Applying the Non-linear Transformation Families to the Lagged-variance of EGARCH and GJR Models

Didit B. Nugroho, Tundjung Mahatma, Yulius Pratomo

Abstract—**This study performed four non-linear transformations, including Tukey, Exponential, Modulus, and Yeo–Johnson to the lagged-variance, and four skewed and/or heavy-tailed distributions, including Student-t, Hansen's skewed Student-t (HST), Fernández–Steel's skewed Student-t (FST), and skewed generalized error distribution (SGED) to error distribution in estimating GARCH, GJR, and EGARCH models. The Adaptive Random Walk Metropolis (ARWM) method was constructed in the Markov Chain Monte Carlo (MCMC) algorithm to perform the Bayesian inference of models. On the basis of empirical study of Swiss Market Index, the results provide strong evidence that ARWM method can be a statistically efficient MCMC method for the Bayesian inference of the GARCH-type models. On comparing the estimated models based on the Log-likelihood Ratio Test (LRT) and Deviance Information Criterion (DIC), we found the superiority of volatility fitting for the models with non-linear transformation. In all considered distributions, the exponential transformation produces the best performance in fitting the GARCH and GJR models, meanwhile the modulus transformation yields the best performance in fitting the other one. DIC suggested the EGARCH models with HST distribution based on the modulus transformation in the lagged-variance as the best fit model.**

Index Terms—**GARCH, heavy-tailedness, non-linear transformations, skewness.**

I. Introduction

In time series and econometrics modeling, the class of Generalized Autoregressive Conditional Heteroskedasticity (known as GARCH) introduced in [1] is the most popular model to describe the time-varying variances of the observations. Models from the GARCH class specifies the current variance as an exact function of past errors and variances. Let R_t be the observable return of an asset at discrete time t with the associated conditional volatility of σ_t.

Manuscript received March 13, 2021; revised August 4, 2021. This work was funded by: (1) Higher Education Excellent Fundamental Research Grant of The Ministry of Research, Technology and Higher Education of the Republic of Indonesia for the 2020 fiscal year (Contract Number 001/LL6/PG/SP2H/AMD/PL.1/2020), and (2) Universitas Kristen Satya Wacana's Internal Funding for the 2020 research (Contract Number 140.1/SPK-PDUPT/PR V/6/2020).

D. B. Nugroho is an Associate Professor of Mathematics and Data Science Department, Universitas Kristen Satya Wacana, Jl. Diponegoro 52–60 Salatiga 50711, Jawa Tengah, Indonesia (e-mail: didit.budinugroho@uksw.edu; didit.budinugroho@staff.uksw.edu). ORCID iD: https://orcid.org/0000-0002-3693-0950.

T. Mahatma is an Assistant Professor of Mathematics and Data Science Department, Universitas Kristen Satya Wacana, Jl. Diponegoro 52–60 Salatiga 50711, Jawa Tengah, Indonesia (e-mail: tundjung.mahatma@uksw.edu)

Y. Pratomo is an Assistant Professor of Economics Department, Universitas Kristen Satya Wacana, Jl. Diponegoro 52–60 Salatiga 50711, Jawa Tengah, Indonesia (e-mail: yulius.pratomo@uksw.edu). ORCID iD: https://orcid.org/0000-0002-0406-3047

A general specification for the returns process that exhibits time-varying volatilities is given by

$$R_t = \sigma_t \epsilon_t,$$

in which ϵ_t (the errors) are i.i.d. (independent and identically distributed) with expectation 0 and variance 1. The basic GARCH(1,1) model is the most popular to model conditional variance (squared volatility) in the empirical study, where the process is given as:

$$\sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2. \qquad (1)$$

The GARCH(1,1) process in Eq. (1) defines the conditional variance at current as a linear function of returns and variances from the past. In the last three decades, several non-linear variants have been derived from the basic GARCH model to make good estimate and forecast of future variance in financial markets. Most of them are developed to capture such aspects as the asymmetric relation between conditional variance and previous observations, see [2] for a brief survey and discussion. The two most popular asymmetric GARCH models are perhaps the exponential GARCH (EGARCH) model studied in [3] and the GJR model given in [4]. This study therefore focuses on the EGARCH and GJR models. The EGARCH model takes the natural logarithmic value of the conditional variance and allows observations of different signs to have a different effect on the natural logarithmic of variance. Meanwhile, the GJR model extends the basic GARCH model by adding a term which allows for asymmetric effects of variance in term of the sign of the past observation.

Another version of nonlinear ARCH model was offered in [5] by taking the Box–Cox power transformation for the current variance and the past observation. This model is then known as NARCH model. Similar to that approach, study in [6] developed the APARCH (asymmetric power ARCH) model, which is the extension of GJR model, by taking conventional power transformation for the variance on both sides of the GARCH process. Recently, study in [7] applied the Tukey transformation to the lag of the variance process in Eq. (1) and showed that the proposed model better performing than the basic GARCH model. The approach of [7] is similar to [8] applied several non-linear transformations to the lag of log-variance in the context of stochastic volatility process.

When one estimates the conditional variance using the time series models in the basic case, the error process (ϵ_t) follows a Normal distribution. However, in reality, the financial time series data do not come from a Normal distribution, but the series tend to exhibit properties such as skewness and kurtosis. To overcome this problem, many researchers perform skewed and/or heavy-tailed distributions to make GARCH modeling more successful in fitting the time series data. In the case of heavy-tailed distribution, study in [9] showed that the Student-t distribution and generalized error distribution (GED) outperform the Normal distribution. Meanwhile, to model both skewness and kurtosis characteristics, three different distributions have been proposed. The three distributions: generalized Student-t of [10], skewed Student-t of [11], and skewed GED of [12], were proposed to provide a flexible distribution for modeling the empirical distribution of financial time series data in the ARCH, regression, and absolute GARCH-in-mean models, respectively. In particular, study in [13] compared the Normal, Skew Normal of [14], Skewed Student-t, and Skewed GED and found empirical evidence to support skewed distributions for the return error in GARCH model.

Motivated by the previous works, the objective of this study is to introduce an alternative non-linear class to the EGARCH and GJR models. First, we incorporate the non-linear transformations such as Extended Tukey, Extended Box–Cox, Exponential, Modulus, and Yeo-Johnson into the lag of the log-variance in the EGARCH process. These transformations are applied to accommodate negative value of log-variance. Meanwhile, the simple Tukey, Box–Cox, Exponential and Modulus transformations that work well only on positive values are applied to the lagged-variance in the GJR process. Second, the models allow generalized Student-t, skewed Student-t, and skewed GED in the error distribution. Third, we employ the Adaptive Random Walk Metropolis (ARWM) method in [15] in the Markov Chain Monte Carlo (MCMC) algorithm to estimate the proposed models. Finally, the appropriateness of the proposed models is evaluated through the goodness-of-fit statistics provided by models when fitted to the real data. To of our knowledge, this study is considered a pioneer in modeling and estimating the EGARCH and GJR models with skewed and heavy-tailed distributions and non-linear transformations.

The rest of the paper is organized as follows. Section II presents our proposed models, including non-linear transformation families, asymmetric GARCH models, and skewed and heavy-tailed distributions. Section III describes the estimating in the Bayesian framework and explains the measurements used to evaluate parameters and sampling method. Section IV discusses an empirical study on the Swiss Market Index and performs comparison of the competing models based on the likelihood ratio test and deviance information criterion. The last section gives some conclusions.

II. Non-Linear GARCH Models

A. Non-linear Transformation Families

The non-linear transformation families are widely used in statistical study to make data as "Normal" as possible for the purposes of increasing the statistical validity. Also, in the mathematics modeling context, non-linear transformations can be used to generalize linear models to be more powerful non-linear models. Thus, by applying non-linear transformation families that nest linear specification, one can choose the most suitable functional form for a variable process. Recently, the non-linear transformation families were successfully applied to the return data by [16], [17] to provide better fitting than a basic model.

A class of non-linear transformations that are especially popular to generalize the standard linear form is power transformations. Firstly, a family of Tukey's conventional power transformations (called Simple Tukey (ST)) takes the following form [18]:

$$g^{ST}(x,\delta) = \begin{cases} x^{\delta}, & \delta > 0 \\ \log(x), & \delta = 0 \end{cases} \tag{2}$$

for $x > 0$. For dealing with negative observations, the ST transformation has an extension called Extended Tukey (ET) transformation, which is defined as in [17]:

$$g^{ET}(x,\delta) = \operatorname{sgn}(x)|x|^{\delta}, \delta > 0, \tag{3}$$

where the signum function, $\operatorname{sgn}(x)$, is defined as 1 for $x > 0$, 0 for $x = 0$, and -1 for $x < 0$. We include the case of $\delta = 0$ by defining

$$g^{ET}(x,\delta) = \operatorname{sgn}(x)\log(|x|), \delta = 0. \tag{4}$$

When $\delta = 1$, the standard linear form is produced as a special case.

Notice that the ST family has a discontinuity at $\delta = 0$. This problem was rectified by the Box–Cox (BC) transformation, which takes he following form:

$$g^{BC}(x,\delta) = \begin{cases} \dfrac{x^{\delta}-1}{\delta}, & \delta \neq 0 \\ \log(x), & \delta = 0. \end{cases} \tag{5}$$

To concern with real values, we could consider the Extended Box–Cox (EBC) transformation, with the addition of the $\delta = 0$ case, which is given by [17]:

$$g^{EBC}(x,\delta) = \begin{cases} \dfrac{\operatorname{sgn}(x)|x|^{\delta}-1}{\delta}, & \delta \neq 0 \\ \operatorname{sgn}(x)\log(x), & \delta = 0. \end{cases} \tag{6}$$

There have existed several popular alternative versions of the BC transformations which are defined for any real number (see [19]). For example, the Exponential (abbreviated as Exp) transformation takes the following form:

$$g^{Exp}(x,\delta) = \begin{cases} \dfrac{\exp\{\delta x\}-1}{\delta}, & \delta \neq 0 \\ x, & \delta = 0, \end{cases} \tag{7}$$

the form of the Modulus (abbreviated as Mod) transformation is given by:

$$g^{Mod}(x,\delta) = \begin{cases} \dfrac{\operatorname{sgn}(x)[(|x|+1)^{\delta}-1]}{\delta}, & \delta \neq 0 \\ \operatorname{sgn}(x)\log(|x|+1), & \delta = 0, \end{cases} \tag{8}$$

and a modification of the Mod transformation, called Yeo–Johnson (abbreviated as YJ) transformation, takes the following form:

$$g^{YJ}(x,\delta) = \begin{cases} \dfrac{(x+1)^{\delta}-1}{\delta}, & x \geq 0,\ \delta \neq 0 \\ \log(x+1), & x \geq 0,\ \delta = 0 \\ \dfrac{(1-x)^{2-\delta}-1}{\delta-2}, & x < 0,\ \delta \neq 2 \\ -\log(-x+1), & x < 0,\ \delta = 2 \end{cases} \tag{9}$$

If x is strictly positive, the YJ transformation is the same as the Mod transformation.

B. Non-linearities in Asymmetric GARCH Models

GARCH-type models are popular since this class considers the time-varying volatility in asset returns as an unobserved variable. This makes uncertainty about the specifications of volatility process whether linear or non-linear. Since the previous power transformations include the case of linear form, applying such transformations in the lagged-volatility process allows the unobserved variable to select the adoption between the linear and non-linear specifications. For these objectives this study propose a class of asymmetric GARCH(1,1) models which incorporates a family of power transformations in the lagged-volatility.

As the part of our model, the return of financial time series R_t at time t is assumed to be modelled by

$$R_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1), \quad t = 1, \ldots, T, \tag{10}$$

in which $\mathcal{N}$ denotes the Normal distribution. When return R_t is positive (negative), it describes good (bad) news in an asset market. The GJR model introduced in [4] is based on a modification of the conditional variance equation of the basic GARCH specification, which assumes that the asymmetric behaviour of current conditional variance depends on the sign of the past returns. Our general non-linear GJR(1,1) model, called the NL1-GJR(1,1) model, is given by

$$\sigma_t^2 = \omega + (\alpha + \gamma \mathbf{1}_{[R_{t-1}>0]}) R_{t-1}^2 + \beta g_{t-1}^{NL}, \tag{11}$$

for $t = 2, \ldots, T$, where $\mathbf{1}_{[A]} = 1$ if condition A holds and 0 otherwise, and $g_t^{NL} \equiv g^{NL}(\sigma_t^2, \delta)$ is a power transformation for σ_t^2. When $\gamma = 0$, we get the symmetric NL1-GARCH(1,1) model. This means negative return and positive return have the same effect on the conditional variance. If $\gamma > 0$ (or $\gamma < 0$), bad (good) news will increase the volatility more than good (bad) returns (according to [20]). The case of $\gamma > 0$ reflects a phenomenon commonly known as the "leverage effect", signifying that a negative return produces higher future volatility (variance) than a positive return of same magnitude (according to [21]).

Since the variance σ_t^2 in Eq. (11) is always positive for $t = 1, \ldots, T-1$, we employ three transformation families: ST, Exp, and Mod transformations. Notice that, in this case, the ST power transformation can be viewed as a simplified version of the BC transformation and also the Mod and YJ transformations have the same functional form. When the values of $\delta = 0$ in the case of Exp transformation and $\delta = 1$ in other cases, the standard GJR(1,1) model is produced. Notice that the power function g_t^{NL} is always positive for $\sigma_t^2 > 0$. Therefore, the GJR(1,1) model conditions on the parameters of the NL1-GJR(1,1) model are still valid and can be used, i.e. $\omega > 0$, $\alpha + \gamma \geq 0$, and $\beta > 0$ to ensure the non-negativity of the conditional variance and $\alpha + \beta + 0.5\gamma < 1$ to satisfy the covariance-stationary condition. The last condition was derived by [22] based on the following theorem.

Theorem 2.1: If $E|\epsilon_t|^{2m} < \infty$ and $E[\omega]^{2m} < \infty$, then the necessary and sufficient condition for the existence of the $2m$-th moment of the solution R_t is $E\left[\beta + \left(\alpha + \gamma \mathbf{1}_{[R_{t-1}>0]}\right)\right]^{2m} < 1$, where m is a positive integer

In contrast to the GARCH specification, [3] proposed the EGARCH (Exponential GARCH) model by taking the logarithm on the conditional variance so that it does not require non-negativity parameters constraint. The mathematical form of our general non-linear EGARCH(1,1) model, called the NL1-EGARCH(1,1), is described as:

$$h_t = \omega + \alpha \left| \frac{R_{t-1}}{\sigma_{t-1}} \right| + \beta g_{t-1}^{NL} + \gamma \frac{R_{t-1}}{\sigma_{t-1}}, \tag{12}$$

for $t = 2, \ldots, T$, where $h_t = \log\left(\sigma_t^2\right)$ and $g_t^{NL} \equiv g\left(h_t, \delta\right)$. Since h_t is real-valued variable, we can apply the ET, Exp, Mod, and YJ transformations. Notice that, in this case, the ET transformation can be viewed as a simplified version of the EBC transformation. The above model also enables the capture of the asymmetric behavior of volatility in asset return, which is represented by the parameter γ. If $\gamma = 0$, it indicates a perfect symmetric model, with asymmetry given by $\gamma \neq 0$. In this case, the existence of leverage effect is reflected by $\gamma < 0$.

C. Distributions Allowing the Skewness and Kurtosis

One of the flexible distributions which generalizes Student's t distribution and allows to capture both skewness and kurtosis (a measure of tail thickness) was derived in [10]. The Hansen's skewed Student's t (HST) distribution for a random variable y with expectation 0 and standard deviation $\sigma > 0$ can be defined by

$$\begin{aligned} \text{HST}(y|\sigma,\nu,\lambda) = bc \Bigg[1 \\ + \frac{1}{\nu-2}\left(\frac{by\sigma^{-1}+a}{1+\text{sgn}(by\sigma^{-1}+a)\lambda}\right)^2 \Bigg]^{-0.5(\nu+1)}. \end{aligned} \tag{13}$$

where

$$a = 4\lambda c \frac{\nu-2}{\nu-1}, b^2 = 1 + 3\lambda^2 - a^2, c = \frac{\Gamma(0.5(\nu+1))}{\sqrt{\pi(\nu-2)}\Gamma(0.5\nu)}.$$

Here, $2 < \nu < \infty$ is the degrees of freedom which controls the height and tails of the density and $-1 < \lambda < 1$ determines degree of skewness. A left (negative) skewed distribution is indicated by $\lambda \in (-1,0)$, whereas a right (positive) skewed distribution is indicated by $\lambda \in (0,1)$. So, if $\lambda = 0$, HST distribution reduces to the traditional Student's t distribution in which the smaller ν gives heavier tails.

A different version of the skewed Student's t distribution was proposed in [11]. For a random variable y with expectation 0 and standard deviation $\sigma > 0$, the Fernández–Steel's skewed Student's t (FST) distribution can be defined by

$$\begin{aligned} \text{FST}(y|\sigma,\nu,\lambda) = \frac{2}{\lambda+\lambda^{-1}} \frac{\Gamma(0.5(\nu+1))}{\Gamma(0.5\nu)} \frac{1}{\sqrt{\pi\sigma^2(\nu-2)}} \Bigg[1 \\ + \frac{1}{\sigma^2(\nu-2)} y^2 \left(\lambda^2 \mathbf{1}_{[y\leq 0]} + \lambda^{-2}\mathbf{1}_{[y>0]}\right) \Bigg]^{-0.5(\nu+1)}. \end{aligned} \tag{14}$$

Clearly, when $\lambda = 1$, the density simplifies to the traditional Student's t distribution. Negative skewness is generated by $\lambda \in (0,1)$, whereas positive skewness corresponds to $\lambda > 1$

Another class of flexible distributions, which covers both asymmetrics and tail-thickness, is the skewed generalized error distribution (SGED) suggested in [12]. In contrast to the two skewed Student's t distributions above, the SGED is a generalization of Skew Normal. The SGED probability density function for the random variable y with expectation 0 and standard deviation $\sigma > 0$ is given by

$$\mathrm{SGED}(y|\sigma,\kappa,\lambda) = \frac{C}{\sigma}\exp\left\{-\frac{|y+u\sigma|^{\kappa}}{[1+\mathrm{sgn}(y+u\sigma)\lambda]q^{\kappa}\sigma^{\kappa}}\right\} \tag{15}$$

where:

$$C = \frac{\kappa}{2q}\cdot\Gamma\left(\frac{1}{\kappa}\right)^{-1}, q = \Gamma\left(\frac{1}{\kappa}\right)^{0.5}\Gamma\left(\frac{3}{\kappa}\right)^{0.5}S(\lambda)^{-1},$$
$$S(\lambda) = \sqrt{1+3\lambda^2-4A^2\lambda^2}, u = \frac{2\lambda A}{S(\lambda)},$$
$$A = \Gamma\left(\frac{2}{\kappa}\right)\Gamma\left(\frac{1}{\kappa}\right)^{-0.5}\Gamma\left(\frac{3}{\kappa}\right)^{-0.5}.$$

Here, κ is the shape parameter with constraint $\kappa > 0$ and controls the height and tails of the density, λ is the skewness parameter with constraint $-1 < \lambda < 1$. When $0 < \kappa < 2$, the tails of the distribution become heavier than the Normal distribution; when $\kappa > 2$, the tails of the distribution become thinner than the Normal distribution. Meanwhile, when λ is more positive, the distribution is more skewed to the right and will be more skewed to the left when λ is more negative. So, for $\kappa = 2$, the SGED reduces to the Normal distribution when $\lambda = 0$ and to the Skew Normal distribution when $\lambda \neq 0$.

III. MCMC Estimation

The Maximum Likelihood Estimation (MLE) and Bayesian MCMC methods are popular statistical methods for estimating the parameters of a statistical model. In particular, the inference in our models is fully based on the Bayesian computational approach. We focus on MCMC algorithm that have been used and become a reference method for analysing computational complexity of model (see [23] for a survey of MCMC). Empirically, the Bayesian MCMC approach performs better than the MLE approach in accuracy (see, for example, [24], [25], [26]).

Bayesian approach relies on likelihood formulation, but in practice it often customarily takes the natural logarithm of the likelihood function. By taking the log (means the natural logarithm), it is not only greatly simplifies the complex mathematical expressions, but it also helps numerically since the product of many small probabilities can be resolved by summing the log probabilities. It could significantly reduce the computer numerical error.

A. Priors and Conditional Log-likelihood Function

In the Bayesian framework, it is further required to complete the specification of statistical model by specifying probability density (referred to as likelihood and denoted by $L(\mathrm{data}|\theta)$) and prior distribution (denoted by $p(\theta)$) for unknown parameter θ. Using Bayesian terminology, the estimated probability of parameter after observing the data is called a "posterior probability" (denoted by $p(\theta|\mathrm{data})$) and it is stated in the following Bayes' theorem [27].

Theorem 3.1: A posterior probability for the parameter θ given the observed data is equal to the conditional probability of data given θ, multiplied by the probability of θ without any given conditions and divided by the probability of data without any given conditions. In the proportional form, it is expressed as follows:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior},$$

where the symbol "$\propto$" means "proportional to".

In this study we choose the exponential prior distribution on the shape parameter k and the degrees of freedom parameter ν as in [28] and the truncated Normal priors on other parameters as in [29]:

$$p(x_1) \propto \exp(-0.5x_1^2/1000)\mathbf{1}_{[x_1>0]}, \tag{16}$$
$$p(x_2) \propto \exp(-0.5x_2^2/1000), \tag{17}$$
$$p(\lambda) \propto \exp(-0.5\lambda^2/1000)\mathbf{1}_{[|\lambda|<1]}, \tag{18}$$
$$p(\kappa) \propto \exp(-0.01\kappa)\mathbf{1}_{[\nu>0]}, \tag{19}$$
$$p(\nu) \propto \exp(-0.01(\nu-2))\mathbf{1}_{[\nu>2]}, \tag{20}$$

where $x_1 \in \{\omega,\alpha,\beta\}$ and $x_2 \in \{\gamma,\delta\}$.

Our purpose is to draw statistical inference based on a sequence of observed returns, $\mathbf{R} = \{R_1, R_2, \ldots, R_T\}$, with the sequence of associated conditional variances is $\sigma_1, \sigma_2, \ldots, \sigma_T$. First, let us consider the HST specification. We define

$$W_t = \left(bR_t\sigma_t^{-1}+a\right)/(1+I_W\lambda)$$
$$W\widetilde{W}_t = 1+W_t^2/(\nu-2),$$

where I_W takes the value of 1 if $bR_t\sigma_t^{-1} > 0$ and -1 otherwise. Under the HST specification, the conditional log-likelihood function of returns with the conditional variance σ_t^2 follows either NL1-EGARCH(1,1) or NL1-GJR(1,1) models can be expressed as

$$\begin{aligned}\mathscr{L}(\mathbf{R}|\theta_1) &= \log\left(L(\mathbf{R}|\theta_1)\right)\\ &= T[\log(b)+\log(c)]\\ &\quad -\frac{\nu+1}{2}\sum_{t=1}^{T}\log(W\widetilde{W}_t),\end{aligned} \tag{21}$$

where $\theta_1 = (\omega,\alpha,\beta,\gamma,\nu,\lambda,\delta)$.

Second, we consider the FST specification. Let us define $Z_t = 1+z_\lambda R_t^2\sigma_t^{-2}(\nu-2)^{-1}$ where z_λ takes the value of λ^2 if $R_t \leq 0$ and λ^{-2} otherwise. Under the FST specification, the conditional log-likelihood function of returns for the NL1-EGARCH(1,1) and NL1-GJR(1,1) models is given by

$$\begin{aligned}\mathscr{L}(\mathbf{R}|\theta_1) &= T\Bigg[\log(2)-\log(\lambda+\frac{1}{\lambda})\\ &\quad +\log\left(\Gamma\left(\frac{1}{2}(\nu+1)\right)\right)\\ &\quad -\log\left(\Gamma\left(\frac{1}{2}\nu\right)\right)-\frac{1}{2}\log(\pi(\nu-2))\Bigg]\\ &\quad -\sum_{t=1}^{T}\log\left(\sigma_t\right)-\frac{\nu+1}{2}\sum_{t=1}^{T}\log(Z_t).\end{aligned} \tag{22}$$

Finally, the SGED specification is considered. The conditional log-likelihood function of returns for the NL1-EGARCH(1,1) and NL1-GJR(1,1) models is given by

$$\mathscr{L}(\mathbf{R}|\theta_2) = T\log(C)-\sum_{t=1}^{T}\log(\sigma_t)-\sum_{t=1}^{T}\frac{|R_t\sigma_t^{-1}+u|^{\kappa}}{(1+s\lambda)^{\kappa}q^{\kappa}}, \tag{23}$$

where $\theta_2 = (\omega,\alpha,\beta,\gamma,\kappa,\lambda,\delta)$ and s takes the value of 1 if $R_t\sigma_t^{-1}+u > 0$ and -1 otherwise.

B. The ARWM Sampler

In MCMC simulation, there are two typically distinct steps. As a first step it simulates a Markov chain by using a sampler. The second step is then to use the Monte Carlo approximation to estimate the descriptive statistics measures (such as mean, standard deviation, Bayesian credible interval, and inefficiency factor) for the Markov chain. There are several samplers that have been proposed. Here we employ the ARWM method of [15]. This sampler was successfully applied to GARCH(1,1)-type models by [30], [31], [32].

The ARWM method is designed to improve the sampling efficiency of the random walk Metropolis, which is the simplest and one of the most common sampler types in practical use. For a parameter θ and at iteration $n \in \mathbb{N}$, schematically, the sampler involves the following ways:

(i) Suppose that at iteration $(n-1)$th, we have $\theta^{(n-1)}$ and step-size $\Delta^{(n-1)}$.

(ii) Draw a new value $\theta^* = \theta^{(n-1)} + \mathcal{N}(0, \Delta^{(n-1)})$.

(iii) Calculate the Metropolis ratio:

$$r\left(\theta^{(n-1)}, \theta^*\right) = \mathscr{L}(\mathbf{R}|\theta^*) + \log(p(\theta^*)) - \mathscr{L}(\mathbf{R}|\theta^{(n-1)}) - \log(p(\theta^{(n-1)})). \quad (24)$$

(iv) Draw $u \sim \mathcal{U}(0,1)$, in which $\mathcal{U}$ is the Uniform distribution. If $u < \exp\left\{r\left(\theta^{(n-1)}, \theta^*\right)\right\}$, we accept θ^* and set $\theta^{(n)} = \theta^*$. Otherwise, we set $\theta^{(n)} = \theta^{(n-1)}$.

(v) Update the step-size $\Delta^{(n)} \in [\Delta_{\min}, \Delta_{\max}]$ and compute:

$$\Delta = \max\left\{\Delta_{\min}, \Delta^{(n)} \frac{\frac{m(\theta^*)}{n} - \bar{\tau}}{n^\eta}\right\}, \quad (25)$$

where $m(\theta^*)$ denotes the number of acceptance for θ^*. If $\Delta < \Delta_{\max}$, we set $\Delta^{(n)} = \Delta$, but otherwise we set $\Delta^{(n)} = \Delta_{\max}$.

This study sets $\Delta_{\min} = 10^{-4}$, $\Delta_{\max} = 1000$, $\eta = 0,6$ as in [15] and chooses a targeted acceptance rate close to $\bar{\tau} = 0.44$ as in [33].

In the case of models with the HST and FST specifications, the MCMC scheme works in the following steps:

(0) Choose a starting value $\theta_1^{(0)}$, the number of iterations N, the burn-in-period J, and set $n = 1$.

(1)
  (i) Update a sequence of parameter values $\omega^{(n)}$, $\alpha^{(n)}$, $\beta^{(n)}$, $\gamma^{(n)}$, $\delta^{(n)}$, $\nu^{(n)}$, $\lambda^{(n)}$ by using the ARWM method for each parameter.
  (ii) If $n > J$, record the values of the parameters.
  (iii) If $n < N$, set $n = n + 1$ and then go to step 2.

(2) Using $N - J$ recorded draws, calculate various descriptive statistics measures to summary the Bayesian inference.

The similar way is also employed for the models with the SGED specification. The burn-in-period refers to a certain number of iterations at the beginning of the chain in order to minimize the effect of starting values. In our algorithm, N and J are specified as 15000 and 5000 iterations, respectively. Begin with the starting values: $\omega_0 = 0.01$, $\alpha_0 = 0.2$, $\beta_0 = 0.7$, $\gamma_0 = 0$, $\nu_0 = 10$, $\kappa_0 = 2$, dan $\gamma_0 = 0.5$, the 10000 remaining draws on each parameter are used to calculate the mean, integrated autocorrelation time (IACT), and 95% highest posterior density (HPD) interval.

C. HPD Interval and IACT

The HPD interval is one of the Bayesian credible intervals, where it is not the same as the equal-tailed interval when the distribution is not unimodal and symmetric. Following [34], the 95% HPD interval of $\{\theta^{(j)}\}_{j=1,\ldots,M}$, where $M = 10000$ (in our case), is constructed in the following steps:

(1) Compute $M_{cut} = [0.05 * M]$ and $M_{span} = M - M_{cut}$, where $[x]$ denotes the standard rounding function for x.
(2) Sort draws to obtain the ordered values: $\{\theta_k\}_{k=1}^M$, where $\theta_1 \le \theta_2 \le \cdots \le \theta_M$
(3) Compute $\boldsymbol{\theta}^* = \{\theta_k\}_{k=M_{span}}^M - \{\theta_k\}_{k=1}^{M_{cut}}$.
(4) Find a position k^* such that $\boldsymbol{\theta}^*(k^*)$ is a minimum value.
(5) The 95% HPD intervals:

$$\left(\theta_{k^*}, \theta_{k^*+M_{span}}\right).$$

Two desirable fundamental properties of a Markov chain are high efficiency and good chain mixing. These issues can be resolved with the help of convergence diagnostics. The convergence properties of the draw update can be assessed graphically using a trace plot. When the draw update stays in the same value for too many consecutives iterations, it indicates a poor mixing. Ideally, a chain have had well mixing or reached stationary if the trace plot of draws fluctuates around a mean value with a relatively constant variance. To measure the mixing/convergence speed of a Markov chain, this study calculates the autocorrelation of the recorded draws after discarding a burn-in period. We look at the IACT which is defined as:

$$\text{IACT} = \tau_{int} = \sum_{t=-\infty}^{\infty} \rho(t), \quad (26)$$

where $\rho(t)$ is the normalized autocorrelation function at time t. The value of τ_{int} is estimated using Sokal's adaptive truncated periodogram estimator by applying the Fast Fourier transform (FFT) to compute autocorrelations from the recorded draws. The procedure to estimate τ_{int} for the individual chain $\theta = \{\theta^{(i)}\}_{i=1,2,\ldots,M}$ was implemented in Matlab by [35] and described as follows:

(1) Calculate $x = FFT(\theta)$ and take the real and imaginary parts of x, namely $x_R = Re(x)$ and $x_I = Im(x)$, respectively.
(2) Compute the squared magnitude $x_P = x_R^2 + x_I^2$ and then set $x_P^{(1)} = 0$.
(3) Take $y = Re(FFT(x_P))$, and compute $\hat{\rho} = \frac{y}{y^{(1)}}$.
(4) Compute $\tau_k = -\frac{1}{3} + \sum_{i=1}^k \left(y^{(i)} - \frac{1}{6}\right)$ where $\tau_k < 0$.
(5) The value of IACT: $2\left(\tau_k + \frac{k-1}{6}\right)$.

The value of IACT gives a reasonable measure of the efficiency of a sampler since it can be interpreted as the number of iterations required to produce an independent draw from the sampler. This means that a sampler with a smaller IACT leads to lower autocorrelation and better mixing, requiring shorter chains per independent draw and generating a high efficiency.

IV. An Application: Swiss Market Index

This section addresses the testing for ARWM sampler and estimation of the parameters of the extended GARCH models. Each model is fitted to the Swiss Market Index (SMI) as one of the proxies of the global market.

A. The Data Set and Descriptive Statistics

In this study, we evaluate the performance of presented models by fitting the models to daily returns of the SMI for a period covering from January 2000 to December 2017. For this period, there are 4443 daily returns based on available trading days. The daily SMI data are extracted from the Oxford-Man Institute of Quantitative Finance (Realized Library Version: 0.2) which provides it for public. The SMI is traded on the SIX Swiss Exchange and contains the 20 stocks with the largest market values. This index have been used in other studies, such as in [36] which applied the GARCH(1,1) model to estimate the implied volatility process from the prices of options for the years 1992–1996 and in [37] which fitted the index for the years 2008–2009 to the AR(1)-GARCH(1,1) model.

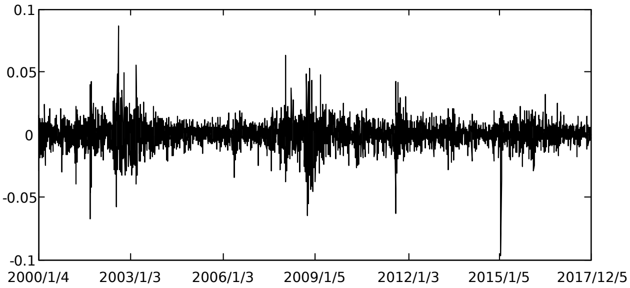

Fig. 1. Daily returns series of SMI

Fig. 1 displays the time series of returns for the SMI for the observation period. The daily returns were calculated by $R_t = 100 \times (\log(P_t) - \log(P_{t-1}))$, where P_t is the closing price of the SMI on day t. Notice that the average of the return series appears to be close to zero (-0.019), whereas the variance of the return series clearly changes over time. The return data have been checked for several standard statistics. We found that the SMI returns have a negative skewness of -0.3046 and a kurtosis of 11.50. These values indicate that the distribution of SMI returns has fatter tails than the Normal distribution and a longer left tail. This deviation from the Normal distribution is confirmed by the Jarque–Bera (JB) normality test with a statistic of 1343.5 ($p = 0.000$) and critical value of 5.98, at the 5% significance level. Hence, it provides a motivation to exploit several alternative non-Normal distributions such as HST, FST, and SGED specifications.

B. Testing the Efficiency of Estimator

It is essential to assess the efficiency of sampling method that involves a large number of parameters and a complicated likelihood function. Table I reports the IACT values for the parameters of distribution and transformation. Notice that the values for all the considered parameters are less than 100, except for the case of the Modulus transformation which is applied to the EGARCHhst and EGARCHfst models. The IACT values is then used to compute Effective Sample Size (ESS) which interprets the number of independent samples drawn through the Markov Chain. Since the ESS is calculated as the MC length (after the burn-in) divided by the IACT, here we have ESSs larger than 100 (excluding two cases) which can be considered good according to [38]. It indicates that the consecutive samples drawn by our sampler have lower correlation or higher efficiency, which is recommended to make reliable inferences on the interest parameter. Therefore, we say that our MCMC algorithm is well behaved.

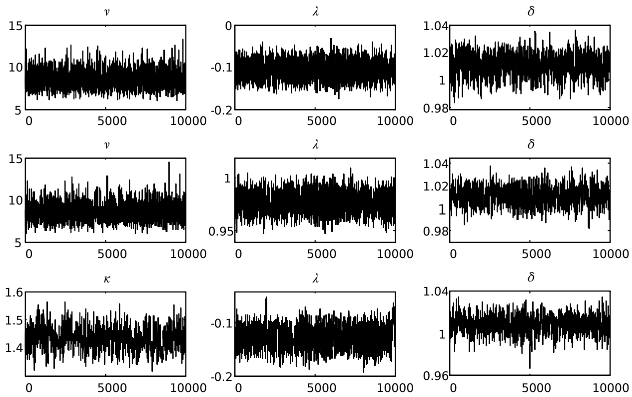

Fig. 2. Trace plots of posterior samples for parameter YJ and parameters of the HST (top), FST (middle), and SGED (bottom) distributions in the YJ1-EGARCH(1,1) models

Performance of a sampler can also be assessed by visual inspection of the trace plots for the posterior samples. For example, Fig. 2 displays the trace plots for the Markov chains of three parameters constructing the YJ1-EGARCH(1,1)

TABLE I
PERFORMANCE OF ARWM SAMPLER AS MEASURED BY IACT.

	GARCHt(1,1)		GARCHhst(1,1)			GARCHfst(1,1)			GARCHsged(1,1)		
Trans.	ν	δ	ν	λ	δ	ν	λ	δ	κ	λ	δ
NoT	10.88		6.39	7.38		7.88	6.24		29.43	10.27	
ST	7.29	41.27	6.50	4.83	96.59	7.12	6.22	56.06	38.30	11.63	45.98
Exp	6.97	14.90	7.59	5.30	14.12	6.60	5.84	15.28	40.43	5.71	20.36
Mod/YJ	6.99	31.93	6.12	6.91	47.32	5.69	8.11	23.37	40.82	6.12	68.58
	GJRt(1,1)		GJRhst(1,1)			GJRfst(1,1)			GJRsged(1,1)		
Trans.	ν	δ	ν	λ	δ	ν	λ	δ	κ	λ	δ
NoT	5.98		6.99	6.96		6.61	6.92		32.21	5.57	
ST	6.15	46.52	6.27	7.63	85.71	5.39	5.60	58.98	27.64	7.61	57.57
Exp	6.90	20.46	6.42	7.22	11.89	5.83	7.90	19.46	26.23	5.67	45.87
Mod/YJ	6.58	30.81	7.81	6.64	25.94	5.65	6.82	27.55	27.23	9.59	38.58
	EGARCHt(1,1)		EGARCHhst(1,1)			EGARCHfst(1,1)			EGARCHsged(1,1)		
Trans.	ν	δ	ν	λ	δ	ν	λ	δ	κ	λ	δ
NoT	5.39		6.75	6.17		5.97	7.16		48.22	6.23	
ST	7.85	25.85	7.75	6.29	13.54	7.26	6.75	7.66	32.20	5.41	11.43
Exp	9.93	13.89	6.35	5.51	15.29	5.86	7.61	15.90	31.56	8.33	24.84
Mod	6.16	76.48	5.95	5.95	114.30	6.66	6.78	122.91	34.93	5.26	62.37
YJ	6.51	26.61	6.46	5.27	15.60	6.23	10.50	15.78	28.86	6.74	28.74

models with HST, FST, and SGED distributions. The trace plots show that the chains seem to mix well and the parameter convergence was achieved. It indicates that the ARWM sampler quickly (effectively) moved through the parameter space.

C. Discussion on the Model Output

In this section, we verify the presence of non-linearity in the lagged-variance, asymmetry effects in the variance equation, and skewness and kurtosis in the asset returns distribution. The model fitting results are presented in Tables II–V. In these tables, for each model, we report the estimates of the posterior means, standard deviations, and 95% HPD intervals for the key parameters only.

1) On Non-linear Transformations: In the model specification, the non-linear transformation is replaced by a linear transformation making it a basic model. Regarding the GJR(1,1) models, the non-linearity assumption is fully guaranteed by all transformations in all distribution specifications since the δ parameter's HPD interval at the 5% significant level diverges from the linear assumption, referred to as $\delta = 0$ for the Exponential transformation and $\delta = 1$ for the others. We find that even 99% HPD interval of δ deviates from the linear assumption in the Student-t specification. Thus, the basic GJR(1,1) model is rejected and the data provide significant evidence of non-linear lagged variance.

When looking at the results from the EGARCH(1,1) models, the ET and Mod transformations demonstrate significant evidence against the basic EGARCH(1,1) model in all distribution specifications. Specifically, the δ parameter estimates for both transformations are significantly different from 1 at the 1% and 10% significant levels respectively in the Student-t and SGED cases. In the case of HST and FST distributions, the parameter of ET transformation has a higher significance level than the parameter of Mod transformation. The parameter δ is significant at a 5% level when the ET transformation is applied, meanwhile the parameter δ in applying the Mod transformation is significant at a 1% level which is preferable. This represents strong evidence that the Mod transformation significantly outperforms the others.

On the basis of the above results, the application of Tukey and Mod transformations in the lagged variance works well for the GJR(1,1) and EGARCH(1,1) processes in each distribution case and for the GARCH(1,1) model, but only in the HST specification. Meanwhile the Exp transformation works well to transform the lagged variance in the GJR(1,1) process when non-Normal distributions is applied and in the GARCH(1,1) process with the HST and SGED specifications. The YJ transformation, as well as the Exp transformation, does not perform well in the EGARCH(1,1) models. Therefore, we cannot make any conclusions based on the parameter significance that a non-linear transformation family is definitely a best choice for any models and distributions. However, the statistical significance of the δ parameter estimates provides sufficient evidence to reject the null hypothesis of a linear process for conditional variance and suggest the use of a non-linear transformation family for the lagged-variance.

2) On Asymmetric Effects: In order to capture the asymmetric effect of returns to the conditional variance, the GJR(1,1) and EGARCH(1,1) model were selected under three different non-Normal distributions. In terms of HPD interval, daily SMI return series provides significant evidence of the presence of asymmetric effect in both models and in all distribution cases. In particular, the parameter γ is statistically significant at the 1% level and positive in the GJR(1,1) model and negative in the EGARCH model. In other words, the leverage effect exists in both models.

We further observe in each asymmetric GARCH(1,1) model that when the non-linear transformation families are applied, the estimated asymmetric parameter across different transformations are close to each other in each of the specified distribution. The estimates for γ are also quite similar to those obtained in the basic asymmetric model. These results indicate that the presence of asymmetric effect is not affected by the transformation of the lagged-variance. In other words, the asymmetric GARCH(1,1) models do not depend on the non-linearity form of lagged-variance.

3) On Flexible Parametric Distributions: We first consider estimation results of the paramaters controling the tail heavyness in four types of distribution, including the shape parameter κ in SGED and the degree of freedom of the parameter ν in the others. All models with Student-t, HST, and FST distributions produce the degrees of freedom smaller than 30, indicating the existence of heavy-tails in the distribution of the SMI returns (according to [39]). Meanwhile, the estimates of κ are less than 2 and significantly different from 2 for each model and transformation, indicated by the 99% HPD interval excluding 2. This significance describes that the distribution of SMI returns has heavier tails than the Normal distribution. So, the four types of distribution is able to adequately capture the kurtosis of the return series of SMI which suggests necessity of heavy-tail feature in explaining leptokurtosis of the SMI data.

Second, we consider the presence of skewness in the SMI returns. Results show that all estimates of λ in the HST, FST, and SGED describe significantly negative skewness. In particularly, the parameter skewness of λ is statistically significant at the 1% level for the HST and SGED specifications since the HPD interval excludes zero and at the 5% level for the FST specification since the HPD interval excludes one. This significance indicates that the distribution of the SMI returns has asymmetric side as observed in our preliminary analysis. So, the three types of distribution is able to adequately capture the skewness of the return series of SMI which suggests necessity of skewness feature in explaining asymmetry of the SMI returns. According to [40], the negative skewness of the SMI return distribution means that the asset generates many small gains and a few large losses in the time period considered. Therefore, although the SMI may provide stable profits, technical investors or traders should be aware of optimal funds invested. This awareness avoids large losses, especially in times of crises like in 2008–2009.

Next, we jointly analize the skewness and heavy-tailedness parameters on flexible distributions. The above analysis shows that a statistically significant presence of both skewness and kurtosis in the HST, FST, and SGED specifications for the SMI return series is confirmed. From those significances, the departures from Normal, Skew Normal, and Student-t in the distribution of the SMI returns are rejected.

TABLE II
ESTIMATED POSTERIOR MEANS (PM), STANDARD DEVIATIONS (SD), AND 95% HPD INTERVAL FOR THE PARAMETERS OF THE GARCH(1,1) MODELS WITH STUDENT-t DISTRIBUTED ERRORS.

Trans.	Stats.	γ	ν	δ
GARCHt(1,1)				
NoT	PM (SD)		7.31 (0.82)	
	HPD		(5.87,8.96)	
ST	PM (SD)		7.37 (0.73)	0.9872 (0.0076)
	HPD		(5.99,8.80)	(0.9733,1.0052)
Exp	PM (SD)		7.40 (0.78)	-0.0111 (0.0063)
	HPD		(6.00,8.96)	$(-0.0224, 0.0022)$
Mod/YJ	PM (SD)		7.40 (0.76)	0.9849 (0.0133)
	HPD		(5.96,8.87)	(0.9577,1.0104)
GJRt(1,1)				
NoT	PM (SD)	0.1418* (0.0143)	7.96 (0.88)	
	HPD	(0.1151,0.1704)	(6.48,9.88)	
ST	PM (SD)	0.1429* (0.0158)	8.11 (0.85)	0.9822*(0.0065)
	HPD	(0.1136,0.1745)	(6.54,9.84)	(0.9695,0.9945)
Exp	PM (SD)	0.1472* (0.0177)	8.10 (0.89)	-0.0138*(0.0058)
	HPD	(0.1131,0.1796)	(6.37,9.76)	$(-0.0249, 0.0030)$
Mod/YJ	PM (SD)	0.1450* (0.0174)	8.11 (0.91)	0.9745*(0.0118)
	HPD	(0.1156,0.1818)	(6.34,9.72)	(0.9527,0.9990)
EGARCHt(1,1)				
NoT	PM (SD)	-0.1079* (0.0101)	8.25 (0.91)	
	HPD	$(-0.1274, -0.0885)$	(6.66,10.18)	
ET	PM (SD)	-0.1139* (0.0104)	8.17 (0.90)	0.9802*(0.0085)
	HPD	$(-0.1354, -0.0943)$	(6.37,9.86)	(0.9637,0.9964)
Exp	PM (SD)	-0.1082* (0.0103)	8.20 (0.90)	0.0066 (0.0046)
	HPD	$(-0.1272, -0.0882)$	(6.54,9.93)	$(-0.0025, 0.0152)$
Mod	PM (SD)	-0.1141* (0.0105)	8.19 (0.92)	0.9155*(0.0298)
	HPD	$(-0.1332, -0.0923)$	(6.50,10.04)	(0.8560,0.9665)
YJ	PM (SD)	-0.1093* (0.0101)	8.21 (0.91)	1.0104 (0.0082)
	HPD	$(-0.1279, -0.0896)$	(6.49,9.99)	(0.9932,1.0254)

Symbol * indicates statistically significant for HPD interval at the 1% level.

TABLE III
ESTIMATED POSTERIOR MEANS, STANDARD DEVIATIONS, AND 95% HPD INTERVAL FOR THE PARAMETERS OF THE GARCH(1,1) MODELS WITH HST DISTRIBUTED ERRORS.

Trans.	Stats.	γ	ν	λ	δ
GARCHhst(1,1)					
NoT	PM (SD)		7.50 (0.79)	-0.0991* (0.0194)	
	HPD		(5.97,8.99)	$(-0.1348, -0.0601)$	
ST	PM (SD)		7.62 (0.78)	-0.0984* (0.0195)	0.9887*** (0.0074)
	HPD		(6.14,9.12)	$(-0.1378, -0.0638)$	(0.9751,1.0035)
Exp	PM (SD)		7.61 (0.80)	-0.0974* (0.0199)	-0.0117** (0.0058)
	HPD		(6.16,9.23)	$(-0.1335, -0.0563)$	$(-0.0236, -0.0005)$
Mod/YJ	PM (SD)		7.68 (0.81)	-0.0973* (0.0198)	0.9824*** (0.0117)
	HPD		(6.25,9.28)	$(-0.1348, -0.0568)$	(0.9605,1.0069)
GJRhst(1,1)					
NoT	PM (SD)	0.1367* (0.0188)	8.21 (0.94)	-0.1030* (0.0204)	
	HPD	(0.0988,0.1705)	(6.56,10.27)	$(-0.1446, -0.0637)$	
ST	PM (SD)	0.1446* (0.0176)	8.26 (0.91)	-0.1034* (0.0211)	0.9841** (0.0073)
	HPD	(0.1088,0.1774)	(6.50,9.97)	$(-0.1422, -0.0617)$	(0.9696,0.9982)
Exp	PM (SD)	0.1430* (0.0177)	8.37 (0.98)	-0.1055* (0.0203)	-0.0155** (0.0058)
	HPD	(0.1036,0.1772)	(6.58,10.27)	$(-0.1457, -0.0677)$	$(-0.0264, -0.0048)$
Mod/YJ	PM (SD)	0.1417* (0.0162)	8.38 (0.97)	-0.1065* (0.0206)	0.9724** (0.0112)
	HPD	(0.1080,0.1701)	(6.60,10.25)	$(-0.1485, -0.0659)$	(0.9508,0.9943)
EGARCHhst(1,1)					
NoT	PM (SD)	-0.1062* (0.0100)	8.59 (0.99)	-0.1068* (0.0199)	
	HPD	$(-0.1257, -0.0869)$	(6.80,10.60)	$(-0.1454, -0.0696)$	
ET	PM (SD)	-0.1137* (0.0109)	8.42 (0.97)	-0.1077* (0.0199)	0.9796** (0.0093)
	HPD	$(-0.1353, -0.0942)$	(6.43,10.29)	$(-0.1471, -0.0689)$	(0.9602,0.9980)
Exp	PM (SD)	-0.1077* (0.0096)	8.55 (1.02)	-0.1066* (0.0200)	0.0075 (0.0046)
	HPD	$(-0.1267, -0.0893)$	(6.88,10.66)	$(-0.1448, -0.0669)$	$(-0.0017, 0.0161)$
Mod	PM (SD)	-0.1128* (0.0102)	8.52 (1.00)	-0.1077* (0.0202)	0.9285* (0.0315)
	HPD	$-0.1321, -0.0933$	(6.81,10.60)	$(-0.1442, -0.0680)$	(0.8672,0.9861)
YJ	PM (SD)	-0.1092* (0.0101)	8.61 (1.03)	-0.1078* (0.0204)	1.0107 (0.0078)
	HPD	$(-0.1279, -0.0892)$	(6.82,10.70)	$(-0.1474, -0.0665)$	(0.9959,1.0266)

Symbols *, **, and *** indicate statistically significant for HPD interval at the 1%, 5%, and 10% levels, respectively.

Furthermore, the estimation results for the distribution parameters in the models with non-linear transformations are close to those in the basic models. This indicates that the estimates of distribution parameters are not affected by transformations.

TABLE IV
ESTIMATED POSTERIOR MEANS, STANDARD DEVIATIONS, AND 95% HPD INTERVAL FOR THE PARAMETERS OF THE GARCH(1,1) MODELS WITH FST DISTRIBUTED ERRORS.

GARCHfst(1,1)					
Trans.	Stats.	γ	ν	λ	δ
NoT	PM (SD)		7.48 (0.82)	0.9772** (0.0096)	
	HPD		(6.00,9.19)	(0.9577,0.9949)	
ST	PM (SD)		7.55 (0.79)	0.9774** (0.0092)	0.9895 (0.0094)
	HPD		(6.19,9.09)	(0.9591,0.9946)	(0.9724,1.0089)
Exp	PM (SD)		7.65 (0.83)	0.9773** (0.0096)	−0.0118 (0.0066)
	HPD		(6.09,9.25)	(0.9570,0.9952)	(−0.0240, 0.0016)
Mod/YJ	PM (SD)		7.66 (0.83)	0.9777** (0.0094)	0.9804 (0.0116)
	HPD		(6.18,9.27)	(0.9572,0.9949)	(0.9564,1.0027)
GJRfst(1,1)					
NoT	PM (SD)	0.1426* (0.0184)	8.19 (0.91)	0.9768** (0.0093)	
	HPD	(0.1056,0.1786)	(6.55,10.02)	(0.9595,0.9957)	
ST	PM (SD)	0.1423* (0.0185)	8.35 (0.94)	0.9772** (0.0097)	0.9830** (0.0070)
	HPD	(0.1066,0.1782)	(6.64,10.18)	(0.9590,0.9967)	(0.9704,0.9984)
Exp	PM (SD)	0.1403* (0.0150)	8.38 (0.99)	0.9769** (0.0095)	−0.0154** (0.0058)
	HPD	(0.1135,0.1717)	(6.72,10.49)	(0.9590,0.9950)	(−0.0270, −0.0041)
Mod/YJ	PM (SD)	0.1444* (0.0153)	8.28 (0.96)	0.9773** (0.0097)	0.9726** (0.0108)
	HPD	(0.1177,0.1773)	(6.54,10.05)	(0.9575,0.9961)	(0.9529,0.9938)
EGARCHfst(1,1)					
NoT	PM (SD)	−0.1083* (0.0100)	8.49 (0.97)	0.9763** (0.0095)	
	HPD	(−0.1272, −0.0882)	(6.65,10.36)	(0.9583,0.9951)	
ET	PM (SD)	−0.1133* (0.0103)	8.44 (0.96)	0.9761** (0.0094)	0.9811** (0.0085)
	HPD	(−0.1335, −0.0934)	(6.64,10.25)	(0.9587,0.9944)	(0.9634,0.9962)
Exp	PM (SD)	−0.1089* (0.0100)	8.54 (0.96)	0.9763** (0.0099)	0.0070 (0.0051)
	HPD	(−0.1292, −0.0897)	(6.84,10.61)	(0.9572,0.9960)	(−0.0026, 0.0173)
Mod	PM (SD)	−0.1140* (0.0104)	8.43 (0.97)	0.9771** (0.0094)	0.9201* (0.0304)
	HPD	(−0.1333, −0.0916)	(6.63,10.36)	(0.9572,0.9942)	(0.8615,0.9762)
YJ	PM (SD)	−0.1085* (0.0103)	8.43 (0.95)	0.9771** (0.0095)	1.0105 (0.0077)
	HPD	(−0.1288, −0.0885)	(6.64,10.27)	(0.9599,0.9966)	(0.9958,1.0254)

Symbols *, **, and *** indicate statistically significant for HPD interval at the 1%, 5%, and 10% levels, respectively.

TABLE V
ESTIMATED POSTERIOR MEANS, STANDARD DEVIATIONS, AND 95% HPD INTERVAL FOR THE PARAMETERS OF THE GARCH(1,1) MODELS WITH SGED DISTRIBUTED ERRORS.

GARCHsged(1,1)					
Trans.	Stats.	γ	κ	λ	δ
NoT	PM (SD)		1.408* (0.037)	−0.1171* (0.0186)	
	HPD		(1.341,1.848)	(−0.1508, −0.0798)	
ST	PM (SD)		1.406* (0.036)	−0.1186* (0.0176)	0.9874 (0.0093)
	HPD		(1.332,1.469)	(−0.151, −0.082)	(0.9709,1.0073)
Exp	PM (SD)		1.400* (0.034)	−0.1184* (0.0171)	−0.0188** (0.0089)
	HPD		(1.336,1.470)	(−0.149, −0.083)	(−0.0348, −0.0011)
Mod/YJ	PM (SD)		1.402* (0.037)	−0.1192* (0.0179)	0.9688 (0.0191)
	HPD		(1.334,1.474)	(−0.1544, −0.0844)	(0.9320,1.0062)
GJRsged(1,1)					
NoT	PM (SD)	0.1219* (0.0142)	1.423* (0.035)	−0.1255* (0.0181)	
	HPD	(0.0970,0.1506)	(1.354,1.490)	(−0.1611, −0.0908)	
ST	PM (SD)	0.1292* (0.0159)	1.432* (0.037)	−0.1247* (0.0182)	0.9809** (0.0078)
	HPD	(0.0963,0.1601)	(1.358,1.501)	(−0.1610, −0.0919)	(0.9653,0.9966)
Exp	PM (SD)	0.1279* (0.0169)	1.423* (0.036)	−0.1247* (0.0184)	−0.0190** (0.0082)
	HPD	(0.0972,0.1598)	(1.358,1.493)	(−0.1582, −0.0853)	(−0.0356, −0.0038)
Mod/YJ	PM (SD)	0.1261* (0.0174)	1.426* (0.036)	−0.1247* (0.0191)	0.9643** (0.0141)
	HPD	(0.0944,0.1594)	(1.356,1.497)	(−0.1669, −0.0919)	(0.9374,0.9918)
EGARCHsged(1,1)					
NoT	PM (SD)	−0.1014* (0.0091)	1.431* (0.036)	−0.1281* (0.0188)	
	HPD	(−0.1204, −0.0818)	(1.355,1.504)	(−0.1678, −0.0945)	
ET	PM (SD)	−0.1044* (0.0100)	1.430* (0.035)	−0.1291* (0.0187)	0.9873*** (0.0077)
	HPD	(−0.1229, −0.0847)	(1.362,1.496)	(−0.1639, −0.0912)	(0.9725,1.0015)
Exp	PM (SD)	−0.1025* (0.0100)	1.433* (0.038)	−0.1271* (0.0179)	0.0050 (0.0055)
	HPD	(−0.1203, −0.0838)	(1.365,1.501)	(−0.1649, −0.0921)	(−0.0059, 0.0149)
Mod	PM (SD)	−0.1048* (0.0097)	1.427* (0.039)	−0.1283* (0.0188)	0.9568*** (0.0218)
	HPD	(−0.1226, −0.0854)	(1.354,1.501)	(−0.1624, −0.0891)	(0.9167,1.0012)
YJ	PM (SD)	−0.1024* (0.0105)	1.429* (0.038)	−0.1275* (0.0191)	1.0085 (0.0083)
	HPD	(−0.1196, −0.0836)	(1.367,1.499)	(−0.1617, −0.0909)	(0.9919,1.0237)

Symbols *, **, and *** indicate statistically significant for HPD interval at the 1%, 5%, and 10% levels, respectively.

D. Testing the Adequacy of the Model

When two models are nested—that is, the simpler model can be derived from the full (more complex) model by placing additional constraints on its parameters [41]—then their goodness-of-fit to describe the observed data can be assessed by using a Log-likelihood Ratio Test (LRT). In our case, the LRT statistic can be used to choose between the basic GARCH-type model (as a simple model M_0, including GARCH, GJR, EGARCH models without transformation) and an NL1-GARCH-type model (as a full model M_1, including GARCH, GJR, EGARCH models with non-linear transformations). The comparison of the fitting performance of these two same type models is performed by conducting the following hypothesis:

$$\begin{aligned} H_0 &: \delta = 0_{\text{Exp}},\ 1_{\text{ST,Mod,YJ}} \text{ (basic model)} \\ H_1 &: \delta \neq 0_{\text{Exp}},\ 1_{\text{ST,Mod,YJ}} \text{ (NL1 model)} \end{aligned}$$

The LRT statistic is calculated as twice the difference in log-likelihoods:

$$\text{LRT} = 2\left(\mathscr{L}(M_1, \hat{\theta}) - \mathscr{L}(M_0, \hat{\theta})\right),$$

where $\hat{\theta}$ is the average of estimated values. The following theorem offers an asymptotic distribution of the LRT statistic [42].

Theorem 4.1: Given some suitable regularity conditions, the distribution of the LRT statistic is approximately $\chi^2_{(d)}$-distributed if sufficient data is acquired, where d is the difference between the number of free parameters.

Table VI contains the results for the log-likelihood values as well as LRT statistics comparing basic and NL1 models. Since the difference in the number of free parameters between the two nested models is 1, those statistics are then compared with the 99%, 95%, and 90% critical values, namely $\chi^2_{(1)} = 6.64, 3.84, 2.71$ respectively. We find that in the case of ST transformation, there is a sufficient evidence to reject the null hypothesis. The exception is for the GARCH models with FST and SGED specifications which are accepted at all canonical significance levels (1%, 5%, and 10%). Applying the Exp transformation, the null hypothesis is rejected in the case of GARCH and GJR models and accepted in the case of EGARCH model. In the case of Mod transformation, the rejection of the null hypothesis is provided by all cases, except for the GARCH with FST distribution. Finally, the YJ transformation is rejected to be applied in the EGARCH model with any distributions. The results conclude that non-linear transformations have the potential to outperform the basic GARCH-type models.

Notice that the results of the superiority of families of transformations confirm the previous findings related to the significance of the transformation parameter. There are some cases in which the non-linear transformation parameter is not statistically significant at all canonical significance levels, but the model with non-linear transformation is superior than the basic model. For example, all NL1-GARCHt models statistically outperform the basic GARCHt model even though the parameter δ is not statistically significant in terms of HPD intervals at all canonical significance levels. In this case, its posterior estimates are primarily less than a value that corresponds to no transformation. In addition, the ST and Mod transformations seems to have similar strengths, while the strength of the Exp transformation is similar to those of the YJ transformation. This is indicated by the similar LRT statistics.

Furthermore, since all competing models are not nested, the DIC (Deviance Information Criterion) values were calculated to compare the performance of all models and find a best fit model. The DIC statistic for each estimated model with parameter θ is calculated as follows [43]:

$$\text{DIC} = -4E_{\theta|\mathbf{R}}[\mathscr{L}(\mathbf{R}|\theta)] + 2\mathscr{L}(\mathbf{R}|\tilde{\theta}), \tag{27}$$

TABLE VI
ESTIMATED LOG-LIKELIHOOD, LRT (GARCH VS NL1-GARCH), AND DIC FOR EACH MODEL.

Trans.	$-\mathscr{L}$	LRT	DIC	$-\mathscr{L}$	LRT	DIC	$-\mathscr{L}$	LRT	DIC
	GARCHt(1,1)			GJRt(1,1)			EGARCHt(1,1)		
NoT	5278.04	-	10561.03_4	5233.09	-	10470.74_4	5221.82	-	10448.70_5
ST	5276.37	3.34^{***}	10557.02_2	5230.48	5.22^{**}	10465.85_2	5219.79	4.06^{**}	10445.59_2
Exp	5275.96	4.16^{**}	10556.99_1	5229.72	6.74^{*}	10465.60_1	5221.21	1.22	10448.15_3
Mod	5276.58	2.92^{***}	10559.17_3	5230.77	4.64^{**}	10467.63_3	5219.28	5.08^{**}	10444.36_1
YJ							5221.30	1.04	10448.69_4
	GARCHhst(1,1)			GJRhst(1,1)			EGARCHhst(1,1)		
NoT	5266.63	-	10537.41_4	5221.02	-	10448.02_4	5209.18	-	10424.41_3
ST	5265.12	3.02^{***}	10535.62_2	5218.34	5.36^{**}	10443.14_3	5207.00	4.36^{**}	10421.23_2
Exp	5264.61	4.04^{**}	10535.09_1	5216.95	8.14^{*}	10440.59_1	5208.86	0.64	10425.15_4
Mod	5265.05	3.16^{***}	10535.76_3	5218.01	6.02^{**}	10442.54_2	5206.32	5.27^{**}	10419.06_1
YJ							5208.86	0.64	10425.15_5
	GARCHfst(1,1)			GJRfst(1,1)			EGARCHfst(1,1)		
NoT	5275.30	-	10555.65_4	5230.98	-	10468.05_4	5219.79	-	10445.97_4
ST	5274.50	1.60	10554.79_3	5228.36	5.24^{**}	10463.42_3	5217.47	4.64^{**}	10441.66_2
Exp	5273.71	3.18^{***}	10553.44_1	5226.79	8.38^{*}	10459.69_1	5219.23	1.12	10445.66_3
Mod	5274.16	2.28	10553.83_2	5228.11	5.74^{**}	10462.94_2	5217.07	5.44^{**}	10440.76_1
YJ							5219.32	0.94	10446.11_5
	GARCHsged(1,1)			GJRsged(1,1)			EGARCHsged(1,1)		
NoT	5279.62	-	10564.82_4	5240.57	-	10486.82_4	5229.76	-	10465.84_5
ST	5278.56	2.12	10562.83_3	5237.74	5.66^{**}	10482.63_3	5228.40	2.72^{**}	10463.71_2
Exp	5277.52	4.20^{**}	10560.32_1	5237.05	7.04^{*}	10481.87_1	5229.34	0.84	10465.17_3
Mod	5278.21	2.82^{***}	10562.19_2	5237.55	6.04^{**}	10482.54_2	5228.09	3.34^{**}	10462.58_1
YJ							5229.19	1.14	10465.21_4

Symbols *, **, and *** indicate statistically significant at the 1%, 5%, and 10% levels, respectively. Meanwhile, subscript denotes the rank of the model.

where $\tilde{\theta}$ is an estimate of θ maximizing the posterior. Given a set of candidate models for the data, the preferred model is the one that produces the smallest DIC.

Table VI also reports the results for the DIC values and the corresponding ranking in a set of models with same type and distribution. According to this grouping, all NL1-GARCH-type models provides a better fit than the basic GARCH-type models, except for the EGARCH models with HST and FST distributions. The results indicate that transforming the lagged-variance in the GARCH and GJR models by using exponential transformation can provide the best fit for any distributions. Meanwhile, the Modulus transformation yields the best performance when it is applied to the EGARCH model with any distributions.

In addition, the models including asymmetry effect in the variance process perform better than GARCH model in each distribution and non-linear transformation, where the EGARCH model is the best one. On comparing four distributions, the models with HST distribution is found to give the best result, followed by the models with FST, Student-t, and SGED specifications. Notice that the models with SGED distribution is outperformed by the models with Student-t distribution although the SGED specification is able to capture both skewness and heavy-tailedness of the return data. In general, this concludes that the best fit volatility model is provided by the Mod1-EGARCHhst model.

V. Conclusions

This study applied four non-linear transformation families, including Tukey, Exponential, Modulus, and Yeo–Johnson, to the lagged-variance in three different GARCH-type models: GARCH(1,1), GJR(1,1), and EGARCH(1,1) models. Here we incorporated four different returns error distributions: Student-t, Hansen's skewed Student-t, Fernández–Steel's skewed Student-t, and skewed generalized error distribution into the investigated models. The Bayesian inference of the models was performed by employing the ARWM method in MCMC scheme. The model and method were applied to the Swiss Market Index over the daily period from January 2000 to December 2017 for a total of 4443 returns.

Our findings can be summarised as follows. First, in terms of IACT as well as ESS, we found the ARWM method to be very efficient statistically to sample our proposed models which have complicated likelihood functions. Second, our findings suggest that the GARCH-type models might have performed better when a non-linear transformation family is applied to the lagged-variance. In particular, the GARCH and GJR models based on the exponential transformation and EGARCH model based on the Modulus transformation have statistically significant ability to fit the data in any considered distributions. Furthermore, since the DIC favored the HST distribution, this study generally concludes that the best fitted model for the SMI data is the EGARCH model based on the Modulus transformation with the return error following the Hansen's skewed Student-t distribution. This means that the model might be suitable for modeling daily returns in a market, such as SMI, which is mostly filled by companies managed by foreign elites [44].

References

[1] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[2] P. H. Franses and D. van Dijk, *Nonlinear time series models in empirical finance*, Cambridge University Press, 2003.

[3] D. B. Nelson, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991.

[4] L. Glosten, R. Jagannathan, and D. E. Runkle, "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, vol. 48, no. 5, pp. 1779–1801, 1993.

[5] M. L. Higgins and A. K. Bera, "A Class of Nonlinear ARCH Models," *International Economic Review*, vol. 33, no. 1, pp. 137–158, 1992.

[6] Z. Ding, C. W. J. Granger, and R. F. Engle, "A Long Memory Property of Stock Market Returns and A New Model," *Journal of Empirical Finance*, vol. 1, no. 1, pp. 83–106, 1993.

[7] D. B. Nugroho, B. Susanto, K. N. P. Prasetia, and R. Rorimpandey, "Modeling of Returns Volatility Using GARCH(1,1) Model under Tukey Transformation," *Jurnal Akuntansi dan Keuangan*, vol. 21, no. 1, pp. 12–20, 2019.

[8] D. B. Nugroho and T. Morimoto, "Realized Non-linear Stochastic Volatility Models with Asymmetric Effects and Generalized Student's t-Distributions," *Journal of The Japan Statistical Society*, vol. 44, no. 1, pp. 83–118, 2014.

[9] L. Feng and Y. Shi, "A Simulation Study on the Distributions of Disturbances in the GARCH Model," *Cogent Economics & Finance*, vol. 5, no. 1,1355503, 2017.

[10] B. E. Hansen, "Autoregressive Conditional Density Estimation," *International Economic Review*, vol. 35, no. 3, pp. 705–730, 1984.

[11] C. Fernández and M. F. J. Steel, "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 359–371, 1998.

[12] T. G. Bali and P. Theodossiou, "Risk Measurement Performance of Alternative Distribution Functions," *The Journal of Risk and Insurance*, vol. 75, no. 2, pp. 411–437, 2008.

[13] R. S. Ehlers, "Computational Tools for Comparing Asymmetric GARCH Models via Bayes Factors," *Mathematics and Computers in Simulation*, vol. 82, no. 5, pp. 858–867, 2012.

[14] A. Azzalini, "Skew-Normal Distribution," in: M. Lovric (Ed.), *International Encyclopedia of Statistical Science*, Springer, pp. 1342–1344, 2011

[15] Y. F. Atchade and J. S. Rosenthal, "On Adaptive Markov Chain Monte Carlo Algorithms," *Bernoulli*, vol. 11, no. 5, pp. 815–828, 2005.

[16] D. B. Nugroho, "GARCH(1,1) model with the Yeo–Johnson transformed returns," *Journal of Physics: Conference Series*, vol. 1320, no. 1, p. 012013, 2019.

[17] D. B. Nugroho, T. Mahatma, and Y. Pratomo, "GARCH Models under Power Transformed Returns: Empirical Evidence from International Stock Indices," *Austrian Journal of Statistics*, vol. 50, no. 4, pp. 1–18, 2021.

[18] T. Daimon, "Box–Cox Tranformation," in: M. Lovric (Eds.), *International Encyclopedia of Statistical Science*, Springer, Berlin, Heidelberg, pp. 176–178, 2011.

[19] E. Charitidou, D. Fouskakis, I. Ntzoufras, "Bayesian Transformation Family Selection: Moving Toward A Transformed Gaussian Universe," *Canadian Journal of Statistics*, vol. 43, no. 4, pp. 600–623, 2015.

[20] A. Dutta, "Model Volatility: Symmetric or Asymmetric GARCH Models," *Journal of Statistics: Advances in Theory and Applications*, vol. 12, pp. 99–108, 2014.

[21] M. A. Carnero and A. Perez, "Outliers and Misleading Leverage Effect in Asymmetric GARCH-type Models," *Studies in Nonlinear Dynamics & Econometrics*, vol. 25, no. 1, 20180073, 2019.

[22] S. Ling and M. McAleer, "Stationarity and the existence of moments of a family of GARCH processes," *Journal of Econometrics*, vol. 106, no. 1, pp. 109–117, 2002.

[23] C. Robert and G. Casella, "A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data," *Statistical Science* vol. 26, no. 1, pp. 102–115, 2011.

[24] T. Miazhynskaia and G. Dorffner, "A Comparison of Bayesian Model Selection Based on MCMC with An Application to GARCH-type Models," *Statistical Papers* vol. 47, no. 4, pp. 525–549, 2006.

[25] H. Wang, C. Wang, Y. Wang, X. Gao, and C. Yu, "Bayesian Forecasting and Uncertainty Quantifying of Stream Flows Using Metropolis–Hastings Markov Chain Monte Carlo Algorithm," *Journal of Hydrology* vol. 549, pp. 476–483, 2017.

[26] L. Rabie and J. Li, "E-Bayesian Estimation for Burr-X Distribution Based on Type-I Hybrid Censoring Scheme," *IAENG International Journal of Applied Mathematics* vol. 48, no. 3, pp. 244–250, 2018.

[27] J. N. Rouder and R. D. Morey, "Teaching Bayes' Theorem: Strength of Evidence as Predictive Accuracy," *American Statistician*, vol. 73, no. 2, pp. 186–190, 2019.

[28] P. Deschamps, "A Flexible Prior Distribution for Markov Switching Autoregressions with Student-t Errors," *Journal of Econometrics*, vol. 133, no. 1, pp. 153–190, 2006.

[29] D. Ardia and L. F. Hoogerheide, "Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations," *The R Journal*, vol. 2, no. 2, pp. 41–47, 2010.
[30] D. B. Nugroho and B. Susanto, "Volatility Modeling for IDR Exchange Rate Through APARCH Model with Student-t Distribution," *AIP Conference Proceedings ICRIEMS*, Yogyakarta, Indonesia, p. 040005, 2017.
[31] D. B. Nugroho, "Comparative Analysis of Three MCMC Methods for Estimating GARCH Models," *IOP Conference Series: Materials Science and Engineering ICEAT*, Mataram, Indonesia, p. 012061, 2018.
[32] D. B. Nugroho, D. Kurniawati, L. P. Panjaitan, Z. Kholil, B. Susanto, and L. R. Sasongko, "Empirical Performance of GARCH, GARCH-m, GJR- GARCH and log-GARCH Models for Returns Volatility," *Journal of Physics: Conference Series*, vol. 1307, no. 1, p. 012003, 2019.
[33] G. O. Roberts and J. S. Rosenthal, "Examples of Adaptive MCMC," *Journal of Computational and Graphical Statistics*, vol. 18, no. 2, pp. 349–367, 2009.
[34] H. Le, U. Pham, P. Nguyen, and T. B. Pham, "Improvement on Monte Carlo Estimation of HPD Intervals," *Communications in Statistics - Simulation and Computation*, vol. 49, no. 8, pp. 2164–2180, 2020.
[35] T. Lampart and I. Sbalzarini, "Implementation and Performance Comparison of an Ensemble Sampler with Affine Invariance,", *Technical report of the MOSAIC group, Institute of Theoretical Computer Science*, 2012.
[36] M. Sabbatini and O. Linton, "A GARCH Model of the Implied Volatility of the Swiss Market Index from Option Prices," *International Journal of Forecasting*, vol. 14, no. 2, pp. 199–213, 1998.
[37] H. Joe, *Dependence Modeling with Copulas*, CRC Press, 2015
[38] G. H. Givens and J. A. Hoeting, '*Computational Statistics*, Second Edition, John Wiley & Sons, 2013.
[39] M. Blangiardo and M. Cameletti, *Spatial and Spatio-temporal Bayesian Models with R-INLA*, John Wiley & Sons, 2015.
[40] S. Jaggia and A. Kelly-Hawke, "Modelling Skewness and Elongation in Financial Returns: The Case of Exchange-traded Funds," *Applied Financial Economics*, vol. 19, no. 6, pp. 1305–1316, 2009.
[41] T. Z. Keith and M. R. Reynolds, "Using Confirmatory Factor Analysis to Aid in Understanding the Constructs Measured by Intelligence Tests," in: D. P. Flanagan and P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, Fourth Edition, The Guilford Press, pp. 853–900, 2018.
[42] Q. Wu and P. Vos, "Inference and Prediction," in: V. N. Gudivada and C. R. Rao (Eds.), *Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications*, North Holland, pp. 131–192, 2018.
[43] C. M. Pooley and G. Marion, "Bayesian Model Evidence as A Practical Alternative to Deviance Information Criterion," *Royal Society Open Science*, vol. 5, no. 3, 171519, 2018.
[44] I. Fortuna, *Le recrutement des élites dirigeantes de nationalité étrangère: Le cas des entreprises du Swiss Market Index*, PhD dissertation, Université de Lausanne, Faculté des sciences sociales et politiques, 2014.