Model risk measures: A review and new proposals on risk forecasting

Fernanda Maria Müller[a,*], Marcelo Brutti Righi[a]

[a]*Business School, Federal University of Rio Grande do Sul, Washington Luiz, 855, Porto Alegre, Brazil, zip 90010-460*

Abstract

In financial decisions, model risk has been recognized as an important source of uncertainty. The revision of the Basel II suggests that financial institutions quantify and manage their model risk. Focusing on risk forecasting literature, we identify two main approaches to quantify model risk: the worst case and loss function. The first approach includes measures that are applied to a set of forecasts and have a similar structure to deviation measures. On the other hand, for the second approach, monetary risk measures are employed under a loss or error function, from some forecasting procedure. Moreover, based on the untapped features of model risk for both approaches we suggest new proposals, which include measures to quantify upside and downside model risk, and average costs associated with risk overestimation and underestimation. We also conduct an empirical assessment of model risk measures using Value at Risk (VaR) and Expected Shortfall (ES) forecasting. Results indicate that model risk estimates change according to measures, sample and functional (VaR and ES). We also conclude that a model with good performance to risk forecasting does not indicate this model has lower model risk. Furthermore, we highlight insights into future research directions regarding this topic.

Keywords: Model risk; Risk forecasting; Model risk measures; Robust finance; Review.

1. Introduction

Financial decisions are based on statistical model outputs for some variables of interest, such as return (linked to mean and median), and risk (linked to variance, Value at Risk (VaR) and Expected Shortfall (ES)). The most common approach is to quantify these functionals through the empirical distribution function. Other approaches often employed are parametric, which includes the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, and semi-parametric, as Quantile Regression and Filtered Historical Simulation (FHS). The accuracy of the results depends on the reliability of the model used. Although there are studies that compare risk forecast models, for instance, Kuester et al. (2006), Righi and Ceretta (2015), Beckers et al. (2017) and Müller and Righi (2018), there is no consensus regarding the most adequate approach or even a ranking

*Corresponding author

Email addresses: `fernanda.muller@ufrgs.br` (Fernanda Maria Müller), `marcelo.righi@ufrgs.br` (Marcelo Brutti Righi)

of model performance. Furthermore, the models are simplified representations of reality and inevitably acknowledge that there is no perfect model (see Daníelsson (2008) and Federal Reserve (2011)). In literature the hazard of working with a potentially incorrect or inadequate model is referred to as being "model risk"[1]. In risk forecasting literature, Kerkhof et al. (2010), Barrieu and Scandolo (2015), Bernard and Vanduffel (2015), Daníelsson et al. (2016), and Kellner and Rösch (2016) discuss this feature in detail.

A body of research uses backtesting procedures, loss functions, Monte Carlo simulations, and metrics, such as bias and root mean squared error to assess model risk. After revising the Basel II Market Risk Framework, which requires that financial institutions quantify and manage their model risk like other types of risk (see Basel Committee on Banking Supervision (2009) and Federal Reserve (2011)), the attention of literature has focused on procedures to quantify it. From that moment, there is also a consensus that better methods for dealing with model risk are pivotal to improving risk management. However, measures to quantify it are not consolidated at the same level as those for market risk. The main approach used to measure model risk is the worst case. In our study, we consider as worst case measures those that are applied to a set of estimates or forecasts and those with a similar structure to deviation measures. We mention Cont (2006), Kerkhof et al. (2010), and Barrieu and Scandolo (2015) as examples of studies that use it. Besides this approach, it is also possible to employ risk measures in a loss or error function, from some estimation or forecasting procedure. Throughout our research, we name this approach as the loss function. We refer Bignozzi and Tsanakas (2015) and Detering and Packham (2016) as examples of studies that explore it.

In this context, our main objective is to present the literature on model risk measures in risk forecasting[2]. We aim to verify if any pattern in the measures proposed in the literature and unify procedures with similar characteristics into one general approach. We only consider works that propose measures. Empirical applications and adaptations are cited where it is convenient. As the second main objective, we suggest new proposals to quantify the model risk of risk forecast-

[1] An important concept, which interlaces with model risk, is model uncertainty (also referred as model ambiguity). Some definitions used to differentiate them are linked to the concepts of risk and uncertainty, which Knight (1921) discusses. The risk stems from a situation where we do not know the result of the scenario. However, it is still possible to assign probabilities for future results. Also, unlike the risk, uncertainty does not enable us to gain all the necessary information to determine the probability of an event occurring.

[2] We do not explore studies that develop model uncertainty measures in this review since it is not the focus of this study. Cont (2006) proposes the only model uncertainty measure which we describe because it can be easily adapted to quantify model risk.

ing models. For the worst case approach, we explore measures that allow us to quantify upside, downside and tail model risk. Concerning loss function approach, we consider risk overestimation and underestimation distribution in the quantification of model risk. We also conduct an empirical analysis, considering financial data, of model risk measures. In our illustration, we assess model risk of VaR and ES forecasting, which we estimate using well-known GARCH models. We investigate the evolution of model risk across years by adopting a rolling window procedure, considering crisis and non-crisis periods. Furthermore, based on our findings, we highlight future directions that can be explored regarding this subject.

Our paper, to the best of our knowledge, is the first to consider a review regarding model risk measures. Previous studies, such as Branger and Schlag (2004), Sibbertsen et al. (2008), and Bannör and Scherer (2014), respectively explore model risk, which relates to the hedging of derivative contracts, a data-driven notion of model risk, and model risk and model uncertainty in stochastic modeling techniques. Thus, it can be asserted that these authors have a different objective from ours. Additionally, our work presents up to date literature investigating the subject, which can be used as a guide in this research topic. This study also provides theoretical support to investors and managers to examine the main tools to measure this type of risk in risk forecasting.

Moreover, our study also contributes to the literature because we are proposing measures to quantify unexplored characteristics of model risk. In risk forecasting, model risk underestimation has a distinct effect than its overestimation. An example is the different costs resulting from risk underestimation and overestimation (see Dhaene et al. (2003) and Laeven and Goovaerts (2004)). Risk overestimation can generate costs arising from unexpected and uncovered losses, while risk overestimation can generate opportunity costs. However, the worst case measures treat both positive and negative deviations from a reference model as model risk, and loss function measures consider the entire distribution of the errors in the quantifying model risk. Based on this perspective, for worst case measures, we propose measures that allow quantifying upside, downside, and tail model risk. Concerning loss function measures, we explore the use of underestimation and overestimation error distribution to determine model risk value.

Another contribution of our paper is that it is the first to illustrate model risk measures proposed in the literature. For instance, previous studies like, Kellner and Rösch (2016), Barrieu and Scandolo (2015), and Daníelsson et al. (2016), only illustrate the measures proposed by them. Although Gianfreda and Scandolo (2018) conduct an empirical evaluation of model risk, they consider only one model risk measure. Our study also differs because we are checking how model risk behaves in times of crisis and non-crisis. Albeit Daníelsson et al. (2016) assess model risk in crisis and non-crisis periods, they consider only the measure proposed by them.

The structure for the remainder of this paper is: Section 2 displays definitions and preliminary information; Section 3 defines model risk measures proposed in the literature and those we propose; Section 4 describes data and preliminary analysis of VaR and ES forecasting, which we use to illustrate model risk measures; Section 5 illustrates worst case and loss function measures; and Section 6 concludes the paper as well as demonstrating directions for future studies on the subject.

2. Preliminaries

Consider a real valued random result $X : \Omega \to \mathbb{R}$ of an asset or portfolio ($X \geq 0$ is a gain, and $X < 0$ is a loss) defined in a vector space of random variables $\mathcal{X} := \mathcal{X}(\Omega,\mathcal{F},\mathbb{P})$. F_X refers to cumulative distribution function of X, and F_X^{-1} is its (left) inverse. We consider $\mathbb{F} := \{F_X : X \in \mathcal{X}\}$ as a set of distribution functions. We define $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$, and 1_A as the indicator function for an event A. We denote by $\mathbb{R}$ the field of real numbers, and by $\bar{\mathbb{R}}$ the extended real line.

A law invariant monetary risk measure is a functional $\rho : \mathcal{X} \to \bar{\mathbb{R}}$ that fulfills the following properties:

$$\text{(MT)} : \text{if } X \leq Y, \text{then } \rho(X) \geq \rho(Y), \forall\, X, Y \in \mathcal{X}.$$

$$\text{(TI)} : \rho(X + C) = \rho(X) - C, \forall\, X \in \mathcal{X},\ \forall\, C \in \mathbb{R}.$$

$$\text{(LI)} : \text{if } F_X = F_Y, \text{then } \rho(X) = \rho(Y), \forall\, X, Y \in \mathcal{X},\ \forall\, F_X, F_Y \in \mathbb{F}.$$

In our study, (MT), (TI) and (LI) refers to Monotonicity, Translation Invariance, and Law Invariance, respectively.

The (MT) indicates that if the losses of a financial position are greater in all situations, then the expectation for its risk is always greater. The second property, (TI), informs that if adding a certain gain (C) to a position X, it is expected that the risk of this position decreases by the same amount. If a risk measure satisfies both properties it is named as a monetary risk measure. The third property, (LI), indicates that two positions with the same probability function (law) have equal risks. When a risk measure fulfills (LI) it is labeled as a law invariant risk measure. Therefore, if a risk measure satisfies (MT), (TI) and (LI) it is known as a law invariant monetary risk measure.

We highlight that a law invariant risk measure can be represented by a functional $R : \mathbb{F} \to \bar{\mathbb{R}}$, defined as:

$$\rho(X) = R(F_X), \forall\, X \in \mathcal{X}.$$

We provide examples of some law invariant monetary risk measures. We present these risk measures because they are common in risk management literature.

(i) $\text{EL}(X) = -E[X]$,

(ii) $\text{VaR}^\alpha(X) = -F_X^{-1}(\alpha)$,

(iii) $\text{ES}^\alpha(X) = -\frac{1}{\alpha}\int_0^\alpha VaR^s(X)ds$,

where $\alpha \in (0, 1)$ is the significance level. The negative sign of risk measures is used to indicate a monetary loss. The first measure refers to Expected Loss (EL), which computes the expected value (mean) of a loss. This measure is the most parsimonious among those considered. The second measure refers to VaR, which quantifies the maximal loss, which we expect will occur for a financial position, for a given period and confidence level. We also present the ES, which quantifies the expected value of the losses that exceed α - quantile.

3. Model risk measures

Consider $\mathcal{I} := \{1, \cdots, n_1\}$ as a finite set of models used to estimate the distribution function of X. In our framework, $T := \{1, \cdots, n_2\}$ is out-of-sample period and $X_T := \{X_1, \cdots, X_{n_2}\}$ are verifying observations. We denote $\mathbb{G}_X := \left\{\hat{F}_{X,i,t} \colon X \in \mathcal{X}, i \in \mathcal{I}, t \in T\right\} \subseteq \mathbb{F}$, with some abuse of notation, as a set of estimates of the distribution function of X, which we obtain with $i \in \mathcal{I}$ for period $t \in T$. We define, for any $X \in \mathcal{X}$ and $\forall\ \hat{F}_{X,i,t} \in \mathbb{G}_X$, $\rho_{i,t}(X) = R\left(\hat{F}_{X,i,t}\right)$ as a risk forecast for period $t \in T$ obtained using model i in X. Thus, for any $X \in \mathcal{X}$, $\forall\ i \in \mathcal{I}$ and $\forall\ t \in T$, $\rho_{\mathcal{I},T}(X) := \{\rho_{1,T}(X), \cdots, \rho_{n_1,T}(X)\} \in \mathbb{R}^{n_1 \times n_2}$ represents a matrix of risk forecasts. In our notation, $\rho_{i,T}(X)$ represents risk forecasts (for$X$) for out-of-sample period (T) obtained by any model $i \in \mathcal{I}$, and $\rho_{\mathcal{I},t}(X)$ represents risk forecasts (for X) obtained by the set of models ($\mathcal{I}$) for period $t \in T$. In practical sense, $\rho_{\mathcal{I},T}(X)$ can be determined in terms of some divergence centered at forecasts computed by a reference model (see Glasserman and Xu (2014), Breuer and Csiszár (2016) and Krajcovicova et al. (2019)).

Moreover, we consider $\sup \rho_{\mathcal{I},t}(X)$ and $\inf \rho_{\mathcal{I},t}(X)$ as the supremum and infimum values, respectively, of $\rho_{\mathcal{I},t}(X)$ for period $t \in T$. Besides, as $\mathcal{I}$ is finite the supremum and infimum, are, of course, a maximum and a minimum, respectively. When $\inf \rho_{\mathcal{I},t}(X) = \sup \rho_{\mathcal{I},t}(X)$ there is not model risk for period $t \in T$. We delimit a weighted average as a functional $\bar{\mu} : \mathbb{R}^{n_1} \to \bar{\mathbb{R}}$, which can be defined by $\bar{\mu}(X) = \sum_{i=1}^{n_1} x_i \mu(i)$, where $\sum_{i=1}^{n_1} \mu(i) = 1$, $\mu(i) \geq 0 \ \forall \ i \in \mathcal{I}$. We define $\rho_t^{\mu}(X) = \bar{\mu}\left(\rho_{\mathcal{I},t}(X)\right)$ as a weighted risk forecasting for period $t \in T$. It is worth noting that $\inf \rho_{\mathcal{I},t}(X) \leq \rho_t^{\mu}(X) \leq \sup \rho_{\mathcal{I},t}(X)$. In addition, we consider $G_T := \{g_1, \cdots, g_{n_2}\}$ and $L_T := \{l_1, \cdots, l_{n_2}\}$ as being non-negative variables that represent, respectively, costs from risk overestimation and underestimation. In our study, these costs are financial rates traded in the market.

We specify a worst case measure as a functional $MR^{WC} : \mathcal{X} \to \bar{\mathbb{R}}$, defined by $MR^{WC}(X) := f_1 \circ \rho_{\mathcal{I},t}(X) = f_1\left(\rho_{\mathcal{I},t}(X)\right)$, where $f_1 : \mathbb{R}^{n_1} \to \bar{\mathbb{R}}$ is an aggregation function. Moreover, we delimit a loss function measure as a functional $MR_i^{LF} : \mathcal{X} \to \bar{\mathbb{R}}$, defined by $MR_i^{LF}(X) := f_2 \circ \rho_{i,T}(X) = f_2\left(\rho_{i,T}(X)\right)$, where $f_2 : \mathbb{R}^{n_2} \to \bar{\mathbb{R}}$ is an aggregation function. Additionally, we feel that we should note that a discussion regarding the theoretical properties of aggregation functions and model risk measures is beyond the scope of this research. For situations where we are not working with a finite set, it is necessary to consider a more complex framework, as in Righi (2018).

3.1. Worst case approach

One can refer to these examples of worst case measures and the respective study that proposes it:

- $MR^{WC_1}(X) := \sup \rho_{\mathcal{I},t}(X) - \inf \rho_{\mathcal{I},t}(X)$. (Cont, 2006).
- $MR^{WC_2}(X) := \sup \rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)$. (Kerkhof et al., 2010).
- $MR^{WC_3}(X) := \rho_t^{\mu}(X) - \inf \rho_{\mathcal{I},t}(X)$. (Breuer and Csiszár, 2016).
- $MR^{WC_4}(X) := \left(\bar{\mu}\left[\mid \rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X) \mid^p\right]\right)^{\frac{1}{p}}$, where $p \in [1, \infty)$. (Krajcovicova et al., 2019).
- $MR^{WC_5}(X) := \dfrac{\sup \rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)}{\rho_t^{\mu}(X)}$. (Barrieu and Scandolo, 2015).
- $MR^{WC_6}(X) := \dfrac{\sup \rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)}{\sup \rho_{\mathcal{I},t}(X) - \inf \rho_{\mathcal{I},t}(X)}$. (Barrieu and Scandolo, 2015).
- $MR^{WC_7}(X) := \dfrac{\sup \rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)}{\sup \rho_{\mathcal{I},t}(X)}$. (Bernard and Vanduffel, 2015).

- $MR^{WC_8}(X) := \dfrac{\rho_t^{\mu}(X) - \inf \rho_{\mathcal{I},t}(X)}{\inf \rho_{\mathcal{I},t}(X)}$. (Bernard and Vanduffel, 2015).

- $MR^{WC_9}(X) := \dfrac{\sup \rho_{\mathcal{I},t}(X)}{\inf \rho_{\mathcal{I},t}(X)}$. (Daníelsson et al., 2016).

- $MR^{WC_{10}}(X) := \dfrac{\bar{\mu}\left[\mid \rho_{\mathcal{I},t}(X) - \bar{\mu}\left[\rho_{\mathcal{I},t}(X)\right] \mid\right]}{\bar{\mu}\left[\rho_{\mathcal{I},t}(X)\right]}$. (Kellner and Rösch, 2016).

The measures MR^{WC_1} to MR^{WC_4} assume values greater than or equal to zero, and their value is equal to zero only when there is no model risk[3]. For these measures, lower values imply a lower model risk. Besides that, these measures have a similar structure to traditional deviation measures, which includes range - based deviations and p - deviation, belonging to the class of generalized deviation measures (see Rockafellar et al. (2006)). In relation to MR^{WC_5} to $MR^{WC_{10}}$, they are constructed by the ratio of one of the first four measures and a statistic to standardize, such as $\rho_t^{\mu}(X)$, $\sup \rho_{\mathcal{I},t}(X)$ and $\inf \rho_{\mathcal{I},t}(X)$. However, MR^{WC_5}, MR^{WC_8}, MR^{WC_9} and $MR^{WC_{10}}$ can assume negative values because the statistics used as denominators can take negative values. Thus, for these measures, the closer to zero their value, the lower the model risk. Besides, to avoid negative values, some authors, such as Barrieu and Scandolo (2015), assert that the denominator can only assume values greater than zero.

We point out that model risk measures of literature are constructed using a forecast computed by a reference model instead of weighted risk forecasting as performed by us[4]. We suggest this adaptation because the choice of a reference model is subject to preferences and the empirical knowledge of the agent (see Jokhadze and Schmidt (2018)). Besides that, using a reference model is a particular example of our structure once a model receives $\mu(i) = 1$, for some $i \in \mathcal{I}$, for each $t \in T$, and other receive $\mu(j) = 0$, $\forall\, j \in \mathcal{I}, i \neq j$, for each $t \in T$. For both structures, the main objective of these measures is quantifying the dispersion of the results from competing models.

Another point that is worth noting is that worst case measures do not quantify the upside, downside and tail model risk. To quantify these features of model risk, we recommend the following measures:

- $MR^{WC_{11}}(X) := \left(\bar{\mu}\left[\left(\left(\rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)\right)^-\right)^p\right]\right)^{\frac{1}{p}}$, where $p \in [1, \infty)$.

- $MR^{WC_{12}}(X) := \left(\bar{\mu}\left[\left(\left(\rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)\right)^+\right)^p\right]\right)^{\frac{1}{p}}$, where $p \in [1, \infty)$.

[3] In the original proposal of MR^{WC_3}, the functional of interest refers to expected payoffs. The authors consider $\rho_t^{\mu} = 0$ because the daily expected payoff is close to zero.

[4] In this situation, $\bar{\mu}(X) := \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$, $\forall\, i \in \mathcal{I}$, is the arithmetic mean.

- $MR^{WC_{13}}(X) := \frac{1}{n(1-\alpha)}\bar{\mu}\left[\left(\rho_{\mathcal{I},t}(X) - \rho_t^{\mu}(X)\right)|1_{\left(\rho_{\mathcal{I},t}(X) > q_{(1-\alpha)}\left(\rho_{\mathcal{I},t}(X)\right)\right)}\right]$, where α is the confidence level, $q_{(1-\alpha)}(\rho_{\mathcal{I},t}(X))$ refers $(1-\alpha)$ % higher risk forecasts from $\rho_{\mathcal{I},t}(X)$.

We suggest $MR^{WC_{11}}$ to quantify the downside model risk. It allows us to computes the average distance of each estimate belonging to the $\rho_{\mathcal{I},t}(X)$ below from $\rho_t^{\mu}(X)$. From another perspective, $MR^{WC_{12}}$ estimate upside model risk. This measure computes the average distance of each estimate belonging to the $\rho_{\mathcal{I},t}(X)$ above from $\rho_t^{\mu}(X)$. The last measure, $MR^{WC_{13}}$, computes tail model risk. It quantifies the average model risk above $(1-\alpha)$ % more aggressive risk forecasts. For $MR^{WC_{13}}$, we focus on the upper tail because worst case measures aim to identify the worst outcome from the set of candidate models. However, one may extend the measure to analyze α % lower risk forecasts from $\rho_{\mathcal{I},t}(X)$. For these measures, there is no model risk when their value is equal to zero. Besides, the higher the value of the measure, the greater is the model risk.

3.2. Loss function approach

Let $\rho : \mathbb{R}^{n_2} \to \bar{\mathbb{R}}$ be a law invariant monetary risk measure, we refer these examples of loss function measures and the respective research that proposes it[5]:

- $MR_i^{LF_1}(X) := \rho\left(X_T - \rho_{i,T}(X)\right)$. (Bignozzi and Tsanakas, 2015).
- $MR_i^{LF_2}(X) := \rho\left[|\left(X_T - \rho_{i,T}(X)\right)|\right]$. (Detering and Packham, 2016).
- $MR_i^{LF_3}(X) := \rho\left[|\left(X_T - \rho_{i,T}(X)\right)^{-}|\right]$. (Detering and Packham, 2016).

The measures $MR_i^{LF_1}$ and $MR_i^{LF_2}$ consider the entire distribution of the errors, while $MR_i^{LF_3}$ consider only a situation wherein the capital reserve is not enough to cover losses. Differently of worst case measures, these measures allow quantifying model risk of an individual model. In addition, their main objective is to assess the precision of each risk forecasting model. Therefore, they can be used jointly with usual statistics applied to rank the quality of the forecasting models. Besides, these measures are not specific to an individual functional and their use is not conditioned to statistical properties, such as Elicitability[6] for risk measures.

To consider the risk overestimation and underestimation distribution in the quantification of model risk, we intend to generalize the proposal of Detering and Packham (2016). Our model risk measures can be represented in this way:

[5] The reader should not confuse ρ, which is a law invariant monetary risk measure, such as EL, VaR and ES and $\rho_{i,T}$, which refers to risk forecasts for X for the out-of-sample period, which we obtain using model $i \in \mathcal{I}$.

[6] A functional is named elicitable when it is the minimizer of expectation of some score function. See Ziegel (2016) and Acerbi and Szekely (2017).

- $MR_i^{LF_4}(X) := \rho\left[(X_T - \rho_{i,T}(X))^+ G_T + (X_T - \rho_{i,T}(X))^- L_T\right]$.
- $MR_i^{LF_5}(X) := \rho\left[|\ (X_T - \rho_{i,T}(X))^+\ |\right]$.
- $MR_i^{LF_6}(X) := \rho_1\left[(X_T - \rho_{i,T}(X))^+\right] + \rho_2\left[(X_T - \rho_{i,T}(X))^-\right]$, where ρ_1 and ρ_2 are monetary risk measures applied to $(X_T - \rho_{i,T}(X))^+$ and $(X_T - \rho_{i,T}(X))^-$, respectively.

Our first measure, $MR_i^{LF_4}$, is inspired in the robust risk measurement approach proposed by Righi et al. (2019), which minimizes the expectation of sum between costs from overestimation and underestimation. We use their score function here as a model risk measure. This measure allows us to identify the model with the best trade-off between the sum of the costs from risk overestimation and underestimation. In periods of greater instability, in which risk underestimation is more punitive than its overestimation, a higher underestimation cost can be considered to penalize the more expensively underestimation errors. A specification of this measure is to consider different weights rather than the costs applied under each error distribution. In this way, this measure can be defined by $\rho\left[\lambda\,(X_T - \rho_{i,T}(X))^+ + (1-\lambda)\,(X_T - \rho_{i,T}(X))^-\right]$, where $\lambda \in (0{,}1)$, and $\lambda \leq \frac{1}{2}$.

The next measure, $MR_i^{LF_5}$, is the complement of $MR_i^{LF_3}$, which only considers the situation in which the results of position are better than the value determinate by risk measure. Differently, for $MR_i^{LF_6}$, we applied a risk measure on the underestimation and overestimation error distribution. We can use as ρ_1 and ρ_2 the same or a different monetary risk measure. In the financial market, usually, the underestimation error is more serious, so it is expected $\rho_2 \geq \rho_1$. Therefore, in periods of greater instability, a more conservative risk measure can be employed on underestimation errors, resulting in higher levels of security regarding model risk. Besides that, according to loss function measures, the closer to zero their value the lower is model risk.

4. Data and preliminary analysis

In this section, we describe data and a preliminary analysis of the risk forecasts used to illustrate model risk measures. For financial position, X, we consider log-returns of S&P500 U.S. market index multiplied by 100, for the period from January 1, 2001, to May 30, 2018, totalizing 4376 observations. We consider this market index because it is frequently used in academic research. We divide the sample into crisis and non-crisis periods to analyze if the model risk changes in periods with greater variability in the return series. To divide the sample we follow what was suggested by Righi and Vieira (2014). The period considered as a crisis starts on August 1, 2007, until September 28, 2012 (1297 observations). The trading days from August 1, 2007, until July

13, 2010, corresponds to the Subprime mortgage crisis and from June 14, 2010, to September 28, 2012, refers to the Eurozone crisis. Trading days from January 1, 2001, until the Subprime crisis (1660 observations) and after the Eurozone crisis until May 2018 (1419 observations) are non-crisis periods. We refer these periods as before crisis and after crisis periods, respectively[7].

In Table 1, we describe descriptive statistics of log-returns. In the crisis period, the log-returns have negative average value (-0.005) and the highest standard deviation (1.662), suggesting that in this period there is a greater risk. Except for the before crisis period, log-returns display negative asymmetry. Moreover, we observe the presence of heavy-tailed behavior, which indicates a greater probability of extreme values when compared to a normal distribution. These characteristics are commonly observed in daily stock returns data. The log-returns evolution of the index depicted in Figure 1 shows that the series display periods of calm and greater instability, which generally coincide with the non-crisis and crisis periods.

We use VaR and ES, defined in Section 2, as a functional base to be predicted. We choose these measures because they are currently the most common risk measures on risk forecasting literature. We compute these measures using AR(p) (auto-regressive) - GARCH(q,s) model, which can be described in this manner:

$$
\begin{aligned}
X_t &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t, \\
\epsilon_t &= \sigma_t z_t, \qquad z_t \sim i.i.d.\, F(0,1), \\
\sigma_t^2 &= a_0 + \sum_{j=1}^{q} a_j \epsilon_{t-j}^2 + \sum_{k=1}^{s} b_k \sigma_{t-k}^2, \qquad (1)
\end{aligned}
$$

where $t = 1, \cdots, N$ is the period[8], X_t is the return, ϕ_i, for $i = 0, 1, \cdots, p$, are parameters of auto-regressive model, ϵ_t is the innovation in expectation, z_t is a white noise process with distribution F. σ_t^2 is the conditional variance, and a_j, for $j = 0, 1, \cdots, q$, as well as b_k, for $k = 1, \cdots, s$, are parameters of the GARCH model. The parameters are estimated through the Quasi-Maximum Likelihood. The form of the likelihood depends on distributing innovations (z_t), i.e., of F. For F we assume normal (GARCH$_{\text{norm}}$), skewed normal (GARCH$_{\text{snorm}}$), Student-t

[7]BenMim and BenSaïda (2019) use similar periods for the beginning of the Subprime crisis and end of the Eurozone crisis. However, we do not intend to state that the crisis occurred exactly in these periods. We performed the division of the sample in crisis and non-crisis periods because Daníelsson et al. (2016) show that in crisis periods model risk increases.

[8]This model is computed using the information we have in - sample period.

($\text{GARCH}_{\text{std}}$), skewed Student-$t$ ($\text{GARCH}_{\text{sstd}}$), generalized error ($\text{GARCH}_{\text{ged}}$), skewed generalized error ($\text{GARCH}_{\text{sged}}$), normal inverse Gaussian ($\text{GARCH}_{\text{nig}}$), and Johnson SU ($\text{GARCH}_{\text{jsu}}$) distributions. We consider this set of distributions because they are frequently used in the empirical analysis. Besides the normal distribution, we select the distributions that can capture asymmetry and heavy tail. The model used is AR(1)-GARCH(1,1). We select the number of lags through the Akaike information criterion (AIC). For simplicity, when we refer to this model we will name it as the GARCH model. So with some abuse of notation, we consider $\mathcal{I} = \{\text{GARCH}_{\text{norm}}, \text{GARCH}_{\text{snorm}}, \text{GARCH}_{\text{std}}, \text{GARCH}_{\text{sstd}}, \text{GARCH}_{\text{ged}}, \text{GARCH}_{\text{sged}}, \text{GARCH}_{\text{nig}}, \text{GARCH}_{\text{jsu}}\}$. We conduct all computational implementations using `R` programming language (R Core Team, 2019), and the package for estimating the model parameters is `rugarch` (Ghalanos, 2019).

Given a certain distribution assumption for z, VaR and ES forecasts, for each period $t \in T$, are quantified in this way:

$$\begin{aligned} \text{VaR}^{\alpha}_{i,t} &= -\left(\mu_{i,t} + \sigma_{i,t}\text{VaR}^{\alpha}(z_{i,t})\right), \\ \text{ES}^{\alpha}_{i,t} &= -\left(\mu_{i,t} + \sigma_{i,t}\text{ES}^{\alpha}(z_{i,t})\right), \end{aligned} \tag{2}$$

where $\mu_{i,t}$ and $\sigma_{i,t}$ are, respectively, the conditional mean forecasting and conditional standard deviation forecasting, which we compute by $i \in \mathcal{I}$ for period $t \in T$.

As α values, we employ 0.01 for VaR, and 0.025 for ES because the Basel Committee on Banking Supervision (see Basel Committee on Banking Supervision (2013)) recommends these values. In the estimation process, we use a rolling estimation window of 250 days[9]. At each step, we obtain one-step-ahead risk forecasts. In this sense, for each day in the out-sample period, we use the last 250 observations to compute the risk measures. For descriptive analysis of point forecasts, we present in Table 2 average value (Mean), standard deviation (SD), and realized loss ($\mathcal{L}_\rho$), which for VaR and ES, respectively, we compute as (see Gneiting (2011) and Fissler and Ziegel (2016)):

$$\mathcal{L}_{\text{VaR}^{\alpha}_{i}} = \frac{1}{n_2}\sum_{t=1}^{n_2}\left[\alpha\left(X_t + \text{VaR}^{\alpha}_{i,t}\right)^{+} + (1-\alpha)\left(X_t + \text{VaR}^{\alpha}_{i,t}\right)^{-}\right],$$

$$\mathcal{L}_{\text{ES}^{\alpha}_{i}} = \frac{1}{n_2}\sum_{t=1}^{n_2}[(I_{i,t}-\alpha)\left(-\text{VaR}^{\alpha}_{i,t}\right) - I_{i,t}X_t + e^{\left(-\text{ES}^{\alpha}_{i,t}\right)} \times \left(\left(-\text{ES}^{\alpha}_{i,t}\right) + \text{VaR}^{\alpha}_{i,t} + \frac{I_{i,t}}{\alpha}\left(\left(-\text{VaR}^{\alpha}_{i,t}\right) - X_t\right)\right)$$

[9]We select this rolling estimation window because is common in risk forecasting literature (see Bayer (2018) and Argyropoulos and Panopoulou (2019)) and it is recommended by the Basel Committee (see Basel Committee on Banking Supervision (2013)).

$$- e^{\left(-\mathrm{ES}_{i,t}^{\alpha}\right)} + 1 - \log\left(1-\alpha\right)], \tag{3}$$

where $T := \{1, \cdots, n_2\}$ is out-of-sample period, $X_t \in X_T$, $I_{i,t} = 1_{X_t < -\mathrm{VaR}_{i,t}^{\alpha}}$, $\forall\, t \in T$.

We observe in Table 2, except for $\mathrm{GARCH}_{\mathrm{nig}}$, that average and standard deviation value of VaR and ES forecasts are higher in the crisis period. For $\mathrm{GARCH}_{\mathrm{nig}}$ the higher values of these statistics are found in the before crisis period. On the other hand, during the after crisis period risk forecasts present the lowest standard deviation and average values. Moreover, we can state that these results are consistent with the descriptive analysis of the log-returns. For VaR forecasts, $\mathrm{GARCH}_{\mathrm{ged}}$ has lower realized loss. This model also has the best performance concerning ES forecasting throughout the whole sample and after the crisis period. During the before crisis period, for ES forecasts, $\mathrm{GARCH}_{\mathrm{sged}}$ and $\mathrm{GARCH}_{\mathrm{sstd}}$ have a lower realized loss, while during crisis period only $\mathrm{GARCH}_{\mathrm{sged}}$ has. Contrastingly, $\mathrm{GARCH}_{\mathrm{nig}}$ displays the worst performance for both risk measures.

We provide, in Figure 2, the evolution of S&P500 log-returns, and VaR and ES forecasting with converted signal considering the whole sample. For brevity, we omit the illustrations of risk forecasts for the sub-samples. They are available upon request[10]. In line with the descriptive analysis, as seen, the evolution of forecasts obtained considering $\mathrm{GARCH}_{\mathrm{nig}}$ are far from the evolution of log-returns as well as the results of the other models. This behavior corroborates with the realized loss values of forecasts obtained for this model.

5. Empirical Results

5.1. Worst case measures

Aiming to illustrate worst case measures, we use the risk forecasts of VaR and ES described in Section 4. For each one-step-ahead, we compute model risk using $MR^{WC_m}, m = 1, \cdots, 13$. We quantify ρ_t^{μ} by means of equally weighted scheme, $\mu(i) = \frac{1}{8}$, $\forall\ i \in \mathcal{I}$, $\forall\ t \in T$. We also consider the situation in which instead of ρ_t^{μ} we have forecasts obtained from a reference model. These measures are defined by $MR_0^{WC_m}, m = 2, 3, 4, 5, 6, 7, 8, 11, 12, 13$. Our reference model, as performed by Kerkhof et al. (2010) and Krajcovicova et al. (2019), is the model that follows a normal distribution, i.e., $\mathrm{GARCH}_{\mathrm{norm}}$. For MR^{WC_4}, $MR_0^{WC_4}$, $MR^{WC_{11}}$, $MR_0^{WC_{11}}$, $MR^{WC_{12}}$ and $MR_0^{WC_{12}}$ we just describe the results with $p = 1$. For $MR^{WC_{13}}$ and $MR_0^{WC_{13}}$ we use $\alpha\% = 30\%$, and so $(1-\alpha)\% = 70\%$. For each model risk measure, we present the average and standard

[10]To keep the pattern, we will present the figures, in all illustrations, only for the whole sample.

deviation value of model risk estimates, and the ratio of model risk about weighted risk forecasting in % (Prop (%)). Prop (%) allows us to assess the proportion of model risk in relation to weighted risk measure[11]. We expose, considering the whole sample and sub-samples, these results, for VaR and ES forecasts, in Tables 3 and 4, respectively.

It can be highlighted that the mean values of model risk differ according to measures, sample and the functional (VaR or ES). The standard deviation values indicate variability of the model risk estimates over the period analyzed. For VaR forecasts, we note that the highest average model risk value is computed by MR^{WC_1}, being its highest estimate observed in the before crisis period (7.073). In this period, MR^{WC_1} has the highest standard deviation (13.320) and it assumes value more than twice as high as ρ_t^{μ} (Prop (%) = 236.092%). Regarding ES forecasts, the highest averages values are quantified by MR^{WC_1} and $MR^{WC_{10}}$ in the whole sample (5.129) and after crisis period (43.680), respectively. Results from MR^{WC_1} allow us to conclude there is a large amplitude between the minimum and maximum values of $\rho_{\mathcal{I},t}(X)$. This dispersion can lead to an unbalanced regulatory environment, once the individual risk measures determine different amounts of regulatory capital. The magnitude of the results of these measures become clearer when we see Figures 3 and 4, which evolve model risk and weighted risk forecasting (gray line), computed for VaR and ES forecasting, respectively (whole sample).

Among the periods (samples) investigated, we perceive that for most measures, for instance MR^{WC_1}, $MR_0^{WC_2}$, MR^{WC_2}, $MR_0^{WC_5}$, MR^{WC_5}, $MR_0^{WC_6}$, MR^{WC_6}, $MR_0^{WC_8}$, MR^{WC_9}, $MR^{WC_{10}}$, $MR_0^{WC_{12}}$, $MR^{WC_{12}}$, model risk estimates (in absolute value) are significantly higher in the before crisis period than crisis and after crisis period[12]. Unlike our findings, in the study of Daníelsson et al. (2016), using MR^{WC_9}, it was found that crisis period increases model risk. Our results differ, possibly, due to the atypical behavior of $\text{GARCH}_{\text{nig}}$ estimates. The worst case measures are sensitive to the influence of a particular model[13], which is one of the main criticisms of these measures. When

[11] For worst case measures estimated with ρ_t^{μ} Prop (%) $= \left(\frac{MR^{WC_m}(X)}{\rho_t^{\mu}(X)}\right) \times 100$, , $m = 1, \cdots, 13$, while worst case measures estimated with reference model Prop (%) $= \left(\frac{MR_0^{WC_m}(X)}{\rho_t^{\mu}(X)}\right) \times 100$, $m = 2, 3, 4, 5, 6, 7, 8, 11, 12, 13$.

[12] We used Mann - Whitney U test to investigate whether the model risk estimates are significantly higher in the before crisis period than crisis period and after crisis period. Therefore, we compare the model risk estimates (from the before crisis period) by testing it against the other periods. The null hypothesis is true location shift is equal to 0 and the alternative hypothesis true location shift is greater than 0. For brevity, p-value and test statistics are not displayed. We use the Mann-Whitney U test because it is a non-parametric test and it is based on fewer assumptions regarding sample distributions. They are available upon request.

[13] In our study, we did not perform the analysis without $\text{GARCH}_{\text{nig}}$ because we intend to illustrate how the different model risk measures behave.

analyzing the temporal evolution of MR^{WC_9}, for both VaR and ES, we identify that two periods with high estimates coincide with the higher volatility of risk forecasts quantify by $\text{GARCH}_{\text{nig}}$. According to Figures 3 and 4, besides MR^{WC_9}, model risk estimates from MR^{WC_1}, MR^{WC_2}, $MR_0^{WC_2}$, MR^{WC_3}, $MR_0^{WC_3}$, MR^{WC_8}, $MR_0^{WC_8}$ and $MR^{WC_{10}}$ are also affected by $\text{GARCH}_{\text{nig}}$ estimates. Besides, Daníelsson et al. (2016) consider different methodological characteristics (for example, financial position, size of the sample and rolling window estimation). Furthermore, we observe model risk is significantly lower (in absolute value) in the after crisis period than before crisis and crisis period[14]. One of the possible explanations for this finding is due to the lower variability of the risk forecasts for this period.

In relation to the measures computed using ρ_t^{μ} and a reference model, we verify some patterns in the results of these measures. For measures proposed by Kerkhof et al. (2010) $\left(MR^{WC_2}\right)$ and by Barrieu and Scandolo (2015) $\left(MR^{WC_5}\text{ and }MR^{WC_6}\right)$ model risk average value tends to be higher when estimated with the reference model. For measures of Breuer and Csiszár (2013) $\left(MR^{WC_3}\right)$ and of Krajcovicova et al. (2019) $\left(MR^{WC^4}\right)$ we identify an opposing result. Regarding the results of VaR and ES, we realize that they are conditioned to worst case measure used. In the study by Kellner and Rösch (2016), $MR^{WC_{10}}$ is used to assess model risk. Since this measure is standardized by $\bar{\mu}\left[\rho_{\mathcal{I},t}(X)\right]$, it allows the comparison of VaR and ES results. The higher the result of this measure, the more dispersed is the capital requirements of a financial asset. Our results indicate that ES is more sensitive to regulatory arbitrage[15]. For instance, in the whole sample, for VaR, $MR^{WC_{10}}$ has a value equal to -0.084, while for ES this measure assumes a value equal to 12.536. Regarding the sub-samples, the measure also presents higher values for the ES forecasts. Similar results are verified by Kellner and Rösch (2016).

5.2. Loss function measures

We also consider VaR and ES forecasting, to illustrate loss function measures. Additionally, since we are considering negative results as losses, we first correct the sign of each risk forecasts $\rho_{i,t}(X) \in \rho_{\mathcal{I},T}(X)$. For risk forecasting obtained by each model, $\rho_{i,T}(X)$, we compute model risk using $MR_i^{LF^m}, m = 1, \cdots, 7$. We use EL as law invariant monetary risk measure, which we

[14] We used Mann - Whitney U test to investigate whether the model risk estimates are significantly lower in the after crisis period than before crisis and crisis period. For brevity, p-value and test statistics are not displayed. They are available upon request.

[15] In the sense used, regulatory arbitrage refers to two institutions with the same portfolio and uses different internal models, approved by the regulator, and so quantify different amounts of capital requirement. As they keep the same portfolio, they must hold the same or at least almost the same amount of regulatory capital.

quantify non-parametrically, i.e., $EL(X) = \frac{1}{n_2}\sum_{t=1}^{n_2} x_t$. We choose this risk measure to maintain the estimation pattern of functions used to quantify realized loss (see formulation (3)). For $MR_i^{LF_6}$, we opted in to use $\rho_1 = \rho_2 =$ EL. As costs of risk overestimation G_T and underestimation L_T, we used daily yield rates of the U.S. Treasury Bill with a maturity of three months and the U.S. Dollar based Overnight London Interbank Offered Rate (LIBOR). These assets are commonly used in the literature. Moreover, these rates reflect a risk-free investment with liquidity, where the surplus over capital requirement can be invested, and a rate for loans, when the capital requirement is not enough, respectively. We convert both yield rates to a daily frequency. Figure 5 presents the temporal evolution of these series multiplied by 100. We observe a huge change with their dynamics in the early of 2005 and the end of 2008. This change is possibly due to economic events that the Subprime crisis generated. For each model risk measure, we present the EL, which is the average value of loss function measure, and standard deviation value. These results are reported in Table 5, for VaR, and in Table 6, for ES.

We identify that the forecasting model with the highest model risk coincides with the model's worst performance according to realized loss, which refers to $\text{GARCH}_{\text{nig}}$ model. On the other hand, the forecast model with the lowest model risk does not match with the model with lower realized loss. For example, in the whole sample, for VaR, the risk forecasts with lowest realized loss are quantified with $\text{GARCH}_{\text{ged}}$ (see Table 2); while forecasts with lower model risk according $MR_i^{LF_3}$ are estimated with $\text{GARCH}_{\text{std}}$. For $MR_i^{LF_1}$, $MR_i^{LF_2}$, $MR_i^{LF_5}$, and $MR_i^{LF_6}$, we notice advantages of risk forecasts estimated with $\text{GARCH}_{\text{snorm}}$. Regarding $MR_i^{LF_4}$, forecasts obtained by $\text{GARCH}_{\text{norm}}$ and $\text{GARCH}_{\text{snorm}}$ have the smallest model risk. Thus, it is worth stressing that the best performance to forecast risk measures does not necessarily mean a low model risk. A similar result was verified by Gianfreda and Scandolo (2018).

We also note that model risk estimates are significantly higher in crisis period than before crisis and after crisis period[16]. During this period, we also observe a higher standard deviation. Thus, according to these results, we conclude that during crisis period the model risk increases. This increase primarily occurs, as it is exposed by Daníelsson (2008), due to the assumption, in most statistical risk modeling, that the basic statistical properties of financial series during calm periods remain or about the same as periods of instability (crisis). In periods with greater market

[16]We also use Mann - Whitney U test to investigate whether the model risk estimates are significantly higher in the crisis period than before crisis, and after crisis period. For brevity, p-value and test statistics are also not displayed. They are available upon request.

uncertainty, model estimates tend to be least reliable. However, these results do not corroborate with those identified for worst case measures. A possible explanation for distinct results between both approaches is the conceptual differences between them. Worst case measures evaluate the dispersion among forecasts from different models, while loss function measures evaluate the precision of each model. On the other hand, we verify, similar to the worst case approach measures, except for $MR_i^{LF_3}$ and $MR_i^{LF_4}$, that model risk is significantly lower during after crisis period than before crisis and crisis period[17].

Regarding the individual results of each measure, we observe that $MR_i^{LF_3}$ has lower values than $MR_i^{LF_5}$. This result indicates that on average the position has a better result than risk forecasting. This is a characteristic commonly observed in risk measures forecasting with GARCH models. See Hwang and Valls Pereira (2006), Carnero et al. (2007) and Müller and Righi (2018). This result also justifies the similarity between the results of $MR_i^{LF_5}$ and $MR_i^{LF_6}$.

Concerning $MR_i^{LF_4}$, we note that on average its value is close to zero. This is also observed when we see the evolution of their values in Figures 6 and 7, which display results for VaR and ES, respectively[18]. The advantage of this measure about the other is that it allows quantifying the average costs from risk overestimation and underestimation. Generally, among the models, for both VaR and ES, $\text{GARCH}_{\text{nig}}$ has the greatest cost. In the after the crisis period, $\text{GARCH}_{\text{std}}$ presents the worst result. Curiously, this model, in this period, has the lowest realized loss for VaR forecasts. We can explain this difference by the fact of the realized loss computed from elicitable functions, especially for VaR (see formulation (3)), penalizes more heavily the observations for which we note returns showing risk estimates exceedance. Moreover, unlike our model risk measure, elicitable loss functions only consider forecasting errors, rather than the costs associated with such errors.

6. Conclusions and future directions

We provide a review regarding model risk measures. Based on our findings, we propose new model risk measures to capture unexplored characteristics of model risk. We also conduct an empirical assessment of model risk measures using VaR and ES forecasting obtained by well-known GARCH models. Moreover, we will highlight insights for future research directions regarding this

[17]For brevity, p-value and test statistics of Mann - Whitney U test are also not displayed. They are available upon request.

[18]For brevity, these figures display the results for $MR_i^{LF_1}$ and $MR_i^{LF_4}$ considering the eight models used and whole sample. The illustrations of the other loss function measures are available upon request.

topic.

We realize that the model risk measures should be categorized into two main groups, which refer to the worst case and loss function approach. The first group has a similar structure to deviation measures, and they are applied to a set of forecasts. Furthermore, due to the fact that model risk underestimation and overestimation have a distinct effect, we suggest model risk measures to quantify upside, downside and tail model risk. On the other hand, the second group employs monetary risk measures in an error or loss function, from some forecasting procedure. As advantages of these measures, we refer to the possibility of employing them as a complementary criterion for forecast model selection. For this approach, we recommend measures that consider risk overestimation and underestimation distribution in the quantification of model risk. Our empirical results indicate, according to these measures, that the model risk increases during the crisis period. Contrastingly, worst case measures show that before crisis period has the highest model risk values. The conflicting results between the two approaches can be explained by the conceptual differences between each of them. We also observed that a model with good performance to risk forecasting, i.e., with lower realized loss, does not indicate this model has lower model risk.

Additionally, we verify that current studies focus on model risk measures coming from individual models. However, a point that deserves attention, is to consider the model risk originally from aggregate models. As Federal Reserve (2011), the model risk is affected by interaction and dependencies among models and determining its magnitude might help to manage it correctly.

Another point, which remains open is the formalization of the theoretical properties of model risk measures. Theoretical discussions gained a boost, in the risk management literature, after Artzner et al. (1999)'s pioneering work. Cont (2006), Barrieu and Scandolo (2015), Lazar and Zhang (2019) present an initial discussion regarding this topic. However, these studies focus on the individual characteristics of their measure. One possibility to circumvent some of these limitations is to formalize a theoretical framework for worst case measures using as basis deviation measures literature. Furthermore, we can confidently state that this framework is consistent and easily interpretable in a model risk context. However, we do not claim that the theoretical properties of deviation measures are perfect in a model risk context. Naturally, one can think of imposing another axiomatic body. Although, they are at least a starting point to gain a more robust and solid discussion regarding using model risk measures in financial analysis. Therefore, future works need to conduct a more detailed investigation to identify a complete axiomatic structure for model risk measures.

References

Acerbi, C., Szekely, B., 2017. General properties of backtestable statistics. Working Paper URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2905109.

Argyropoulos, C., Panopoulou, E., 2019. Backtesting VaR and ES under the magnifying glass. International Review of Financial Analysis 64, 22–37.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Bannör, K.F., Scherer, M., 2014. Model risk and uncertainty - illustrated with examples from mathematical finance, in: Risk-A Multidisciplinary Introduction. Springer, pp. 279–306.

Barrieu, P., Scandolo, G., 2015. Assessing financial model risk. European Journal of Operational Research 242, 546–556.

Basel Committee on Banking Supervision, 2009. Revisions to the Basel II market risk framework. Bank for International Settlements. URL: https://www.bis.org/publ/bcbs158.htm.

Basel Committee on Banking Supervision, 2013. Fundamental review of the trading book: A revised market risk framework. Consultative Document, October URL: https://www.bis.org/publ/bcbs265.pdf.

Bayer, S., 2018. Combining value-at-risk forecasts using penalized quantile regressions. Econometrics and statistics 8, 56–77.

Beckers, B., Herwartz, H., Seidel, M., 2017. Risk forecasting in (T) GARCH models with uncorrelated dependent innovations. Quantitative Finance 17, 121–137.

BenMim, I., BenSaïda, A., 2019. Financial contagion across major stock markets: A study during crisis episodes. The North American Journal of Economics and Finance 48, 187–201.

Bernard, C., Vanduffel, S., 2015. A new approach to assessing model risk in high dimensions. Journal of Banking & Finance 58, 166–178.

Bignozzi, V., Tsanakas, A., 2015. Parameter uncertainty and residual estimation risk. The Journal of Risk and Insurance 83, 949– 978.

Branger, N., Schlag, C., 2004. Model risk: A conceptual framework for risk measurement and hedging. Conference Paper URL: `https://ssrn.com/abstract=493482orhttp://dx.doi.org/10.2139/ssrn.493482`.

Breuer, T., Csiszár, I., 2013. Systematic stress tests with entropic plausibility constraints. Journal of Banking & Finance 37, 1552–1559.

Breuer, T., Csiszár, I., 2016. Measuring distribution model risk. Mathematical Finance 26, 395–411.

Carnero, M.A., Pena, D., Ruiz, E., 2007. Effects of outliers on the identification and estimation of GARCH models. Journal of Time Series Analysis 28, 471–497.

Cont, R., 2006. Model uncertainty and its impact on the pricing of derivative instruments. Mathematical Finance 16, 519–547.

Daníelsson, J., 2008. Blame the models. Journal of Financial Stability 4, 321–328.

Daníelsson, J., James, K.R., Valenzuela, M., Zer, I., 2016. Model risk of risk models. Journal of Financial Stability 23, 79–91.

Detering, N., Packham, N., 2016. Model risk of contingent claims. Quantitative Finance 16, 1357–1374.

Dhaene, J., Goovaerts, M.J., Kaas, R., 2003. Economic capital allocation derived from risk measures. North American Actuarial Journal 7, 44–56.

Federal Reserve, 2011. Supervisory guidance on model risk management. Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency , 11–7URL: `https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf`.

Fissler, T., Ziegel, J.F., 2016. Higher order elicitability and osbands principle. The Annals of Statistics 44, 1680–1707.

Ghalanos, A., 2019. Package rugarch URL: `https://cran.r-project.org/web/packages/rugarch/rugarch.pdf`.

Gianfreda, A., Scandolo, G., 2018. Measuring model risk in the european energy exchange, in: Handbook of Recent Advances in Commodity and Financial Modeling. Springer, pp. 89–110.

Glasserman, P., Xu, X., 2014. Robust risk measurement and model risk. Quantitative Finance 14, 29–58.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106, 746–762.

Hwang, S., Valls Pereira, P.L., 2006. Small sample properties of GARCH estimates and persistence. The European Journal of Finance 12, 473–494.

Jokhadze, V., Schmidt, W.M., 2018. Measuring model risk in financial risk management and pricing. Working Paper URL: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3113139`.

Kellner, R., Rösch, D., 2016. Quantifying market risk with value-at-risk or expected shortfall?–consequences for capital requirements and model risk. Journal of Economic Dynamics and Control 68, 45–63.

Kerkhof, J., Melenberg, B., Schumacher, H., 2010. Model risk and capital reserves. Journal of Banking & Finance 34, 267–279.

Knight, F.H., 1921. Uncertainty and profit.

Krajcovicova, Z., Perez-Velasco, P.P., Vázquez Cendón, C., 2019. A new approach to the quantification of model risk for practitioners. Journal of Computational Finance 23.

Kuester, K., Mittnik, S., Paolella, M.S., 2006. Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4, 53–89.

Laeven, R.J., Goovaerts, M.J., 2004. An optimization approach to the dynamic allocation of economic capital. Insurance: Mathematics and Economics 35, 299–319.

Lazar, E., Zhang, N., 2019. Model risk of expected shortfall. Journal of Banking & Finance 105, 74 – 93.

Müller, F.M., Righi, M.B., 2018. Numerical comparison of multivariate models to forecasting risk measures. Risk Management 20, 29–5.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Righi, M.B., 2018. A theory for combinations of risk measures. Working Paper URL: https://arxiv.org/abs/1807.01977.

Righi, M.B., Ceretta, P.S., 2015. A comparison of expected shortfall estimation models. Journal of Economics and Business 78, 14–47.

Righi, M.B., Müller, F.M., Moresco, M.R., 2019. A robust approach for minimization of risk measurement errors. Working Paper URL: https://arxiv.org/abs/1707.09829.

Righi, M.B., Vieira, K.M., 2014. Liquidity spillover in international stock markets through distinct time scales. Plos One 9.

Rockafellar, R.T., Uryasev, S., Zabarankin, M., 2006. Generalized deviations in risk analysis. Finance and Stochastics 10, 51–74.

Sibbertsen, P., Stahl, G., Luedtke, C., 2008. Measuring model risk. Journal of Risk Model Validation 2, 65–81.

Ziegel, J.F., 2016. Coherence and elicitability. Mathematical Finance 26, 901–918.

Table 1: Summary statistics of the S&P500 log-returns (in %) for the whole sample (January 1, 2001 to May 30, 2018) and sub-samples (before crisis, January 1, 2001 until July 30, 2007, crisis, August 1, 2007 until September 28, 2012, and after crisis, September 29, 2012 until May 30, 2018).

Statistics	Whole Sample	Before Crisis	Crisis	After Crisis
Mean	0.015	0.008	-0.005	0.043
Minimum	-9.470	-5.047	-9.470	-4.184
Maximum	10.957	5.574	10.957	5.321
Standard Deviation	1.199	1.050	1.662	0.788
Skewness	-0.208	0.116	-0.232	-0.324
Excess kurtosis	9.357	2.990	6.674	4.214

Table 2: Average value (Mean), standard deviation (SD), and realized loss ($\mathcal{L}_\rho$) for VaR$^{0.01}$ and ES$^{0.025}$ forecasts. The results are exposed to the whole sample (January 1, 2001, to May 30, 2018) and sub-samples (before crisis, January 1, 2001 until July 30, 2007, crisis, August 1, 2007 until September 28, 2012, and after crisis, September 29, 2012 until May 30, 2018) for S&P500 log-returns (in %).

	VaR$^{0.01}$			ES$^{0.025}$		
Whole Sample	Mean	SD	$\mathcal{L}_{\text{VaR}_i^\alpha}$	Mean	SD	$\mathcal{L}_{\text{ES}_i^\alpha}$
$\text{GARCH}_{\text{norm}}$	2.372	1.447	0.036	2.384	1.454	1.020
$\text{GARCH}_{\text{snorm}}$	2.231	1.368	0.038	2.241	1.374	1.037
$\text{GARCH}_{\text{std}}$	3.181	1.924	0.036	3.306	2.026	1.030
$\text{GARCH}_{\text{sstd}}$	2.446	1.442	0.036	2.522	1.474	1.013
$\text{GARCH}_{\text{ged}}$	2.629	1.582	**0.034**	2.662	1.600	**1.012**
$\text{GARCH}_{\text{sged}}$	2.464	1.464	0.035	2.494	1.480	1.014
$\text{GARCH}_{\text{nig}}$	4.607	9.965	0.552	3.987	7.413	> 100.000
$\text{GARCH}_{\text{jsu}}$	2.432	1.436	0.036	2.487	1.462	1.017
	VaR$^{0.01}$			ES$^{0.025}$		
Before Crisis	Mean	SD	$\mathcal{L}_{\text{VaR}_i^\alpha}$	Mean	SD	$\mathcal{L}_{\text{ES}_i^\alpha}$
$\text{GARCH}_{\text{norm}}$	2.088	0.920	0.026	2.099	0.925	0.968
$\text{GARCH}_{\text{snorm}}$	2.062	0.974	0.027	2.072	0.979	0.969
$\text{GARCH}_{\text{std}}$	2.361	1.090	0.026	2.403	1.120	0.971
$\text{GARCH}_{\text{sstd}}$	2.164	1.024	0.026	2.199	1.039	**0.960**
$\text{GARCH}_{\text{ged}}$	2.183	0.965	**0.025**	2.200	0.973	0.963
$\text{GARCH}_{\text{sged}}$	2.142	1.004	0.026	2.159	1.011	**0.960**
$\text{GARCH}_{\text{nig}}$	8.796	13.980	0.114	6.861	8.879	3.995
$\text{GARCH}_{\text{jsu}}$	2.170	1.032	0.026	2.198	1.047	0.961
	VaR$^{0.01}$			ES$^{0.025}$		
Crisis	Mean	SD	$\mathcal{L}_{\text{VaR}_i^\alpha}$	Mean	SD	$\mathcal{L}_{\text{ES}_i^\alpha}$
$\text{GARCH}_{\text{norm}}$	3.435	2.161	0.053	3.452	2.171	1.096
$\text{GARCH}_{\text{snorm}}$	3.146	2.037	0.057	3.159	2.046	1.106
$\text{GARCH}_{\text{std}}$	4.595	2.557	0.053	4.761	2.632	1.116
$\text{GARCH}_{\text{sstd}}$	3.430	2.100	0.053	3.523	2.130	1.087
$\text{GARCH}_{\text{ged}}$	3.824	2.291	**0.051**	3.870	2.311	1.090
$\text{GARCH}_{\text{sged}}$	3.483	2.120	0.052	3.523	2.140	**1.086**
$\text{GARCH}_{\text{nig}}$	4.983	4.149	0.057	5.068	4.201	1.124
$\text{GARCH}_{\text{jsu}}$	3.379	2.092	0.053	3.445	2.116	1.088
	VaR$^{0.01}$			ES$^{0.025}$		
After Crisis	Mean	SD	$\mathcal{L}_{\text{VaR}_i^\alpha}$	Mean	SD	$\mathcal{L}_{\text{ES}_i^\alpha}$
$\text{GARCH}_{\text{norm}}$	1.766	0.725	0.031	1.774	0.728	1.002
$\text{GARCH}_{\text{snorm}}$	1.638	0.695	0.034	1.645	0.698	1.032
$\text{GARCH}_{\text{std}}$	2.635	1.308	**0.030**	2.778	1.461	1.004
$\text{GARCH}_{\text{sstd}}$	1.874	0.819	0.032	1.967	0.888	0.999
$\text{GARCH}_{\text{ged}}$	2.007	0.861	**0.030**	2.039	0.878	**0.990**
$\text{GARCH}_{\text{sged}}$	1.895	0.882	0.032	1.925	0.901	1.002
$\text{GARCH}_{\text{nig}}$	0.040	7.445	1.744	0.079	7.475	>100.000
$\text{GARCH}_{\text{jsu}}$	1.871	0.857	0.033	1.936	0.909	1.010

Note: The bold values refers to the model with the best performance for VaR$^{0.01}$ and ES$^{0.025}$ forecasting (lower realized loss). Risk forecasting are estimated using GARCH model model with normal ($\text{GARCH}_{\text{norm}}$), skewed normal ($\text{GARCH}_{\text{snorm}}$), Student-$t$ ($\text{GARCH}_{\text{std}}$), skewed Student-$t$ ($\text{GARCH}_{\text{sstd}}$), generalized error ($\text{GARCH}_{\text{ged}}$), skewed generalized error ($\text{GARCH}_{\text{sged}}$), normal inverse Gaussian ($\text{GARCH}_{\text{nig}}$), and Johnson SU ($\text{GARCH}_{\text{jsu}}$) distributions. The rolling estimation window is of 250 observations.

Table 3: Descriptive analysis of worst case measures for VaR$^{0.01}$ forecasts. The results are exposed to the whole sample (January 1, 2001, to May 30, 2018) and sub-samples (before crisis, January 1, 2001 until July 30, 2007, crisis, August 1, 2007 until September 28, 2012, and after crisis, September 29, 2012 until May 30, 2018) for S&P500 log-returns (in %).

	Whole Sample			Before Crisis			Crisis			After Crisis		
Measures	Mean	SD	Prop (%)	Mean	SD	Prop (%)	Mean	SD	Prop (%)	Mean	SD	Prop (%)
MR^{WC_1}	4.136	9.100	147.956	7.073	13.320	236.092	2.575	2.040	68.051	2.991	7.603	174.325
$MR_0^{WC_2}$	3.386	8.315	121.118	6.912	13.380	230.703	2.279	1.956	60.234	0.896	0.785	52.243
MR^{WC_2}	2.963	7.260	105.972	6.004	11.678	200.411	1.930	1.679	51.009	0.946	1.144	55.147
$MR_0^{WC_3}$	0.750	4.128	26.838	0.161	0.406	5.389	0.296	0.205	7.818	2.095	7.576	122.082
MR^{WC_3}	1.174	3.694	41.983	1.069	1.672	35.680	0.645	0.385	17.043	2.045	6.593	119.178
$MR_0^{WC_4}$	0.670	1.144	23.951	0.976	1.679	32.578	0.507	0.318	13.395	0.551	0.967	32.103
MR^{WC_4}	0.903	1.972	32.313	1.536	2.905	51.261	0.568	0.413	15.016	0.656	1.619	38.256
$MR_0^{WC_5}$	1.274	2.517	45.561	2.604	3.957	86.913	0.634	0.284	16.761	0.511	0.314	29.774
MR^{WC_5}	1.112	2.199	39.789	2.263	3.456	75.539	0.531	0.232	14.043	0.505	0.384	29.419
$MR_0^{WC_6}$	0.844	0.228	30.199	0.888	0.253	29.624	0.853	0.097	22.548	0.764	0.291	44.520
MR^{WC_6}	0.690	0.153	24.697	0.756	0.176	25.235	0.718	0.063	18.982	0.597	0.158	34.808
$MR_0^{WC_7}$	0.408	0.190	14.601	0.526	0.250	17.542	0.372	0.095	9.832	0.316	0.109	18.440
MR^{WC_7}	0.364	0.210	13.037	0.457	0.211	15.247	0.313	0.075	8.265	0.334	0.274	19.479
$MR_0^{WC_8}$	-3.046	196.722	-108.950	-9.358	336.310	-312.370	0.102	0.060	2.684	0.404	12.608	23.566
MR^{WC_8}	-2.342	165.017	-83.780	-7.488	282.118	-249.927	0.212	0.051	5.605	0.466	10.411	27.175
MR^{WC_9}	-0.752	200.415	-26.895	-5.877	342.546	-196.159	1.791	0.260	47.320	2.048	16.027	119.392
$MR^{WC_{10}}$	-0.084	34.805	-3.008	0.337	0.292	11.261	0.146	0.033	3.850	-0.894	65.401	-52.105
$MR_0^{WC_{11}}$	0.123	0.530	4.402	0.034	0.065	1.143	0.079	0.074	2.085	0.300	0.967	17.504
$MR^{WC_{11}}$	0.452	0.986	16.157	0.768	1.452	25.630	0.284	0.206	7.508	0.328	0.810	19.128
$MR_0^{WC_{12}}$	0.546	1.051	19.548	0.942	1.692	31.435	0.428	0.301	11.310	0.250	0.218	14.599
$MR^{WC_{12}}$	0.452	0.986	16.157	0.768	1.452	25.630	0.284	0.206	7.508	0.328	0.810	19.128
$MR_0^{WC_{13}}$	1.352	2.780	48.357	2.429	4.483	81.081	1.071	0.776	28.302	0.527	0.415	30.697
$MR^{WC_{13}}$	0.928	1.781	33.211	1.522	2.778	50.790	0.722	0.501	19.077	0.577	0.983	33.601

Note: This table presents results from worst case measures. Measures are applied in VaR$^{0.01}$ forecasts obtained using GARCH model with normal (GARCH$_{\text{norm}}$), skewed normal (GARCH$_{\text{snorm}}$), Student-t (GARCH$_{\text{std}}$), skewed Student-t (GARCH$_{\text{sstd}}$), generalized error (GARCH$_{\text{ged}}$), skewed generalized error (GARCH$_{\text{sged}}$), normal inverse Gaussian (GARCH$_{\text{nig}}$), and Johnson SU (GARCH$_{\text{jsu}}$) distributions. In the estimation process, we use a rolling estimation of 250 observations. MR^{WC_m}, $m = 1, \cdots, 13$, are model risk measures estimated with weighted risk forecasting. When model risk measure is estimated using a reference model, which refers to GARCH$_{\text{norm}}$, we call it $MR_0^{WC_m}$, $m = 2, 3, 4, 5, 5, 7, 8, 11, 12, 13$.

Table 4: Descriptive analysis of worst case measures for $\mathrm{ES}^{0.025}$ forecasts. The results are exposed to the whole sample (January 1, 2001, to May 30, 2018) and sub-samples (before crisis, January 1, 2001 until July 30, 2007, crisis, August 1, 2007 until September 28, 2012, and after crisis, September 29, 2012 until May 30, 2018) for S&P500 log-returns (in %).

	Whole Sample			Before Crisis			Crisis			After Crisis		
Measures	Mean	SD	Prop (%)	Mean	SD	Prop (%)	Mean	SD	Prop (%)	Mean	SD	Prop (%)
MR^{WC_1}	3.575	6.496	129.510	5.129	8.210	184.919	2.732	2.093	70.946	3.123	7.628	176.654
$MR_0^{WC_2}$	2.827	5.243	102.405	4.973	8.270	179.267	2.435	2.004	63.246	1.028	0.962	58.169
MR^{WC_2}	2.450	4.585	88.759	4.297	7.222	154.922	2.037	1.710	52.903	1.034	1.216	58.515
$MR_0^{WC_3}$	0.748	4.138	27.104	0.157	0.394	5.652	0.296	0.207	7.701	2.095	7.595	118.485
MR^{WC_3}	1.125	3.625	40.750	0.832	1.033	29.997	0.695	0.408	18.043	2.089	6.604	118.139
$MR_0^{WC_4}$	0.616	0.820	22.312	0.742	1.027	26.732	0.542	0.331	14.075	0.590	0.973	33.369
MR^{WC_4}	0.780	1.398	28.251	1.112	1.789	40.097	0.600	0.421	15.596	0.686	1.622	38.802
$MR_0^{WC_5}$	1.121	1.688	40.608	2.030	2.614	73.191	0.683	0.300	17.747	0.583	0.381	32.982
MR^{WC_5}	0.968	1.479	35.063	1.756	2.287	63.307	0.564	0.241	14.659	0.552	0.408	31.202
$MR_0^{WC_6}$	0.853	0.224	30.899	0.890	0.249	32.080	0.863	0.092	22.409	0.778	0.288	44.021
MR^{WC_6}	0.687	0.151	24.891	0.749	0.173	26.991	0.716	0.062	18.601	0.597	0.157	33.746
$MR_0^{WC_7}$	0.417	0.180	15.112	0.509	0.239	18.350	0.389	0.097	10.107	0.342	0.115	19.338
MR^{WC_7}	0.366	0.202	13.263	0.438	0.204	15.806	0.323	0.076	8.377	0.347	0.269	19.631
$MR_0^{WC_8}$	-0.077	8.621	-2.799	-0.378	14.544	-13.640	0.101	0.060	2.621	0.053	2.671	2.989
MR^{WC_8}	0.141	7.251	5.110	-0.023	12.237	-0.828	0.229	0.060	5.945	0.202	2.231	11.410
MR^{WC_9}	2.094	9.048	75.878	2.690	15.117	96.978	1.843	0.279	47.880	1.677	3.540	94.845
$MR^{WC_{10}}$	12.536	776.507	454.145	0.303	0.248	10.912	0.153	0.033	3.962	43.680	1458.807	2470.739
$MR_0^{WC_{11}}$	0.120	0.531	4.333	0.033	0.064	1.194	0.072	0.070	1.866	0.298	0.970	16.857
$MR^{WC_{11}}$	0.390	0.699	14.125	0.556	0.894	20.049	0.300	0.211	7.798	0.343	0.811	19.401
$MR_0^{WC_{12}}$	0.496	0.670	17.979	0.708	1.041	25.538	0.470	0.318	12.209	0.292	0.262	16.512
$MR^{WC_{12}}$	0.390	0.699	14.125	0.556	0.894	20.049	0.300	0.211	7.798	0.343	0.811	19.401
$MR_0^{WC_{13}}$	1.197	1.761	43.347	1.797	2.758	64.775	1.162	0.806	30.189	0.609	0.514	34.473
$MR^{WC_{13}}$	0.820	1.186	29.701	1.121	1.706	40.430	0.764	0.512	19.847	0.616	0.994	34.819

Note: This table presents results from worst case measures. Measures are applied in $\mathrm{ES}^{0.025}$ forecasts obtained using GARCH model with normal ($\mathrm{GARCH}_{\mathrm{norm}}$), skewed normal ($\mathrm{GARCH}_{\mathrm{snorm}}$), Student-$t$ ($\mathrm{GARCH}_{\mathrm{std}}$), skewed Student-$t$ ($\mathrm{GARCH}_{\mathrm{sstd}}$), generalized error ($\mathrm{GARCH}_{\mathrm{ged}}$), skewed generalized error ($\mathrm{GARCH}_{\mathrm{sged}}$), normal inverse Gaussian ($\mathrm{GARCH}_{\mathrm{nig}}$), and Johnson SU ($\mathrm{GARCH}_{\mathrm{jsu}}$) distributions. In the estimation process, we use a rolling estimation of 250 observations. MR^{WC_m}, $m = 1, \cdots, 13$, are model risk measures estimated with weighted risk forecasting. When model risk measure is estimated using a reference model, which refers to $\mathrm{GARCH}_{\mathrm{norm}}$, we call it $MR_0^{WC_m}$, $m = 2, 3, 4, 5, 5, 7, 8, 11, 12, 13$.

Table 5: Descriptive analysis of loss function measures for VaR$^{0.01}$ forecasts. The results are exposed to the whole sample (January 1, 2001, to May 30, 2018) and sub-samples (before crisis, January 1, 2001 until July 30, 2007, crisis, August 1, 2007 until September 28, 2012, and after crisis, September 29, 2012 until May 30, 2018) for S&P500 log-returns (in %). Results from $MR_i^{LF_4}$ are multiplied by 100.

	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
Whole sample	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$\text{GARCH}_{\text{norm}}$	2.391	1.878	2.414	1.848	0.012	0.131	0.010	0.014	2.402	1.859	2.414	1.848
$\text{GARCH}_{\text{snorm}}$	2.250	1.817	2.281	1.777	0.016	0.146	0.010	0.013	2.266	1.791	2.281	1.777
$\text{GARCH}_{\text{std}}$	3.200	2.266	3.207	2.256	0.004	0.081	0.014	0.020	3.203	2.260	3.207	2.256
$\text{GARCH}_{\text{sstd}}$	2.465	1.873	2.486	1.844	0.011	0.122	0.011	0.014	2.476	1.855	2.487	1.844
$\text{GARCH}_{\text{ged}}$	2.648	1.983	2.663	1.962	0.007	0.107	0.011	0.015	2.656	1.970	2.663	1.963
$\text{GARCH}_{\text{sged}}$	2.483	1.889	2.503	1.8862	0.010	0.118	0.011	0.014	2.493	1.872	2.503	1.862
$\text{GARCH}_{\text{nig}}$	4.626	10.049	5.638	9.518	0.506	3.787	0.024	0.045	5.132	9.025	5.638	9.518
$\text{GARCH}_{\text{jsu}}$	2.451	1.871	2.474	1.840	0.011	0.123	0.011	0.014	2.462	1.851	2.474	1.840
	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
Before Crisis	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$\text{GARCH}_{\text{norm}}$	2.105	1.365	2.115	1.347	0.005	0.076	0.020	0.013	2.110	1.354	2.116	1.348
$\text{GARCH}_{\text{snorm}}$	2.079	1.405	2.091	1.386	0.006	0.081	0.019	0.013	2.085	1.393	2.091	1.386
$\text{GARCH}_{\text{std}}$	2.377	1.481	2.381	1.474	0.002	0.056	0.023	0.017	2.379	1.476	2.381	1.474
$\text{GARCH}_{\text{sstd}}$	2.180	1.439	2.188	1.426	0.004	0.068	0.021	0.014	2.185	1.431	2.189	1.427
$\text{GARCH}_{\text{ged}}$	2.199	1.394	2.201	1.393	0.003	0.065	0.021	0.014	2.203	1.387	2.206	1.383
$\text{GARCH}_{\text{sged}}$	2.158	1.424	2.167	1.411	0.004	0.069	0.020	0.014	2.163	1.416	2.167	1.411
$\text{GARCH}_{\text{nig}}$	8.813	14.043	8.864	14.011	0.026	0.189	0.057	0.064	8.838	14.026	8.864	14.011
$\text{GARCH}_{\text{jsu}}$	2.186	1.446	2.194	1.433	0.004	0.068	0.021	0.014	2.190	1.438	2.194	1.433
	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
Crisis	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$\text{GARCH}_{\text{norm}}$	3.445	2.772	3.483	2.724	0.019	0.197	0.026	0.022	3.464	2.741	3.483	2.724
$\text{GARCH}_{\text{snorm}}$	3.155	2.671	3.205	2.611	0.025	0.216	0.024	0.021	3.180	2.632	3.205	2.611
$\text{GARCH}_{\text{std}}$	4.605	3.091	4.618	3.070	0.007	0.135	0.034	0.027	4.612	3.077	4.619	3.069
$\text{GARCH}_{\text{sstd}}$	3.440	2.717	3.476	2.671	0.018	0.186	0.026	0.022	3.458	2.688	3.476	2.671
$\text{GARCH}_{\text{ged}}$	3.833	2.871	3.858	2.837	0.013	0.164	0.028	0.024	3.846	2.849	3.858	2.837
$\text{GARCH}_{\text{sged}}$	3.492	2.729	3.526	2.685	0.017	0.182	0.026	0.022	3.509	2.701	3.526	2.685
$\text{GARCH}_{\text{nig}}$	4.993	4.477	5.006	4.462	0.007	0.105	0.037	0.034	4.999	4.468	5.006	4.462
$\text{GARCH}_{\text{jsu}}$	3.388	2.709	3.426	2.661	0.019	0.187	0.025	0.022	3.407	2.679	3.426	2.661
	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
After Crisis	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$\text{GARCH}_{\text{norm}}$	1.804	1.104	1.829	1.062	0.013	0.126	0.015	0.014	1.816	1.076	1.829	1.062
$\text{GARCH}_{\text{snorm}}$	1.677	1.082	1.710	1.028	0.017	0.138	0.014	0.013	1.693	1.046	1.710	1.028
$\text{GARCH}_{\text{std}}$	2.673	1.561	2.680	1.549	0.004	0.056	0.024	0.027	2.677	1.554	2.680	1.549
$\text{GARCH}_{\text{sstd}}$	1.912	1.171	1.939	1.127	0.013	0.117	0.017	0.016	1.925	1.143	1.939	1.127
$\text{GARCH}_{\text{ged}}$	2.045	1.201	2.063	1.169	0.009	0.104	0.017	0.016	2.054	1.181	2.063	1.169
$\text{GARCH}_{\text{sged}}$	1.933	1.218	1.958	1.178	0.012	0.109	0.017	0.017	1.946	1.193	1.958	1.178
$\text{GARCH}_{\text{nig}}$	0.079	7.509	3.565	6.608	1.743	6.960	0.023	0.028	1.822	1.259	3.565	6.608
$\text{GARCH}_{\text{jsu}}$	1.909	1.206	1.938	1.160	0.014	0.117	0.017	0.017	1.923	1.177	1.937	1.160

Note: This table presents EL (average) and standard deviation (SD) of model risk obtained using measures belong loss function approach considering as $\rho = EL$. Measures are applied in VaR$^{0.01}$ forecasts estimated using GARCH model with normal ($\text{GARCH}_{\text{norm}}$), skewed normal ($\text{GARCH}_{\text{snorm}}$), Student-$t$ ($\text{GARCH}_{\text{std}}$), skewed Student-$t$ ($\text{GARCH}_{\text{sstd}}$), generalized error ($\text{GARCH}_{\text{ged}}$), skewed generalized error ($\text{GARCH}_{\text{sged}}$), normal inverse Gaussian ($\text{GARCH}_{\text{nig}}$), and Johnson SU ($\text{GARCH}_{\text{jsu}}$) distributions. The rolling estimation window is of 250 observations.

Table 6: Descriptive analysis loss function measures for $ES^{0.025}$ forecasts. The results are exposed to the whole sample (January 1, 2001, to May 30, 2018) and sub-samples (before crisis, January 1, 2001 until July 30, 2007, crisis, August 1, 2007 until September 28, 2012, and after crisis, September 29, 2012 until May 30, 2018) for S&P500 log-returns (in %). Results from $MR_i^{LF_4}$ are multiplied by 100.

	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
Whole sample	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$GARCH_{norm}$	2.402	1.884	2.425	1.854	0.011	0.130	0.010	0.014	2.414	1.864	2.425	1.854
$GARCH_{snorm}$	2.260	1.822	2.291	1.782	0.016	0.145	0.010	0.013	2.275	1.797	2.291	1.783
$GARCH_{std}$	3.325	2.354	3.331	2.344	0.003	0.077	0.014	0.022	3.328	2.348	3.331	2.345
$GARCH_{sstd}$	2.541	1.899	2.560	1.873	0.009	0.114	0.011	0.015	2.550	1.883	2.560	1.873
$GARCH_{ged}$	2.681	1.997	2.695	1.997	0.007	0.105	0.011	0.016	2.688	1.985	2.695	1.978
$GARCH_{sged}$	2.512	1.902	2.532	1.876	0.010	0.115	0.011	0.015	2.522	1.885	2.532	1.876
$GARCH_{nig}$	4.005	7.525	5.018	6.890	0.507	3.795	0.021	0.034	4.512	6.136	5.018	6.890
$GARCH_{jsu}$	2.506	1.891	2.526	1.862	0.010	0.117	0.011	0.015	2.516	1.873	2.527	1.863
	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
Before Crisis	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$GARCH_{norm}$	2.115	1.368	2.126	1.351	0.005	0.075	0.020	0.013	2.120	1.358	2.126	1.351
$GARCH_{snorm}$	2.089	1.409	2.101	1.390	0.006	0.080	0.019	0.013	2.095	1.397	2.101	1.390
$GARCH_{std}$	2.419	1.502	2.423	1.496	0.002	0.054	0.024	0.018	2.421	1.498	2.423	1.496
$GARCH_{sstd}$	2.215	1.450	2.222	1.438	0.004	0.066	0.021	0.015	2.219	1.443	2.222	1.439
$GARCH_{ged}$	2.217	1.399	2.223	1.389	0.003	0.064	0.021	0.014	2.220	1.393	2.223	1.389
$GARCH_{sged}$	2.175	1.429	2.185	1.416	0.004	0.068	0.020	0.014	2.179	1.421	2.183	1.417
$GARCH_{nig}$	6.877	8.977	6.926	8.940	0.025	0.185	0.049	0.043	6.902	8.956	6.926	8.939
$GARCH_{jsu}$	2.214	1.457	2.222	1.445	0.004	0.066	0.021	0.014	2.218	1.450	2.222	1.445
	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
Crisis	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$GARCH_{norm}$	3.461	2.780	3.499	2.732	0.019	0.195	0.026	0.022	3.480	2.750	3.499	2.733
$GARCH_{snorm}$	3.169	2.678	3.217	2.618	0.025	0.214	0.024	0.021	3.193	2.640	3.218	2.619
$GARCH_{std}$	4.770	3.154	4.783	3.154	0.007	0.131	0.035	0.028	4.777	3.141	4.783	3.134
$GARCH_{sstd}$	3.532	2.740	3.565	2.697	0.016	0.178	0.026	0.022	3.549	2.713	3.565	2.698
$GARCH_{ged}$	3.880	2.887	3.903	2.854	0.012	0.161	0.029	0.024	3.892	2.866	3.903	2.855
$GARCH_{sged}$	3.532	2.744	3.564	2.702	0.016	0.178	0.026	0.022	3.548	2.717	3.564	2.702
$GARCH_{nig}$	5.078	4.525	5.090	4.510	0.006	0.099	0.037	0.034	5.084	4.517	5.090	4.511
$GARCH_{jsu}$	3.454	2.728	3.490	2.682	0.018	0.181	0.026	0.022	3.472	2.699	3.490	2.683
	$MR_i^{LF_1}$		$MR_i^{LF_2}$		$MR_i^{LF_3}$		$MR_i^{LF_4}$		$MR_i^{LF_5}$		$MR_i^{LF_6}$	
After Crisis	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD	EL	SD
$GARCH_{norm}$	1.812	1.106	1.837	1.064	0.012	0.125	0.015	0.014	1.825	1.078	1.837	1.065
$GARCH_{snorm}$	1.683	1.084	1.717	1.084	0.017	0.137	0.014	0.013	1.700	1.049	1.717	1.031
$GARCH_{std}$	2.817	1.694	2.822	1.684	0.003	0.047	0.026	0.031	2.820	1.689	2.822	1.685
$GARCH_{sstd}$	2.005	1.223	2.028	1.184	0.012	0.105	0.018	0.018	2.017	1.199	2.028	1.184
$GARCH_{ged}$	2.077	1.214	2.094	1.183	0.009	0.101	0.018	0.017	2.086	1.195	2.095	1.184
$GARCH_{sged}$	1.963	1.233	1.986	1.9453	0.012	0.107	0.017	0.017	1.975	1.209	1.987	1.195
$GARCH_{nig}$	0.117	7.538	3.614	6.615	1.748	6.976	0.023	0.028	1.865	1.280	3.614	6.616
$GARCH_{jsu}$	1.974	1.245	1.999	1.203	0.013	0.110	0.018	0.019	1.987	1.220	2.000	1.203

This table presents EL (average) and standard deviation (SD) of model risk obtained using measures belong loss function approach considering as $\rho = EL$. Measures are applied in $ES^{0.025}$ forecasts estimated using GARCH model with normal ($GARCH_{norm}$), skewed normal ($GARCH_{snorm}$), Student-t ($GARCH_{std}$), skewed Student-t ($GARCH_{sstd}$), generalized error ($GARCH_{ged}$), skewed generalized error ($GARCH_{sged}$), normal inverse Gaussian ($GARCH_{nig}$), and Johnson SU ($GARCH_{jsu}$) distributions. The rolling estimation window is of 250 observations.

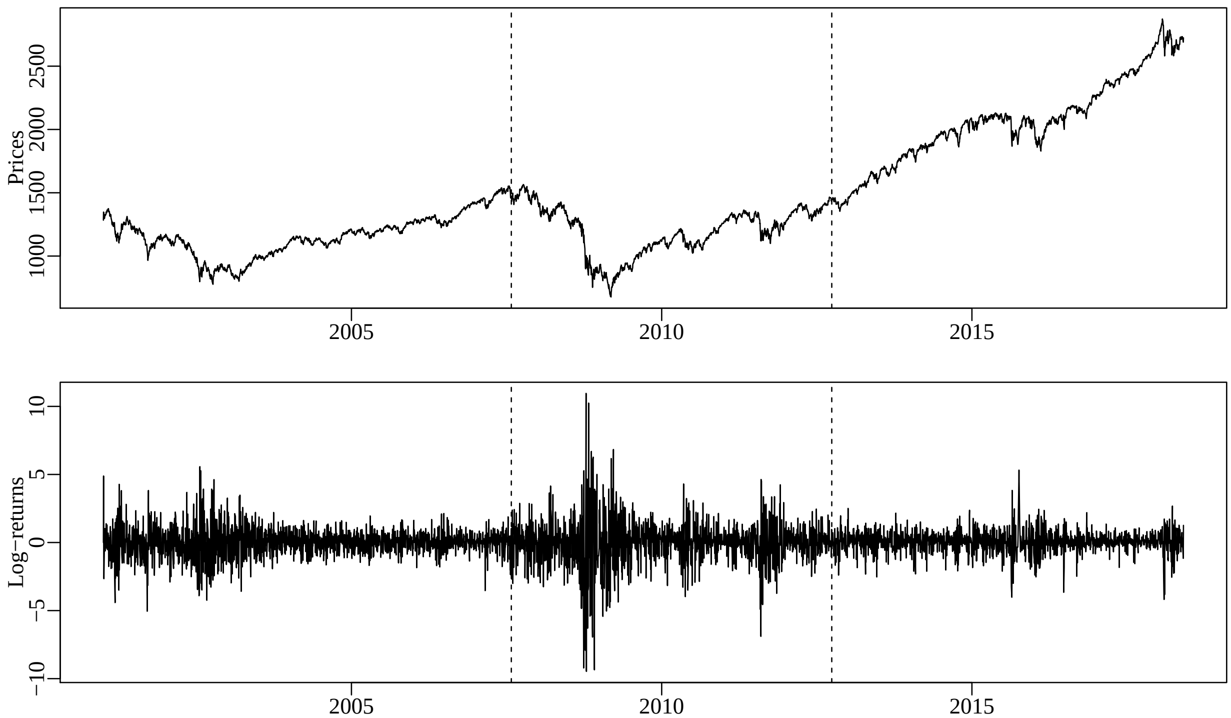

Figure 1: Daily observations from January 1, 2001, to May 30, 2018, of the S&P500 adjusted closing price (Prices) and log-returns in % (Log-returns). The vertical lines represent the subdivision of sample into non-crisis and crisis.

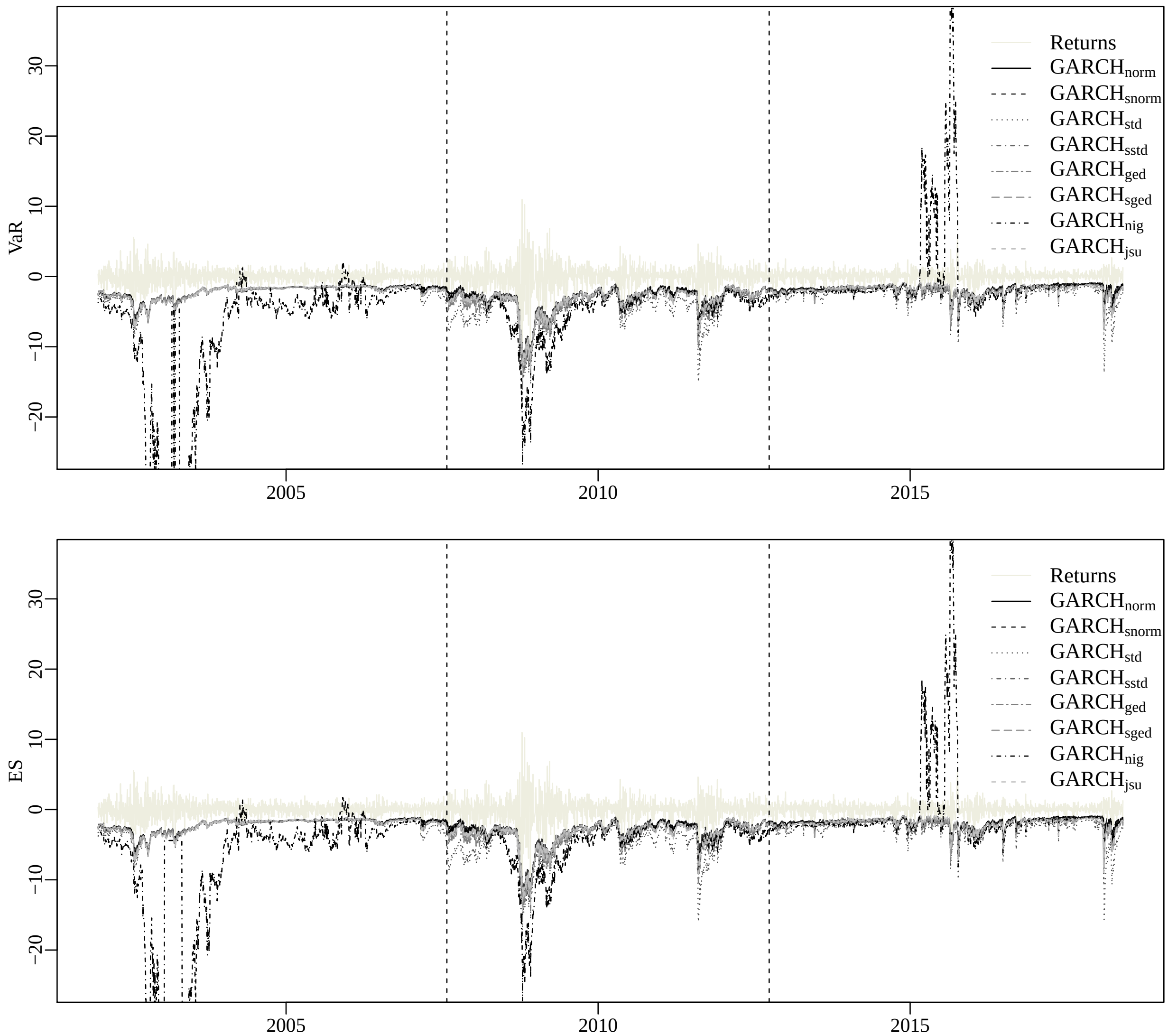

Figure 2: S&P500 log-returns (in %) and $\text{VaR}^{0.01}$ and $\text{ES}^{0.025}$ forecasts, with the corrected signal, considering the whole sample, which refers to January 1, 2001, to May 30, 2018. Thus, forecasts comprehend January 1, 2002 to May 30, 2018. The vertical lines represent the subdivision of sample into non-crisis and crisis..

Note: This figure exposes S&P500 log-returns and $\text{VaR}^{0.01}$ and $\text{ES}^{0.025}$ forecasts, with the corrected signal, considering a rolling estimation window of 250 observations. We estimate risk forecasting using GARCH model with normal ($\text{GARCH}_{\text{norm}}$), skewed normal ($\text{GARCH}_{\text{snorm}}$), Student-$t$ ($\text{GARCH}_{\text{std}}$), skewed Student-$t$ ($\text{GARCH}_{\text{sstd}}$), generalized error ($\text{GARCH}_{\text{ged}}$), skewed generalized error ($\text{GARCH}_{\text{sged}}$), normal inverse Gaussian ($\text{GARCH}_{\text{nig}}$), and Johnson SU ($\text{GARCH}_{\text{jsu}}$) distributions.

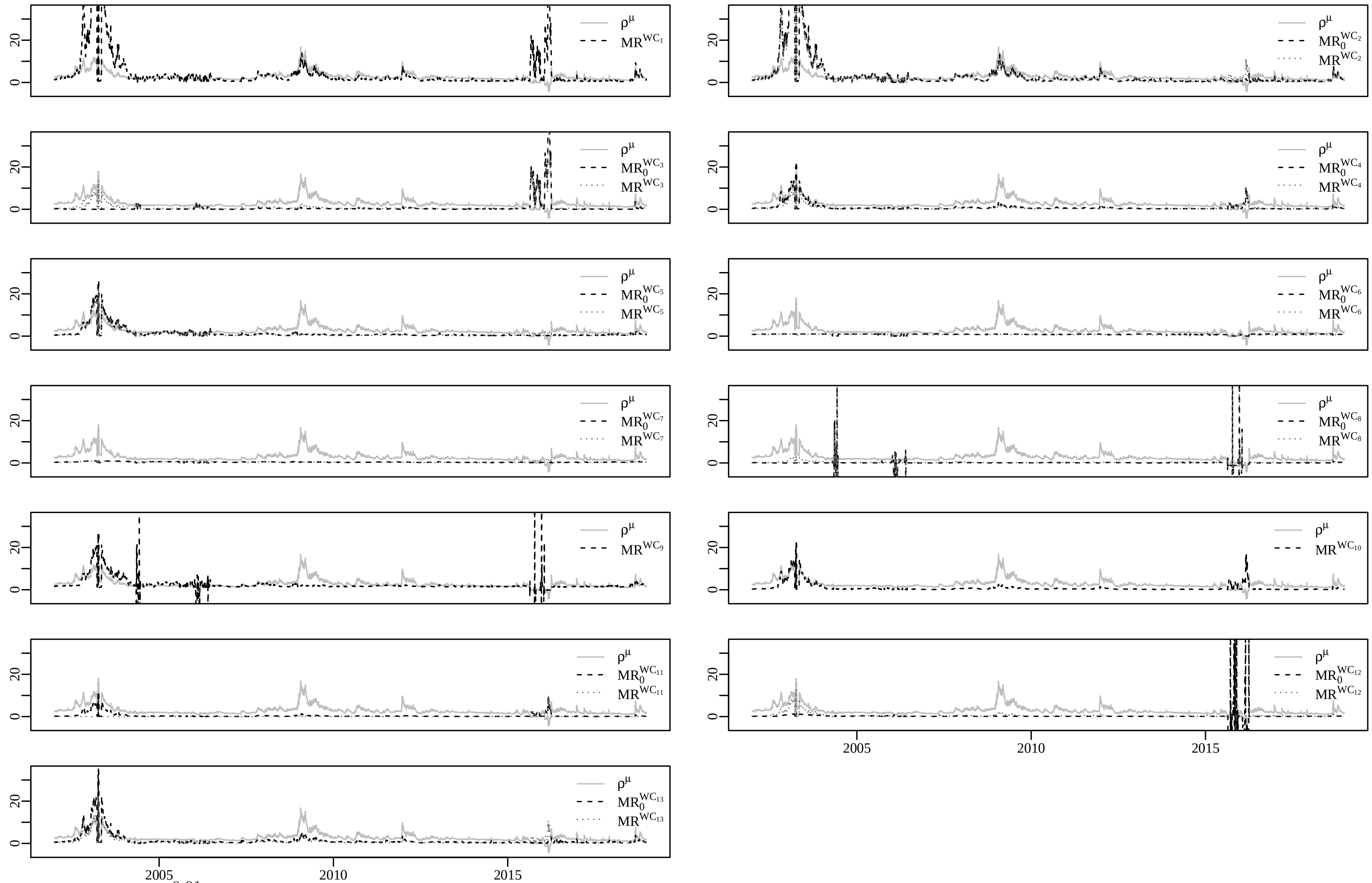

Figure 3: Weighted VaR$^{0.01}$ forecasting (gray line) and model risk measures for S&P500 log-returns (in %) considering the whole sample, which refers to January 1, 2001, to May 30, 2018.

Note: This figure shows the evolution of weighted VaR$^{0.01}$ forecasting (gray line) (ρ_t^{μ}) and model risk considering a rolling estimation window of 250 observations. We compute ρ_t^{μ} by equally weighting the risk forecasting. We estimate VaR$^{0.01}$ using GARCH model with normal (GARCH$_{\text{norm}}$), skewed normal (GARCH$_{\text{snorm}}$), Student-t (GARCH$_{\text{std}}$), skewed Student-t (GARCH$_{\text{sstd}}$), generalized error (GARCH$_{\text{ged}}$), skewed generalized error (GARCH$_{\text{sged}}$), normal inverse Gaussian (GARCH$_{\text{nig}}$), and Johnson SU (GARCH$_{\text{jsu}}$) distributions. MR^{WC_m}, $m = 1, \cdots, 13$, are model risk measures, which are estimated using weighted risk forecasting. When model risk measure is estimated using a reference model, which refers to GARCH model with normal distribution, we call it $MR_0^{WC_m}$, $m = 2, 3, 4, 5, 5, 7, 8, 11, 12, 13$.

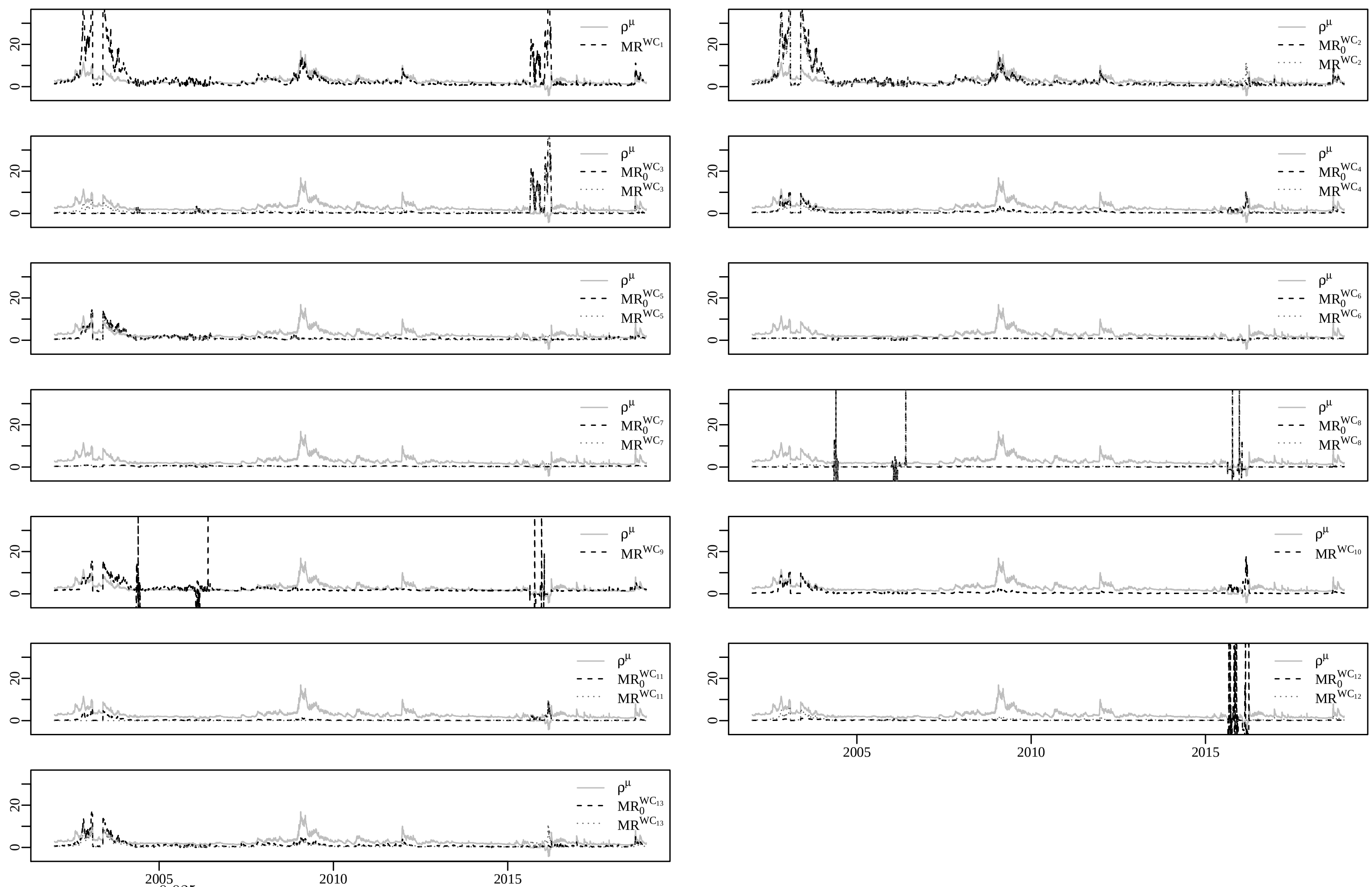

Figure 4: Weighted $\text{ES}^{0.025}$ forecasting (gray line) and model risk measures for S&P500 log-returns (in %) considering the whole sample, which refers to January 1, 2001, to May 30, 2018.

Note: This figure shows the evolution of weighted $\text{ES}^{0.025}$ forecasting (gray line) (ρ_t^{μ}) and model risk considering a rolling estimation window of 250 observations. We compute ρ_t^{μ} by equally weighting the risk forecasting. We estimate $\text{ES}^{0.025}$ using GARCH model with normal ($\text{GARCH}_{\text{norm}}$), skewed normal ($\text{GARCH}_{\text{snorm}}$), Student-$t$ ($\text{GARCH}_{\text{std}}$), skewed Student-$t$ ($\text{GARCH}_{\text{sstd}}$), generalized error ($\text{GARCH}_{\text{ged}}$), skewed generalized error ($\text{GARCH}_{\text{sged}}$), normal inverse Gaussian ($\text{GARCH}_{\text{nig}}$), and Johnson SU ($\text{GARCH}_{\text{jsu}}$) distributions. MR^{WCm}, $m = 1, \cdots, 13$, are model risk measures, which are estimated using weighted risk forecasting. When model risk measure is estimated using a reference model, which refers to GARCH model with normal distribution, we call it MR_0^{WCm}, $m = 2, 3, 4, 5, 5, 7, 8, 11, 12, 13$.

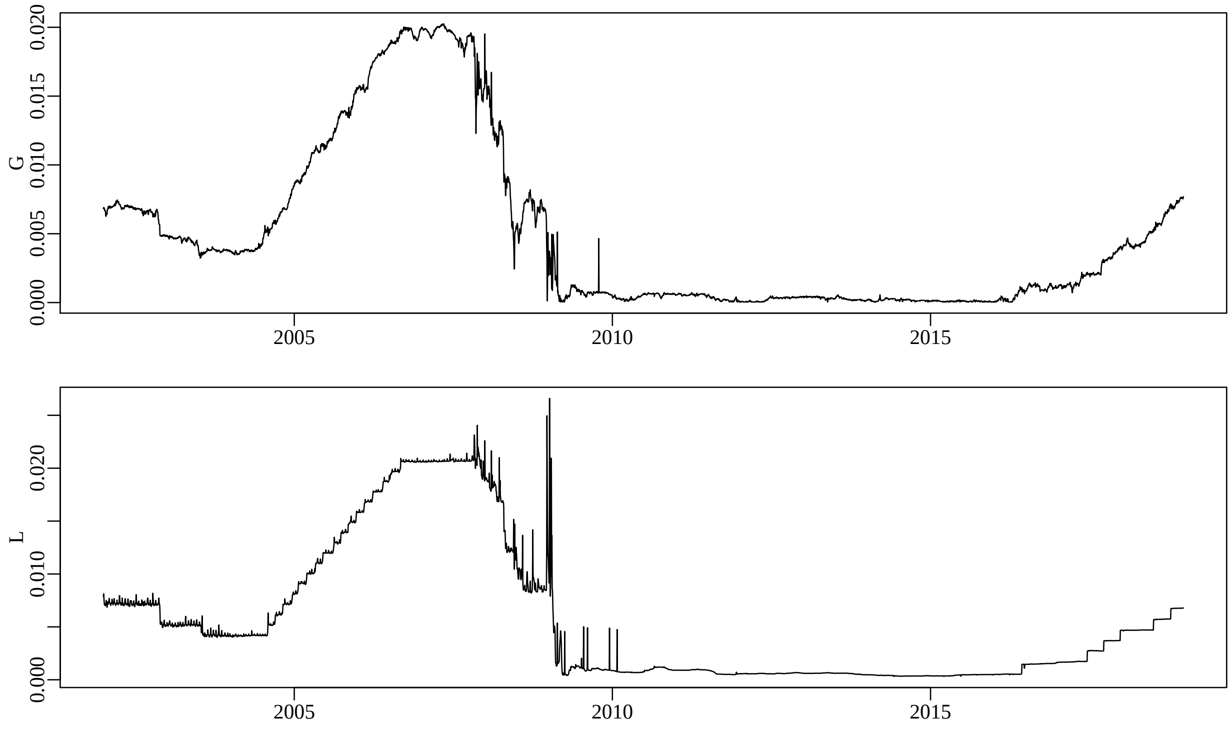

Figure 5: Daily observations (in %) from January 1, 2002, to May 30, 2018, for the annual three months maturity U.S. Treasury Bill yield (G), and yearly U.S. Dollar based Overnight London Interbank Offered Rate (L). Both yield rates are in daily frequency.

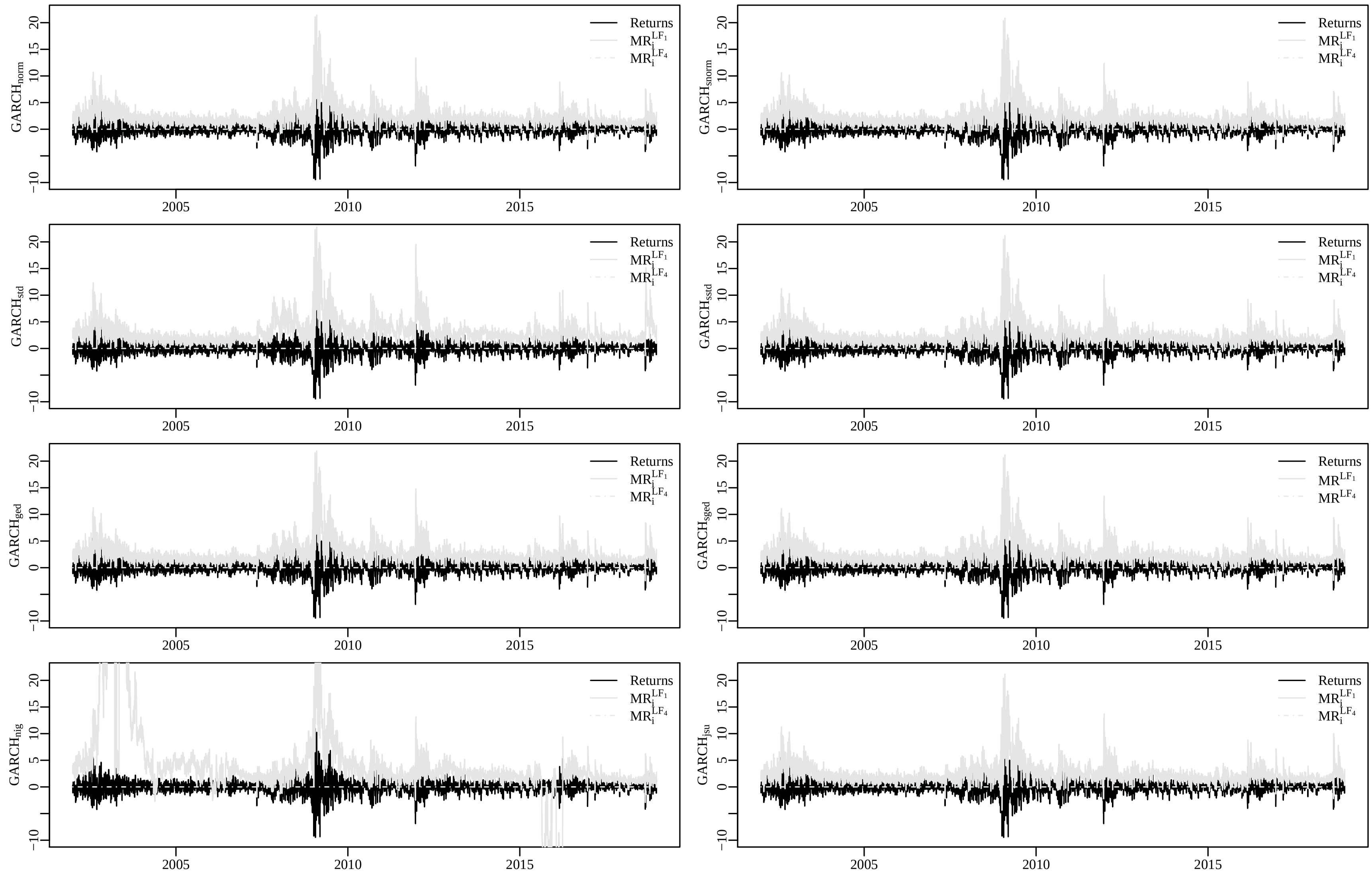

Figure 6: S&P500 log-returns (in %) and loss function measures for $\text{VaR}^{0.01}$ forecasting. These results are exposed to the whole sample, which refers to January 1, 2001, to May 30, 2018

Note: This figure shows the evolution of S&P500 log-returns and loss function measures. We consider MR_i^{LF1} and MR_i^{LF4}.We estimate risk using GARCH model with normal ($\text{GARCH}_{\text{norm}}$), skewed normal ($\text{GARCH}_{\text{snorm}}$), Student-$t$ ($\text{GARCH}_{\text{std}}$), skewed Student-$t$ ($\text{GARCH}_{\text{sstd}}$), generalized error ($\text{GARCH}_{\text{ged}}$), skewed generalized error ($\text{GARCH}_{\text{sged}}$), normal inverse Gaussian ($\text{GARCH}_{\text{nig}}$), and Johnson SU ($\text{GARCH}_{\text{jsu}}$) distributions. In the estimation process, the rolling estimation window is of 250 observations.

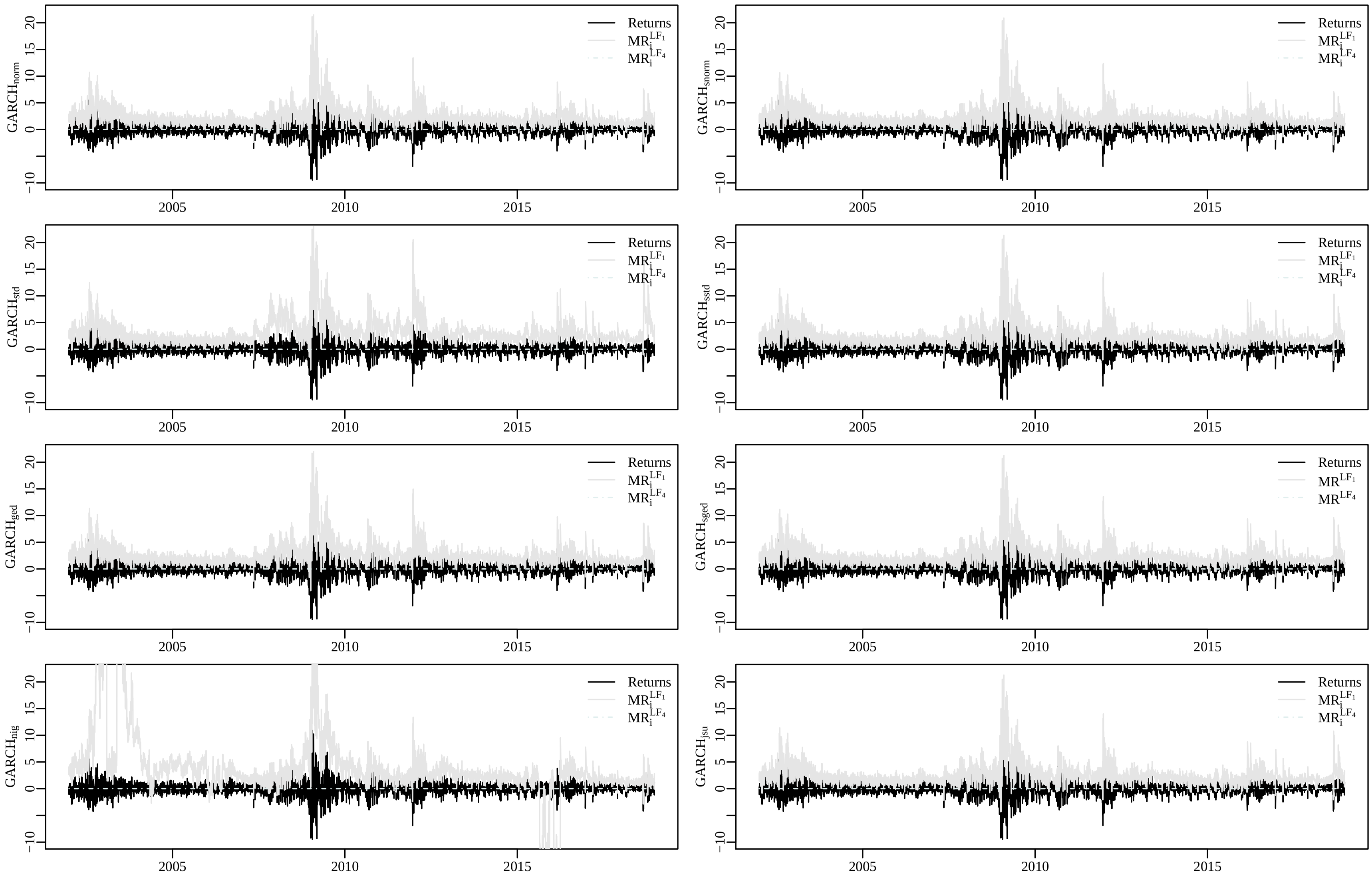

Figure 7: S&P500 log-returns (in %) and loss function measures for $ES^{0.025}$ forecasting. These results are exposed to the whole sample, which refers to January 1, 2001 to May 30, 2018.

Note: This figure shows the evolution of S&P500 log-returns and loss function measures. We consider $MR_i^{LF_1}$ and $MR_i^{LF_4}$. We estimate risk using GARCH model with normal (GARCH$_{\text{norm}}$), skewed normal (GARCH$_{\text{snorm}}$), Student-t (GARCH$_{\text{std}}$), skewed Student-t (GARCH$_{\text{sstd}}$), generalized error (GARCH$_{\text{ged}}$), skewed generalized error (GARCH$_{\text{sged}}$), normal inverse Gaussian (GARCH$_{\text{nig}}$), and Johnson SU (GARCH$_{\text{jsu}}$) distributions. In the estimation process, the rolling estimation window is of 250 observations.