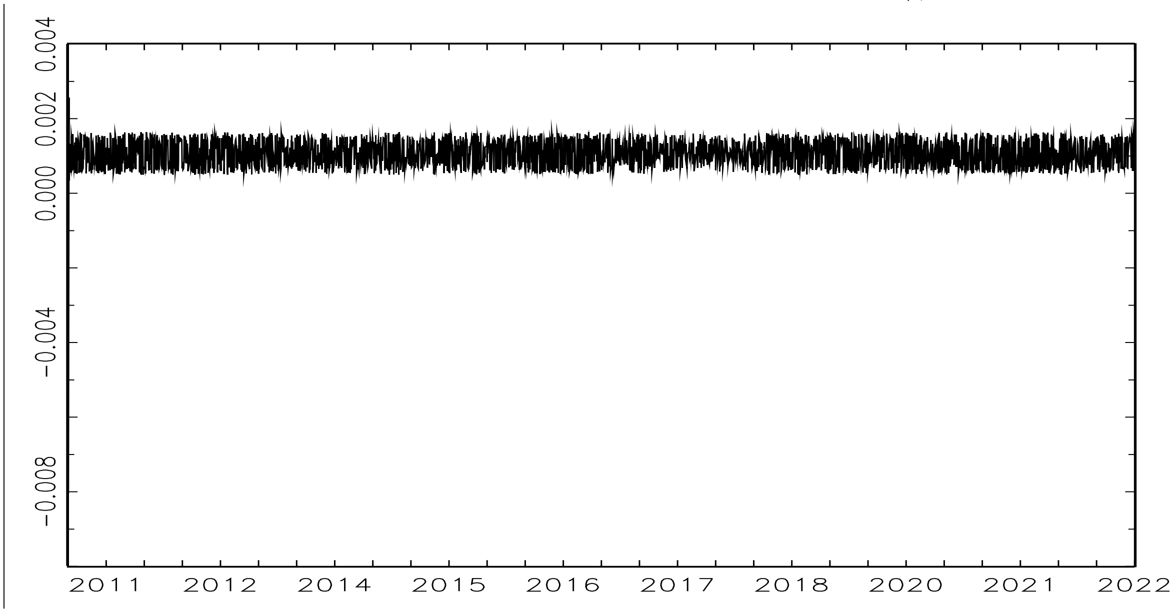


Figure 2. VOO opening price and return for the period of September 10, 2010 to February 9, 2022.

(a). μ_t estimates for the QAR(1) model for all alternatives of $S_{\mu,t}$.



(b). λ_t estimates for the Beta-t-EGARCH(1,1) model for all alternatives of $S_{\lambda,t}$.

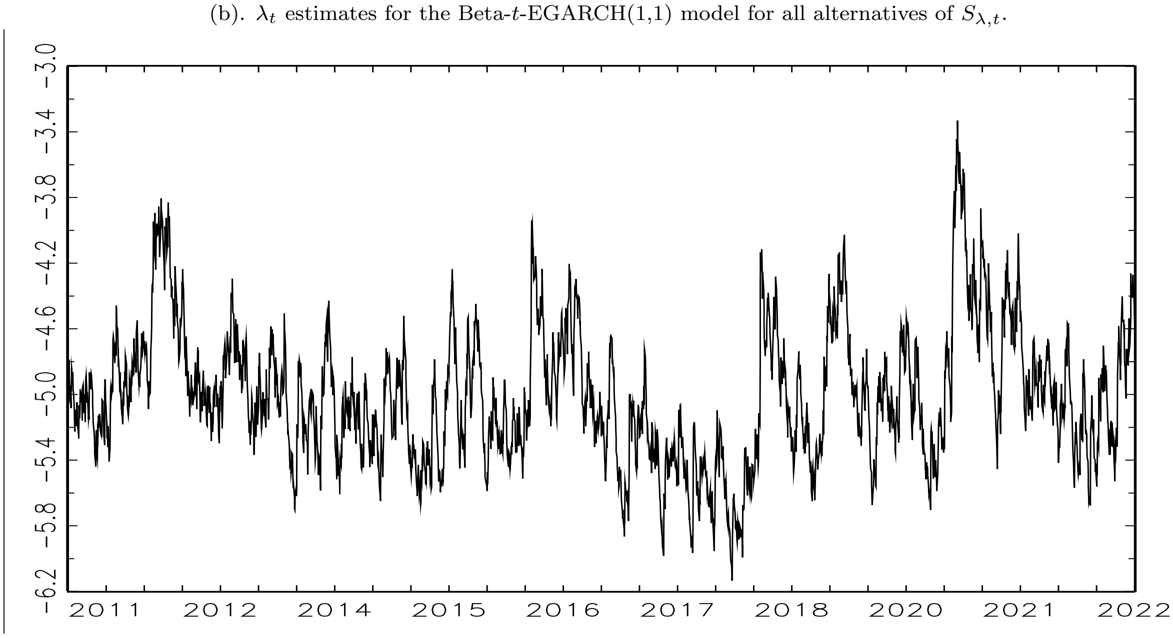
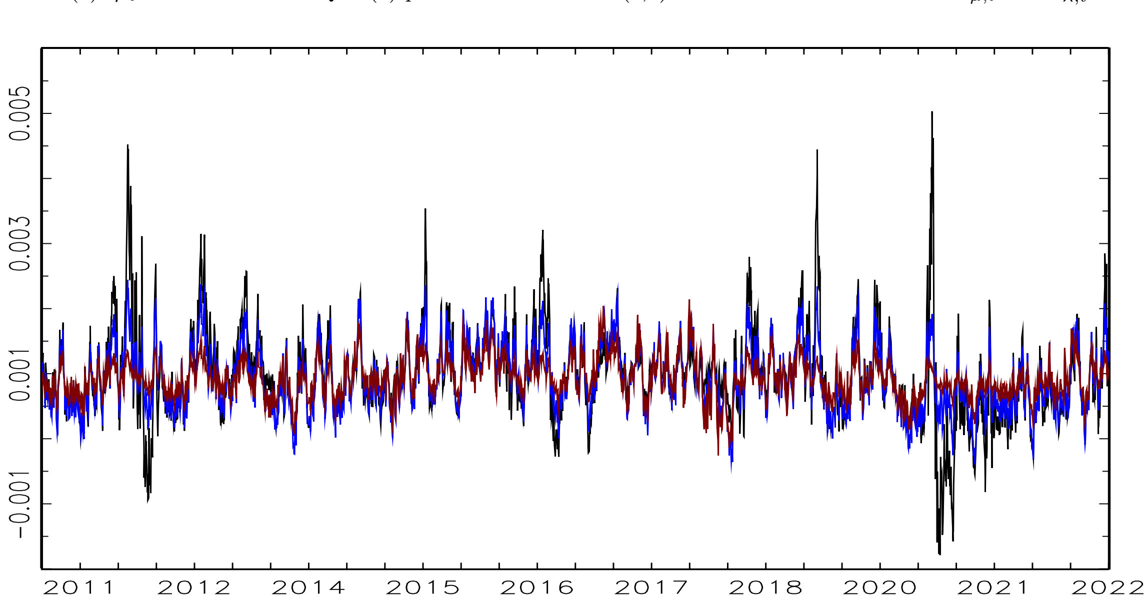


Figure 3. Estimates μ_t and λ_t for score-driven models with time-invariant scaling parameters. Notes: We present all the estimates of μ_t which coincide for all alternatives of $S_{\mu,t}$. We present all the estimates of λ_t which coincide for all alternatives of $S_{\lambda,t}$.

(a). μ_t estimates for the QAR(1) plus Beta-t-EGARCH(1,1) model for all alternatives of $S_{\mu,t}$ and $S_{\lambda,t}$.



(b). λ_t estimates for the QAR(1) plus Beta-t-EGARCH(1,1) model for all alternatives of $S_{\mu,t}$ and $S_{\lambda,t}$.

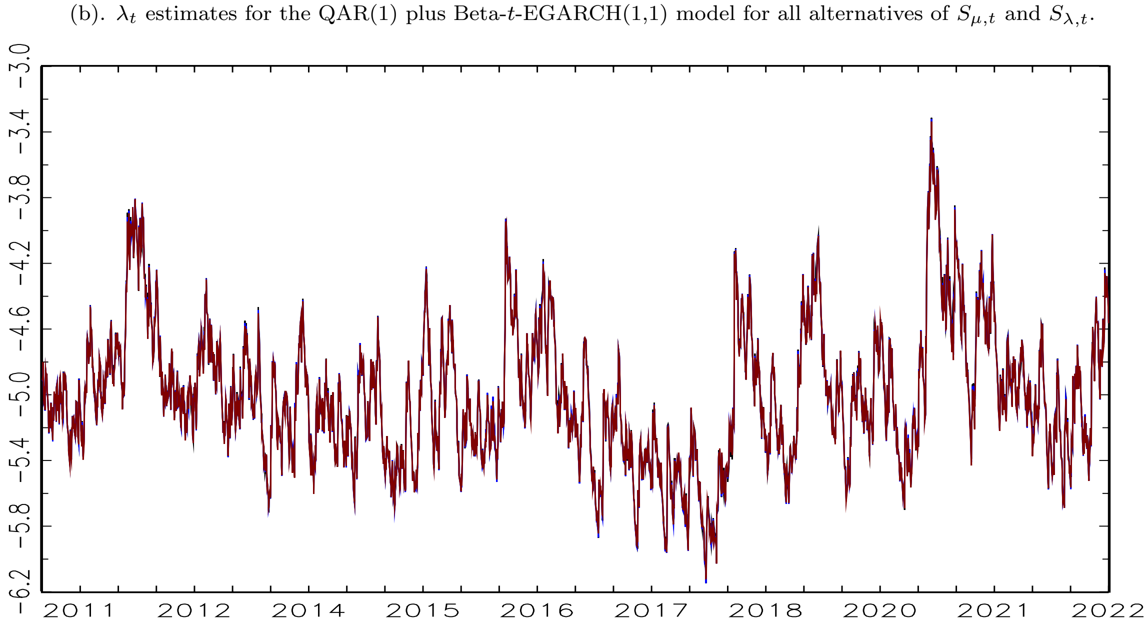


Figure 4. Estimates μ_t and λ_t for score-driven models with with time-varying scaling parameters. Notes: The black line represents all μ_t and λ_t for which $S_{\mu,t} = (I_t^{-1})^1$ (DCS) (Equation (25)) or $S_{\mu,t} = (I_t^{-1})^1$ (GAS) (Equation (26)) and any choice of $S_{\lambda,t}$ (Equations (29) to (31)). The blue line represents all μ_t and λ_t for which $S_{\mu,t} = (I_t^{-1})^{0.5}$ (GAS) (Equation (27)) and any choice of $S_{\lambda,t}$ (Equations (29) to (31)). The red line represents all μ_t and λ_t for which $S_{\mu,t} = (I_t^{-1})^0$ (DCS) (Equation (24)) and any choice of $S_{\lambda,t}$ (Equations (29) to (31)). Ranking according to LL estimates: 1-black, 2-blue, and 3-red. Robust tests statistic (p-value in parentheses) of the difference between μ_t for black and blue: $2.0309^{**}(0.0424)$. Robust tests statistic (p-value in parentheses) of the difference between μ_t for black and red: $1.5040^{+}(0.1327)$. Robust tests statistic (p-value in parentheses) of the difference between λ_t for black and blue: $9.2621^{***}(0.0000)$. Robust tests statistic (p-value in parentheses) of the difference between λ_t for black and red: $8.4833^{***}(0.0000)$. $^{+}$, **, and *** denotes significance at the 15%, 5%, and 1% levels, respectively.