ENERGY RISK MANAGEMENT BY VALUE-AT-RISK

Angelica Gianfreda[100]
Giacomo Scandolo[101]
DOI: https://doi.org/10.31410/EMAN.2018.280

Abstract: *Model risk has an important effect on risk measurements. Indeed, the choice of the underlying probabilistic model can have a significant impact on the risk forecast. The hazard of producing poor risk assessments due to the choice of an unsuited model is known as "model risk". Its detection and quantification are crucial tasks, particularly with energy commodities which require more complex modelling compared to the ones needed in traditional financial markets. Using a normalized measure of model risk for the forecast of daily Value-at-Risk, we focus on a restricted set of plausible models within the GARCH-type class specified with nine different distributions. In this way, we are able to provide a more reliable assessment of model risk for two energy commodities (natural gas and crude oil) over the years from 2001 to 2015.*

Key words: *Relative Measure of Model Risk, VaR, GARCH models, One-step ahead Forecasting, Natural Gas, Crude Oil*

1. INTRODUCTION

The hazard of producing poor risk assessments due to the choice of an unsuited model is known as "model risk". It has important effects on risk measurements according to the choice of the underlying probabilistic models. Therefore, the detection and the quantification of model risk are crucial tasks, particularly with energy commodities which require more complex modelling compared to the one needed in traditional financial markets.

The literature on estimation risk in the context of Value-at-Risk (VaR) and Expected Shortfall (ES) forecasting is well developed. On the contrary, quantifying and managing misspecification risk (due to the wrong distributional assumption) has been less investigated.

Using a normalized measure of model risk for the forecasted daily values of VaR and using the well-established GARCH-type models under different distributional assumptions, we are able to provide a reliable assessment of model risk for two energy commodities (natural gas and crude oil) over more than 10 years.

We obtain a more intrinsic assessment and, more importantly, we never significantly over- or under-estimate risk when averages of the estimates are used. Our empirical results emphasize that the distributional assumptions made in price modelling can produce a relevant discrepancy in risk figures and then trigger substantial model risk.

We also provide empirical evidence that the amount of model risk associated to a given model crucially depends on the estimated risk measure. In particular, under normality we observe more model risk at the more extreme quantiles. Instead, the model risk associated to the daily best distribution is more stable across quantile levels.

[100] Free University of Bozen/Bolzano, Universitätsplatz 1, (39100) Bozen (Italy)
[101] University of Florence, Via delle Pandette 32, (50127) Firenze (Italy)

From a practical point of view, implementing sophisticated models can be time consuming and costly (for instance, in terms of IT equipment, and estimation times). However, our measurement of model risk, even if performed for a limited amount of time and/or from time to time, would provide risk managers with a useful figure that could be complementary to the usual back-testing procedures for VaR. For instance, if model risk under the model which is currently implemented has been consistently high in the recent period, then it is likely that VaR will be underestimated in the subsequent days, if we keep using that model. In this case, proper actions can be taken, such as incrementing the VaR figure by an amount that depends on model risk itself, or preferring an alternative model with lower model risk.

2. METHODOLOGY

We recall that, for a given confidence level *p* the daily Value-at-Risk (VaR) of a portfolio is the amount such that the probability that the daily portfolio loss will exceed it is exactly *p* (here, we assume that the distribution of losses is continuous). We consider the values 1% and 5% for *p*, which are typical in market risk management.

The quantification of VaR obviously depends on the employed probabilistic model. Consequently, competing models may in principle lead to very different VaR forecasts. In this context, model risk arises when the forecasts produced under a set of plausible models are highly divergent. Model risk may be captured by several quantitative measures that have been recently proposed in the literature. We will consider the Relative Measure of Model Risk (henceforth, RMMR) introduced by Barrieu and Scandolo in [1]. At a certain day and for a fixed confidence level *p*, denote by VaR_i the VaR forecast under model *i*. Next, let VaR_0 be the forecast under a fixed *reference* model. Then, the RMMR is defined as the number

$$RMMR = \frac{max_i \, VaR_i - VaR_0}{max_i \, VaR_i - min_i \, VaR_i} \quad (1)$$

If the reference model lies among the set of competing models, then clearly RMMR is in the interval [0,1]. In this case, the closer is RMMR to 1, the lower is VaR_0 with respect to the other competing forecasts and therefore the higher is the amount of model risk involved, as risk may be highly underestimated. As VaR is a positively homogeneous risk metrics, it turns out that RMMR is insensitive to the amount invested in the portfolio; hence, we may focus on returns (alternatively, we may assume that 1 USD is invested in each portfolio). For further properties and discussion about this measure, we refer to [1].

In this work, we consider two portfolios investing in energy-related assets and specifically the Brent crude oil and the ICE UK natural gas. In both cases, we model the (log-)return series using an AR(5)-GARCH(1,1) process. Specifically, if R_t is the return observed at date *t*, we have

$$R_t = \mu_t + \sigma_t Z_t \quad (2)$$

where

$$\mu_t = \overline{\mu} + \sum_{i=1}^{5} \varphi_i R_{t-i} \quad (3)$$

is the conditional mean following an AR(5) process, and

$$\sigma_t^2 = \omega + \alpha (R_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2 \tag{4}$$

is the conditional variance following a GARCH(1,1) model. In this specification, $\bar{\mu}, \varphi_i, \omega, \alpha$ and β are constants that must satisfy known constraints and the Z_t are IID innovations that are assumed to follow any of 9 competing standard distributions. These are the followings: *normal* (NORM) and *skew normal* (SNORM); *Student-t* (STD) and *skew Student-t* (SSTD); *Generalized Error Distribution* (GED) and *skew GED* (SGED); *Johnson's SU family* (JSU); *Normal Inverse Gaussian* (NIG); *Generalized Hyperbolic family* (GH). For a reminder on these distributions and their parametrization, see [3].

Notice that some of the innovations distributions allow for extra-parameters that give control on asymmetry and/or tail behavior. The AR-GARCH specification, coupled with the choice of the innovation distribution, gives a flexible framework to model returns and it is able to capture known features such as heteroschedasticity.

Having fixed a distribution for the innovations, day-by-day we use a rolling window of 256 past daily returns in order to estimate (by Maximum Likelihood) the parameters in the AR-GARCH model, including possible extra-parameters for the distribution of *Z*. This allows us to compute, the conditional mean and variance and, ultimately, the VaR forecast, given by

$$VaR(R_t) = VaR(\mu_t + \sigma_t Z_t) = -\mu_t + \sigma_t VaR(Z_t) \tag{5}$$

This is repeated for all 9 competing models, hence producing, at any date, a set VaR_i (i=1,...,9) of forecasts. Once a reference model is fixed throughout (NORM, for instance), the final output is a daily series of model risk measures, computed as in (1).

In order to improve the approach, we need to consider that not all 9 models perform equally well. In order to pick the *best model* and/or to discard the *worst* ones, we proceed as follows Day-by-day and for each competing model, as an output of the estimation process, we record the maximized log-likelihood λ_i. Then, we are able to compute the Bayesian Information Criterion (BIC), defined as

$$BIC_i = -2\lambda_i + p_i \ln(n) \tag{6}$$

where n is the length of the dataset (n=256 in our analysis) and p_i is the number of parameters in model i. The BIC is a classical measure of fitting ability (the lower is BIC, the better is the fitting) which penalizes over-parametrized models.

We make a twofold use of BIC. First, on a daily basis, we rank the 9 models according to their fitting ability, as measured by BIC, and pick the *daily best* model, as the one providing the lowest BIC value. The daily best value is then used as the reference model in formula (1) for that specific day.

Second, we can build a system of *percentage weights*, one for each model, defined as

$$w_i = \frac{a_i^2}{\sum_{j=1}^{9} a_j^2} \tag{7}$$

where

$$a_i = \frac{max_j BIC_j - BIC_i}{max_j BIC_j - min_j BIC_j} \tag{8}$$

A higher fitting ability is therefore associated with higher weights, since w_i is decreasing in BIC_i. After ranking the models with decreasing weights, we retain all models until the cumulative weight 0.95 is reached, and we compute RMMR as in (1) accordingly. This simple *trimming* procedure allows us to discard, on a daily basis, the few worst fitting models, thus making our approach more sound and robust. See [2] for further discussion on this point and on the choice 0.95 as the threshold level. Sometime it may happen that the reference model is no more among the competing ones. In this case, RMMR is no more confined to the interval [0,1]; still, high values of RMMR (even above 1) signal a danger of severe risk underestimation.

Finally, we use the weights in order to compute an average forecast, naturally defined as

$$VaR_{avg} = \sum_i w_i VaR_i \tag{9}$$

to be used as a possible reference estimate at a given date, i.e. replacing VaR_0 in (1). In (9), discarded models are left out of the average and weights are therefore properly normalized.

3. RESULTS

We have implemented the above procedure for two portfolios investing, separately, in Oil and Gas. At each date, we computed the measure of model risk RMMR as in (1), by considering as reference model either one specific distribution (e.g. NORM, STD, etc), or the daily best, or the average forecast VaR_{avg}. We always discarded the worst fitting models through the procedure presented above (on average, 2 or 3 models are discarded each day). The entire procedure is repeated for the values 1% and 5% for the VaR confidence level p.

Time series of Oil and Gas prices have been retrieved from Datastream, from 01/01/2001 to 31/12/2015, and are quoted on a basis of 5 days per week, for a total of 3914 observations. We employ the R-package *rugarch* for estimating GARCH models and computing BIC values. Some empirical findings are shown next.

Angelica Gianfreda

She is assistant professor at the Faculty of Economics and Management, Free University of Bozen/Bolzano. Previously she was senior fellow at the University of Milano-Bicocca, Marie-Curie Intra-European Fellow at the London Business School and Jean Monnet Fellow at the European University Institute. Her research interests are focused on energy markets and specifically on electricity prices. She has published in journals like Energy Economics, Energy Policy, and The Energy Journal.

First, we give an example regarding the weights construction and the corresponding trimming procedure. On day 22nd January 2002, and considering the Gas series, the weights (as in (7), and already in percentage terms) associated to the 9 models are given in the following Table 1.

Model	GED	GH	JSU	NIG	NORM	SGED	SNORM	SSTD	STD
Weight	23.88	0.00	13.24	16.49	0.00	12.39	0.51	17.90	15.61

Table 1: Model Weights computed as in eq. 7 expressed in percentages

The trimming procedure presented before allows us to discard 3 models (precisely: GH, NORM and SNORM) and to build RMMR with the remaining 6. We remind that the procedure is repeated daily, so that the 3 models are discarded just for the day 22nd January 2002.

The basic choice for the reference model is NORM: this corresponds to the assumption that returns follow a normal GARCH process. It is however well known that normality is often a poor model for innovations. We then search for an "overall best" model for each of the two assets. This is defined, in an admittedly rough way, as the model being the best daily model (i.e. displaying the lowest BIC) in the majority of days. Overall best models, together with "overall worst" models (being the best daily models in the fewest number of days), are shown in Table 2.

Asset	**Best model**	**Worst model**
Oil	STD	GH
Gas	GED	GH

Table 2: Overall Best and Worst Models

We can see that the Generalized Hyperbolic (GH) model, which is specified in terms of 3 extra-parameters, is highly penalized by BIC and turns out to be the worst model for both assets.

In Table 3 we provide some descriptive statistics of the RMMR series (for VaR at confidence level 1%) for different choices of the reference model. We stress that the overall best/worst models are fixed throughout for any given asset (see the previous Table 2); on the contrary, the daily best model may change on a daily base.

reference model	**statistics**	**Oil**	**Gas**
Normal	mean	1.04	0.54
	std. deviation	0.52	0.74
	maximum	5.95	4.75
Overall best	mean	0.52	0.64
	std. deviation	0.32	0.33
	maximum	2.28	3.44
Overall worst	mean	1.59	1.31
	std. deviation	0.46	0.44
	maximum	5.99	5.55
Daily best	mean	0.31	0.20

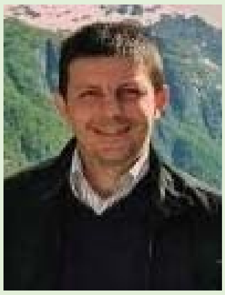
	std. deviation	0.30	0.28
	maximum	1.00	1.00
Average forecast	mean	0.48	0.49
	std. deviation	0.11	0.13
	maximum	0.87	0.88

Table 3: Descriptive statistics of the RMMR series

First, we can notice that using the daily best model provides, on average, the lowest value of model risk. This fact is not surprising, but it is not a simple consequence of the definition of best model. Indeed, being a good model has to do with the fitting ability, as measured by the BIC index; on the contrary, displaying a low model risk has to do with the *conservativeness* of estimates with respect to competing models. As a matter of fact and looking also at the mean levels associated to the overall best and overall worst models, the two features seem to be associated.

Second, by looking at the maximum levels, we can see that using a fixed model can be dangerous in terms of model risk. This is the case not only for the Normal model, which displays RMMR up to nearly 6, but also, a bit surprisingly, for the Overall Best model, for which RMMR can be more than 3.

Giacomo Scandolo



He is associate professor at the University of Florence. Previously he was assistant professor at the University of Verona. He his fellow at the Department of Statistics of the London School of Economics and Political Sciences. He has been Erasmus teaching fellow at the University of Oslo and visiting scholar at Columbia University. His research interests are focused on risk measures, Monte Carlo Methods and applications. He has published in journals as Mathematical Finance, Finance and Stochastics and Quantitative Finance.

Finally, we can see that the average forecast, as expected, gives a level of model risk which is quite stable around 0.5. However, on some days the RMMR almost reached the value 0.9; this level, in our opinion, is extremely high, since discarding the worst models and averaging out the estimates should in principle lead to a great reduction of model risk (toward the value 0.5).

REFERENCES

[1] Barrieu, P., Scandolo, G. (2015) Assessing financial model risk. *European Journal of Operational Research*, 242, 546-556.

[2] Burnham, K.P., Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag.

[3] Rigby, R.A., Stasinopoulos, D.M. (2005) Generalised Additive Models for location, scale, and shape. *Journal of the Royal Statistical Society: Series C*, 54, 507554.