Forecasting Value-at-Risk for Model Risk Analysis in Energy Markets

Angelica Gianfreda and Giacomo Scandolo

Abstract We consider the assessment of mis-specification risk when forecasting Value-at-Risk on a daily horizon. In particular, we focus on Energy Markets (electricity, oil, gas), where the impact of model risk may be relevant. Within an AR-GARCH framework to capture known features of volatility, we consider nine competing distributions for the standardized innovations and we apply a recently proposed measure of model risk to quantify the amount of model uncertainty in the procedure. Our approach is made more robust by discarding, on a daily basis, the worst performing models by using a set of weights built upon the Bayesian Information Criterion. The analysis covers the period 2001-2015, allowing for an in-depth assessment of the dynamics of model risk.

Key words: Building Weights, Risk Management, Electricity, Natural Gas, Brent Crude Oil

1 Introduction

In the financial literature, it has been recognized that the choice of the underlying probabilistic model for the risk factors can have a significant impact on a risk forecast. The hazard of producing a poor risk assessment due to the choice of an unsuited model is usually termed *model risk*. A distinction is usually made between two aspects of model risk: *estimation risk* and *mis-specification risk*. The former one refers to the uncertainty arising from parameters estimation, once a parametric family of distributions has been chosen. Instead, the latter one refers to the choice of

Angelica Gianfreda
Free University of Bozen-Bolzano, Universitätsplatz 1, Bozen e-mail: `angelica.gianfreda@unibz.it`

Giacomo Scandolo
University of Firenze, Via delle Pandette 9, Firenze e-mail: `giacomo.scandolo@unifi.it`

the parametric family itself. Given that quantifying and managing mis-specification risk is more difficult and it has been investigated to a lesser extent[1], we aim at focussing on this issue considering energy markets over almost fifteen years of data (from 2001 to 2015), so that we are able to assess model risk on a long-run basis and depict its historical evolution.

2 Methodology

Given a financial portfolio, the Value-at-Risk $\mathrm{VaR}_{\alpha,t+1}$, from day t to day $t+1$ at level α, is implicitly defined through the equality $P(L_{t+1} > \mathrm{VaR}_{\alpha,t+1}) = \alpha$, where $L_{t+1} = V_t - V_{t+1}$ is the loss from day t to day $t+1$ (here, V_t and V_{t+1} are the portfolio market values at days t and $t+1$). We consider the values $\alpha = 1\%$ and 5%, typical for market risk.

The assessment of VaR naturally depends on a probabilistic model. Henceforth, at any given date, competing models will produce competing VaR forecasts and model risk arises when these forecasts are dispersed. Several measures of model risk have been proposed in the literature. Here, we consider the Relative Measure of Model Risk (henceforth, RMMR) as defined in Barrieu and Scandolo (2015). Let VaR_i be the forecast of $\mathrm{VaR}_{\alpha,t+1}$ under model i, and VaR^* the forecast under a *reference* model; then, the RMMR is defined as the number

$$\mathrm{RMMR} = \frac{\max_i \mathrm{VaR}_i - \mathrm{VaR}^*}{\max_i \mathrm{VaR}_i - \min_i \mathrm{VaR}_i} \tag{1}$$

It easily turns out that this number lies in the interval $[0,1]$ (provided the reference model is among the competing models), and it is insensitive to the amount invested in the portfolio. We observe that the closer is RMMR to 1, the lower is VaR^* with respect to the other competing forecasts and therefore the higher is the amount of model risk involved.

In this work, we consider three portfolios, each of them investing in one of the following energy-related assets: Brent crude oil (*Oil*), ICE UK natural gas (*Gas*), and day–ahead auction prices for electricity observed on the European Energy Exchange (EEX) for delivery in the German/Austrian zones (*Electricity*). The oil and gas series always displayed positive prices, while several occurrences of negative prices were observed for electricity. Therefore, for the former two portfolios we set $L_t = -(e^{X_t} - 1)$, where X_t is the daily log-return observed at day t; for the electricity portfolio, we simply set $L_t = -X_t$, where X_t is the daily price change. In each case, we model the series (X_t) through an AR(5)–GARCH(1,1) process. Specifically, we have $X_t = \mu_t + \sigma_t Z_t$, where $\mu_t = \overline{\mu} + \sum_{i=1}^{5} \phi_i X_{t-i}$ is the conditional mean following an AR(5) process, and $\sigma_t^2 = \omega + \alpha(X_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2$ is the conditional variance following a GARCH(1,1) model. The IID innovation series (Z_t) can follow any of 9 competing standard distributions, which are: *normal* (NORM) and *skew*

[1] See for instance Cont (2006) and Daníelsson et al. (2016); among others.

normal (SNORM); *Student-t* (STD) and *skew Student-t* (SSTD); *Generalized Error Distribution* (GED) and *skew GED* (SGED); *Johnson's S_U family* (JSU); *Normal Inverse Gaussian* (NIG); *Generalized Hyperbolic family* (GHYP). Notice that some of these models are nested[2] and some of them allow for extra-parameters that give control on asymmetry and/or tail behaviour.

For any fixed distribution for the innovations, the parameters of the AR-GARCH model are estimated day-by-day by ML using a rolling window of 256 past daily data. This allows us to obtain a forecast VaR_i ($i = 1, \ldots, 9$) of the daily Value-at-Risk under all competing models. Once a reference model is fixed throughout (NORM, for instance), the final output is a daily series of model risk measures.

As a result of the rolling estimation process, we obtain the maximized log-likelihood $\widehat{\ell}_i$ ($i = 1, \ldots, 9$) for each day and competing model, and then the series of Bayesian Information Criterion, defined as $BIC_i = -2\widehat{\ell}_i + \log(n)p_i$, where n is the length of the dataset ($n = 256$ in our analysis) and p_i is the number of parameters in model i. [3] We observe that BIC is a measure of fitting ability (the lower is BIC, the better is the fitting) which penalizes for over-parametrization, and we use these values in two ways. First, on a daily basis we can rank the 9 models according to their fitting ability. In particular, at each day we single out the *daily best* model, as the one with the lowest BIC value; we then employ this model as the reference one in the computation of RMMR. Second, we can attach to each model a *percentage weight*, defined as

$$w_i = \frac{a_i^2}{\sum_{j=1}^{K} a_j^2}, \tag{2}$$

where

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

Notice that w_i is decreasing in BIC_i: higher fitting ability is therefore associated with higher weights. We use these weights to discard, on a daily basis, the worst fitting models; specifically, after ranking the models with increasing weights, we retain models until the cumulative weight 0.95 is reached. We think this step is crucial in obtaining a measure of model risk that does not strictly depend on the initial choice of the competing models. As a consequence, RMMR do not necessarily lie in the interval $[0, 1]$ if the reference model is among the discarded ones. Finally, we use the weights to obtain an average forecast, naturally defined as $\mathrm{VaR}_{\mathrm{avg}} = \sum_i w_i \mathrm{VaR}_i$, which can be used as a possible reference model throughout (here, discarded models are left out of the average and weights are therefore properly normalized).

---

[2] For instance, NORM is a particular case of SNORM.

[3] For instance, $p_i = 10$ for the STD model: 6 parameters for the AR process, 3 for the GARCH process and 1 for the STD distribution.

3 Empirical results

As anticipated, for each day we obtain various measures of model risk, by considering as reference model either one specific distribution, or the daily best, or the average forecast $\mathrm{VaR}_{\mathrm{avg}}$. We always discard the worst fitting models, as explained above (on average, 2 or 3 models are discarded each day). We repeat the entire procedure for VaR at $\alpha = 1\%$ and $\alpha = 5\%$ and for all three portfolios of individual assets (oil, gas and electricity). All time series have been collected 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 used the R-package `rugarch`. Some of the RMMR series and related empirical findings are shown next.

The "overall best" model for each of the three assets is identified by looking at the majority of days in which it is found to be the best fitting model. Overall best models (and in parenthesis the "overall worst" models, being the best daily models in the fewest number of days) are: STD (GHYP) for oil, GED (GHYP) for gas, and STD (SNORM) for electricity. We observe that the Normal model is not the overall worst model for any of the three assets because the use of BIC penalizes models with many parameters.

In Table 1 we show some descriptive statistics of the RMMR (for $\alpha = 1\%$) for some choices of the reference model. We emphasize that while the overall best/worst models are fixed throughout for any given asset, the daily best model may change on a daily base. As explained before, weights are computed for each model: they are used for discarding the worst fitting models and to compute the average forecast $\mathrm{VaR}_{\mathrm{avg}}$.

Reference model		Oil	Gas	Ele
Normal	mean	1.04	0.54	0.64
	std. dev.	0.52	0.74	0.91
	max	5.95	4.75	5.04
Overall best	mean	0.52	0.64	0.77
	std. dev.	0.32	0.33	0.44
	max	2.28	3.44	8.25
Overall worst	mean	1.59	1.31	0.73
	std. dev.	0.46	0.44	0.50
	max	5.99	5.55	9.41
Daily best	mean	0.31	0.20	0.29
	std. dev.	0.30	0.28	0.34
	max	1.00	1.00	1.00
Average forecast	mean	0.48	0.49	0.44
	std. dev.	0.11	0.13	0.14
	max	0.87	0.88	0.95

Table 1 Some descriptive statistics for RMMR ($\alpha = 1\%$).

Looking at the mean levels, we see that the daily best models provide on average the lowest amount of model risk. The overall best gives less model risk than the overall worst for Oil and Gas; a bit surprisingly, this is not the case for Electricity. The model risk associated to the average VaR is stable around 0.5. Looking at the maximum levels we see that the RMMR sometime reach levels around 10, which signal a huge amount of model risk.

Figure 1 shows the daily dynamics of RMMR for Gas during 2002 (with $\alpha = 1\%$), when using the overall best (GED) and overall worst (GHYP) as reference models. We can see that, even though RMMR with respect to GHYP is consistently higher than for GED, the two series are highly volatile (as confirmed by the standard deviations in Table 1). Therefore, to ease the presentation, we used rolling means of the RMMR computed on a 256-day basis, and we show the dynamics of the RMMR ($\alpha = 1\%$) for two portfolios in Figure 2 (gas and oil shows similar RMMR dynamics). Finally, we also compare the RMMRs for both $\alpha = 1\%$ and $\alpha = 5\%$, showing[4] that the amount of model risk depends on α

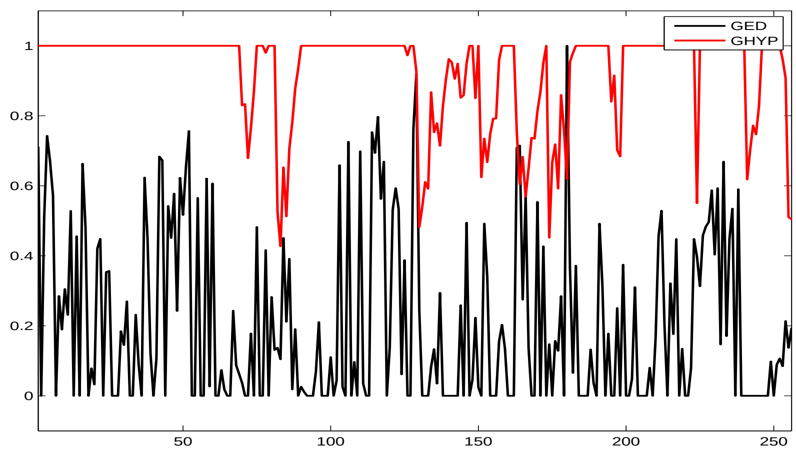

Fig. 1 Dynamics of RMMR for $\mathrm{VaR}_{1\%}$ for Gas in 2002, using the overall best (GED) and the overall worst (GHYP) as reference models.

4 Conclusions

Quantifying and managing mis-specification risk has been less investigated so far, hence we have provided for the first time the empirical assessment of mis-specification risk when studying energy assets. Relaxing the assumption of normality and using a wide range of alternative distributions, we have quantified model risk under the well-established setting of GARCH models. 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. In general, we find that better models tend to produce less model risk. Although not

[4] This figure is omitted for lack of space.

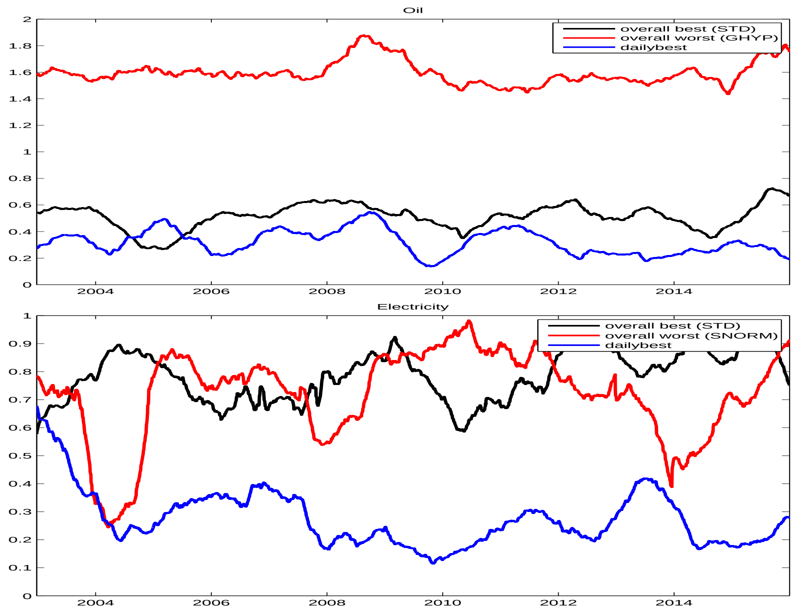

Fig. 2 Dynamics of rolling mean RMMR for $\alpha = 1\%$ using the overall best, the overall worst, and the daily best as reference models for Oil (top panel) and Electricity (bottom panel).

completely surprising, this pattern is quite evident across different assets and levels of VaR.

It is worth highlighting that our analysis intentionally addresses the choice of the innovations distribution, which is one among a number of possible sources of model risk affecting the final VaR figures. Future research may indeed address the misspecification risk due to the choice of the number of lags in the ARMA-GARCH structure or even due to the choice among different types of specification for the conditional mean/volatility.

Acknowledgements The authors kindly acknowledge the Free University of Bozen-Bolzano, for the internal research funding of the project *FOrecasting and MOnitoring electricity Prices, volumes and market Mechanisms*, “FoMoPM”.

References

1. Barrieu, P. and Scandolo, G.: Assessing financial model risk. European Journal of Operational Research. **242**, 546-556 (2015)
2. Cont, R.: Model uncertainty and its impact on the pricing of derivative instruments. Mathematical Finance. **16**, 519–547 (2006)
3. Daníelsson, J., James, K.R., Valenzuela, M., Zer, I.: Model risk of risk models. Journal of Financial Stability. **23**, 79–91 (2016)