(!)

PRELIMINARY DRAFT

PLEASE **DO NOT** QUOTE OR DISTRIBUTE WITHOUT PRIOR PERMISSION OF AUTHORS

STRESS TESTING OF ENERGY-RELATED DERIVATIVE INSTRUMENTS BASED ON CONDITIONAL MARKET RISK MODELS

Abstract

Stress testing has recently been an important risk management tool to both complement and supplement risk measurement methods under different stress scenarios. It is already a part of Basel II capital rules introduced by the *Basel Committee on Banking Supervision*. This paper performs thorough *backtests* under widely used risk models in combination with different conditional volatility frameworks to identify the most suitable risk model specification for *structured energy instruments* mostly traded in financial markets and conducts a *stress test* based on the specified models. We find strong evidence that, for a smaller estimation window, the volatility model that incorporates asymmetric volatility response together with *Student's t* innovations clearly outperforms symmetric ones. However, the *conditional empirical model* is superior when larger estimation windows are considered. Also, it is shown that the *backtest* performances of volatility models are substantially affected by the distribution selection. *Stress test* results clearly indicate that energy shocks can lead greater stress losses than those observed in major currency pairs. Associated with an appropriate initial shock and estimation window size, the stress losses estimated by the selected models favorably compare with past shocks in all but one extraordinary case: the 1st Gulf War scenarios ended in the largest ever price drops. Referring to Basel II capital calculation rules, the results imply the need for significant risk capital buffers for energy portfolios.

Keywords: *Energy Derivatives, Stress testing, Conditional Volatility Models, Value-at-risk*

Özet

Stres testleri, farklı stres senaryoları altında kullanılan tamamlayıcı ve destekleyici risk yönetim metotları için önemli bir risk yönetim aracı haline gelmiştir. Hâlihazırda bu testler *Basel Bankacılık Denetim Komitesi* tarafından düzenlenmiş olan Basel II sermaye kurallarının da bir parçasıdır. Bu çalışma, finansal piyasalarda sıklıkla işlem gören *yapılandırılmış enerji enstrümanları* için en uygun risk model bileşimini belirlemek amacıyla, farklı koşullu oynaklık yapıları ile birlikte çoğunlukla kullanılan risk modelleri üzerinde kapsamlı *geriye dönük testler* sunuyor ve belirlenen modeller üzerinde bir *stres testi* uyguluyor. Göreceli olarak daha kısa olan tahmin penceresi dikkate alındığında, asimetrik oynaklık tepkilerine olanak sağlayan oynaklık modelinin *Student's t* hata terimi varsayımı ile birlikte kullanıldığında simetrik modellerden daha olumlu bir performans sergilediğine dair güçlü bulgulara ulaşıyoruz. Ancak bunun yanında, daha uzun tahmin pencerelerinde *koşullu ampirik model* daha üstün bir görüntü çiziyor. Bununla birlikte, oynaklık modellerinin sergilediği *geriye dönük test* performansların dağılım seçimine paralel olarak önemli ölçüde etkilendiği görülüyor. *Stres testi* sonuçları, enerji şoklarının yol açabileceği stres kayıplarının büyük para birimi çiftlerine ait kur değerlerinde gözlenen kayıplardan daha büyük olabileceğine işaret ediyor. Uygun bir başlangıç şok değeri ve tahmin penceresi uzunluğu ile ilişkilendirildiğinde, seçilen risk modelleri tarafından tahmin edilen stres kayıpları, geçmiş dönemde en yüksek fiyat düşüşleri ile sonuçlanan Birinci Körfez Savaşı senaryolarını içeren olağandışı bir vaka haricinde geçmiş şoklarla makul bir şekilde örtüşüyor. Basel II sermaye hesaplama kuralları dikkate alındığında sonuçlar, enerji portföyleri için ciddi sermaye takviyelerine ihtiyaç duyulduğuna işaret ediyor.

Anahtar kelimeler: *Enerji Türev Ürünler, Stres Testi, Koşullu Oynaklık Modelleri, Riske Maruz Değer*

1. Preamble

The Basel Committee on Banking Supervision revised its 1988 Capital Accord to incorporate market risks and recommended that the senior management review the results of stress testing periodically, use them in the assessment of capital adequacy, and reflect their implications in the policies and limits set by management and the board of directors (2004). Stress tests continue to focus primarily on traded market portfolios. These portfolios are well suited to stress testing as they can be marked to market on a regular basis. As from the end of 1997, banks are required to measure and apply capital charges in respect of their market risks in addition to their credit risks. However, most of the stress testing approaches are still exposed to risk model misspecification threat and even some of them have yet to have an underlying risk model. A survey conducted by the Basel Committee points out that the historical and hypothetical events form the basis of a great majority of the stress tests (Committee on the Global Financial System, 2005)

Increasingly, individual institutions are taking into account information about plausible worst case scenarios and where it is deemed prudent, taking action to avoid the adverse effects of these events. However, yet no coherent stress testing framework has been applied to much more volatile energy-related financial products. Therefore, this paper aims at introducing a model-based approach for the stress testing of a certain set of liquid energy contracts that can incorporate both volatility clustering and heavy-tails assumptions, and explores the potential impact of plausible price shocks that can occur in energy markets on the portfolios held by a variety of institutions for different purposes, varying from commodity positions hedging to even market speculation. The implications of stress test results are also examined in the Basel II context. This helps us enhance the risk-awareness of investors, senior management and stakeholders, as well as regulatory and supervisory authorities, against extreme risks which would arise from low-probability events.

Thus, this study first examines the one-day-ahead forecasting accuracies of three mostly used volatility models under two distributional assumptions using the rolling window approach described by Brooks (2007) and, then performs a stress test based on the risk model(s) which is (are) found relatively more robust. Forecast accuracies are evaluated accordingly with two well-known backtesting statistics. Recursive VaR estimation is expected to reveal the consistency of any risk model under different data samples while the different distributions allow the selection of a model for the return tails. For statistical robustness, this study uses three different energy futures series to avoid outputs rely heavily on one specific series. Having almost 43,000 one-step-ahead VaR values estimated for all series and estimation windows with a specified distribution and volatility model, we produce nearly 1,300,000 one-step-ahead VaR forecasts when all volatility models, presumed distributions and asset positions (long or short) are considered.

The backtest results clearly reveal the fact that Gaussian distribution is not appropriate for the most liquid energy-related futures contracts traded in financial markets. Another interesting result is that, given fat-tailed innovations, the conditional volatility models incorporating asymmetric volatility response are superior to traditional GARCH setting proposed by Bollerslev (1986), particularly at higher confidence levels. It holds for all series and almost all confidence levels considered. This can be interpreted as an evidence for the existence of asymmetric returns in energy markets. Lastly, it is found that the size of the estimation window has significant effects on backtest performances: working with a larger estimation window leads to more accurate results in conditional empirical case, while it causes the "asymmetric GARCH with Student's t innovations" and the "GARCH with Student's t innovations" alternative combinations to behave in a more conservative way than expected.

Model-based stress loss estimations reveals the fact that the risk models that performed better in backtests are able to generate consistent price paths for artificial stressful periods (triggered by different initial tail events), which mostly match the historical price discontinuities observed in previous energy crises. Taking into account a minimum of 10-day risk horizon, which Basel II Capital Accord also requires in internal risk calculations, the selected models predict significant lower bounds for necessary risk capital to absorb potential large losses in energy portfolios.

2. Origins and related studies

The appearance of stress tests in the finance literature is gradual. The roots can be traced back to market risk amendment to the first capital accord (Basel Committee on Banking Supervision, 1996). However, the second accord included a more comprehensive definition of stress testing and gave a special emphasis on it. According to this new accord, supervisory authorities would not approve the use of internal risk models unless they were supplemented with stress tests. Moreover, the institutions would be required not only to follow a routine and rigorous stress testing, but also to reflect the results in the assessment of capital adequacy as well as the management policies. It was also recommended that the stress scenarios covered a range of factors that can create extraordinary losses or gains and make the control of market risk very difficult (Basel Committee on Banking Supervision, 2004).

The green shoots of foundations of the bridge between stress tests and risk models starts with the work of Kupiec (1998) who examined cross market effects resulting from a market shock. Aragones et al. (2001) criticized traditional stress testing approaches for being inevitably subjective, difficult to backtest and not providing probabilistic outcome to allow sound interpretations about their results. Alexander and Sheedy (2008), in their most recent study, proposed a stress testing methodology based on a set of most suitable risk models on which a rigorous set of backtests are conducted to eliminate model risk.

The remaining parts are organized as the following: Sections 3 and 4 shortly introduce the data and risk models included. Then, in Section 5, will follow the backtests and the process for selecting the best-performing risk models. Section 6 covers the stress testing methodology in detail and performs stress tests based on the risk models outperformed in the previous section. The last but not the least, Section 7 deals with the interpretation of the results and draws some implications for risk capital which can be deemed useful for regulatory and supervisory authorities. Finally, Section 8 concludes.

3. Data

The data used in this study are obtained from DataStream and consist of a group of highly liquid energy futures which are traded on NYMEX. Selection is based on the purpose of exploring the effects of a stress event on the value of highly liquid energy contracts and comparing the performance of different risk models in capturing extreme returns in more volatile markets. Among the NYMEX futures contracts, Cushing settled crude oil, natural gas and heating oil futures contract price series are used in analyses. All series include their most recent observations in February 2009. Crude oil series comprise of a total of 6484 observations covering a full daily data starting from the first quarter of 1983, while heating oil has a total of 7300 daily returns from January 1980. Natural gas has a relatively small sample size consisting of 3772 observations starting from January 1994. The price series are converted into returns using the usual logarithmic transformation. Table-1 consists of summary statistics for these three return series. All of them clearly appear to be non-normal as evident leptokurtosis suggests. However, another important point here is that both crude and heating oil returns are negatively skewed implying that the volatility models which allow asymmetry in return distributions are likely to outperform symmetric ones.

Table-1: Summary statistics

	Crude Oil	Natural Gas	Heating Oil
Mean	0.004%	0.019%	0.006%
Standard Deviation	2.44%	3.81%	2.34%
Kurtosis	16.82	6.99	20.33
Skewness	-0.94	0.07	-1.53
Min/Max	-40.00/16.40%	-37.60/32.40%	-39.10/14.00%

At this point, it is worth noting that the data should include at least one turbulent period so that we can make sure that the model is provided with sufficient input to generate extreme enough outputs at the end of the stress horizon.

4. Market risk models

4.1. Conditional normal

This set of symmetric risk models included in this study consists of simple GARCH model which presumes symmetric returns around zero and Gaussian innovations. Specifically, the mean adjusted returns are assumed to be conditionally normally distributed with conditional variance following the symmetric GARCH(1,1) process of Bollerslev (1986):

$$\sigma_t^2 = \xi_1 + \xi_2 \sigma_{t-1}^2 + \xi_3 \varepsilon_{t-1}^2 \tag{1}$$

$$\xi_1 \geq 0 \tag{2}$$

$$\xi_2, \xi_3 > 0 \tag{3}$$

$$\xi_2 + \xi_3 < 1 \tag{4}$$

where $\varepsilon_t \sim N(0, \sigma_t^2)$.

To calculate VaR in one-step-ahead forecasts for backtesting purposes, we simply produce volatility forecasts using the appropriate model parameters as well as the most recent conditional variance and innovation terms. That is, for a long (short) position, $VaR_{\alpha,1}$ corresponds to the absolute value of the lower (upper) $100 * \alpha$ percentile of the innovations drawn from the assumed conditional distribution plus the mean equation constant. However, in calculation process of *s-day* stress loss, following Alexander and Sheedy (2008), we do not simply plug the *s-day* variance in the GARCH forecast process since this method assumes that all the innovations are 'typical' during the stressful market conditions. This alternative would suffer from the lack of ability to produce large shocks during stress horizon which would in turn not cause volatility increases. Thus, we employ Monte Carlo simulation to eliminate this problem that will be discussed later on.

4.2. Conditional Student's t

Heavy-tailed distributions have been mostly combined with GARCH models in VaR estimation literature. The method is identical to the one described in the previous part, but now innovations are drawn from Student's t distribution with v degrees of freedom. That is:

$$\varepsilon_t (v/(v-2))^{0.5} \sim t_v \tag{5}$$

Using a heavy-tailed distribution is supposed to help us capture the conditional excess kurtosis in empirical data.

4.3. Conditional empirical

Using past returns directly to forecast future changes in portfolio value is popular in the industry, though unconditionally. Here we adopt a slightly different approach which is similar to the one described in Barone-Adesi et al. (1998). We make no distributional assumption about the standardized past returns, other than the assumption that there is a mild dependence between them. We first fit either a normal or a Student's t GARCH process to historical data before we standardize each return in the sample by subtracting mean and then dividing them by the corresponding in-sample conditional standard deviation estimate. Once we standardize them, the returns are scaled to the conditional standard deviation estimate for the day on which the VaR is estimated. That is, the standardized returns are multiplied by the current standard deviation estimate to obtain the sample of scaled returns. To calculate $VaR_{\alpha,1}$, we simply draw the lower (upper) $100*\alpha$ percentile of scaled returns based on whether the asset is long (short). To calculate *s-day* stress loss, the GARCH model is simulated forward over the *s-day* risk horizon using innovations that are sample from scaled past returns.

4.4. Asymmetric[1] conditional normal

Unlike the conventional approach in currency markets, we do not discard the use of asymmetric GARCH models for two reasons. First, it is not a stylized fact to use only symmetric volatility models in energy markets. Second, all of the series considered in this study, except natural gas, exhibit significant skewness (even when means are subtracted) implying the need for including asymmetric models as well as symmetric ones. We use the exponential GARCH (EGARCH) model proposed by Nelson (1991) where the disturbance terms are assumed to fit either to Normal or Student's t density. The conditional variance in the EGARCH model can be expressed by in a variety of ways, but one among them can be written as (under normal innovations assumption):

$$\ln(\sigma_t^2) = \xi_1 + \xi_2 \ln(\sigma_{t-1}^2) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \omega \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \tag{6}$$

4.5. Asymmetric conditional Student's t

An alternative way of expressing the asymmetric model introduced above would be (under Student's t innovations assumption):

[1] The volatility responses are not restricted but allowed to be asymmetric.

$$\ln(\sigma_t^2) = \xi_1 + \xi_2 \ln(\sigma_{t-1}^2) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \omega \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{v-2}{\pi}} \frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})} \right] \quad (7)$$

where γ is the leverage parameter.

The EGARCH model also allows for potential asymmetries via γ parameter while necessitates no non-negativity constraints since σ_t^2 term would be positive even though $\ln(\sigma_t^2)$ would not.

Before completing model discussion, it is worth noting that we employed only (p,q)=(1,1) type of conditional volatility models since the former studies do not suggest any evidence about the superiority of more sophisticated models. Hansen and Lunde (2005) compared 330 ARCH-type models in terms of their ability to fit conditional volatility and found no evidence that more complicated settings could cast shadow over the performance of the GARCH(1,1) setting. However, they concluded that the GARCH(1,1) was obviously inferior to the ones that can incorporate leverage effects (i.e. asymmetries).

5. Backtests

This section is devoted to the efforts of selecting the most suitable model(s) for stress testing purposes. To better reflect the effects of a stress event on the value of futures contracts written on energy commodities, we include low probability regions (99.5% and 99.8%) as well as typical 99%. This also helps us reveal some distinguishing skills of heavy-tailed distributions which are argued to perform better in tail extremes. The inclusion of a relatively shorter estimation window takes into account the findings of Hoppe (1998) who argued that smaller sample sizes could lead to more accurate VaR estimates than larger ones. Frey and Michaud (1997) also argued that a small sample size is better in capturing structural changes. However, the Basel regulations insist on an estimation window of more than or equal to 1 year (around 250 trading days). Therefore, we consider a range of estimation windows including both larger and smaller ones: 250, 1000 and 2000 trading days.

The most suitable ones among candidate models were filtered by two different test statistics: the test for coverage by Kupiec (1995) and the test for conditional coverage by Christoffersen (1998). The former has a null hypothesis that the actual number of violations is equal to the expected number of violations while the latter tests the null hypothesis that the violations are not clustered over time.

We followed a thorough backtest methodology to reduce the model risk in the risk models that we will use for stress testing purposes. We recursively calculate one-day-ahead VaR estimates using rolling window method where each estimation window has 250, 1000 or 2000 daily observations. Thus, considering an estimation window of 250 days, the first volatility estimate occurs on 251st trading day and goes up to February 2009. This method allows us to obtain nearly 43,000 one-step-ahead VaR projections for all series considered and all estimation windows included, implying almost 1,300,000 one-step-ahead VaR estimates which are used in backtest process when all distributions, volatility models and asset positions are taken into account. At the end of each estimation period, actual profits/losses are calculated and any case in which the actual loss is greater than VaR estimate in absolute value was recorded as an exceedance. Afterwards these exceedances, together with their probable time-varying frequencies were used as inputs for tests for conditional and unconditional coverage. Results are shown in Table 2 and Table 3 for each of the five distribution-variance model specifications. The greater the p-values, the more confident the null hypotheses are accepted and the more the model is suitable for stress testing.

Table-2: Backtest results (estimation window: 250 days)

		(1) Conditional normal			(2) Asymmetric Conditional normal			(3) Conditional Student's t			(4) Asymmetric Conditional Student's t			(5) Conditional empirical		
	100*(1-α)	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc
long Crude Oil	99	1.91	0.00	0.00	2.20	0.00	0.00	0.85	10.86	22.28	0.96	6.96	76.53	1.48	0.03	0.04
	99.5	1.30	0.00	0.00	1.41	0.00	0.00	0.47	28.89	69.40	0.56	34.29	49.95	0.87	0.02	0.02
	99.8	0.85	0.00	0.00	1.03	0.00	0.00	0.26	60.56	33.73	0.32	13.68	4.97	0.66	0.00	0.00
short Crude Oil	99	1.56	0.01	0.00	1.57	0.01	0.00	0.59	0.11	0.05	0.79	3.73	7.79	1.57	0.01	0.00
	99.5	1.09	0.00	0.00	1.16	0.00	0.00	0.35	4.15	8.22	0.40	46.85	25.15	0.75	2.03	0.82
	99.8	0.77	0.00	0.00	0.69	0.00	0.00	0.16	2.99	46.89	0.22	88.48	66.98	0.48	0.00	0.00
long Natural Gas	99	1.33	8.71	5.75	1.76	0.02	0.00	0.43	0.05	0.01	0.77	15.61	14.70	1.25	20.52	15.18
	99.5	0.88	1.17	0.39	1.19	0.00	0.00	0.26	7.43	2.32	0.31	22.93	8.99	0.62	52.26	31.22
	99.8	0.51	0.24	0.06	0.74	0.00	0.00	0.03	1.66	0.42	0.14	71.36	41.63	0.31	37.35	16.80
short Natural Gas	99	2.13	0.00	0.00	2.22	0.00	0.00	0.74	21.62	10.19	0.82	43.63	27.80	1.16	39.07	33.92
	99.5	1.25	0.00	0.00	1.59	0.00	0.00	0.37	48.96	24.85	0.45	86.17	69.69	0.68	29.75	14.77
	99.8	0.94	0.00	0.00	1.08	0.00	0.00	0.23	92.24	72.36	0.28	56.07	29.42	0.45	1.42	0.38
long Heating Oil	99	1.96	0.00	0.00	2.21	NA	NA	0.89	57.27	36.12	1.01	94.80	95.14	1.46	0.12	0.03
	99.5	1.56	0.00	0.00	1.77	0.00	0.00	0.51	82.46	89.89	0.60	28.31	26.83	0.81	0.27	0.08
	99.8	0.99	0.00	0.00	1.29	0.00	0.00	0.24	72.49	45.37	0.34	5.21	1.65	0.54	0.00	0.00
short Heating Oil	99	1.65	0.00	0.00	1.77	0.00	0.00	0.60	0.06	0.02	0.65	0.45	0.17	1.26	7.79	3.33
	99.5	1.04	0.00	0.00	1.21	0.00	0.00	0.34	2.59	4.38	0.44	24.14	46.42	0.81	0.08	0.08
	99.8	0.65	0.00	0.00	0.77	0.00	0.00	0.07	1.98	0.51	0.20	97.22	97.91	0.48	0.00	0.00

%: proportion of observed model violations, **p-tfcc**: p-value for the test of conditional coverage, **p-tfuc**: p-value for the test of unconditional coverage

As p-values in parts (1) and (2) of Table-2 suggest, the assumption of conditionally normal-distributed returns cannot be justified, either in the symmetric or asymmetric models. Also the estimation window size cannot change this result. The null hypotheses that (i) the actual number of violations is equal to the expected number of violations and (ii) the violations are spread evenly over time are comfortably rejected in most of the cases. Column (3) of Table-2 shows that the symmetric GARCH model under heavy-tails assumption gives promising results to some extent except some of the situations where the crude and heating oil contracts are short and natural gas contract is long. This situation can be attributed to asymmetric returns that we evidenced in the series. That is, negatively skewed crude and heating oil returns reduce the number of potential shocks in the upside causing symmetric models to appear too conservative in short portfolio case. Similar interpretation can be made for slightly positively skewed natural gas case. Moreover, in a great majority of the cases in which the hypothesis that the actual number of model exceedances is equal to the expected violations is rejected, the p-values of the test for conditional coverage also indicate that the exceedances are clustered. Column (4) of Table-2 shows that the EGARCH model eliminates the violation clustering problem resulted from the strongly asymmetric structure of crude and heating oil returns. At 95% confidence level, in 16 out of 18 cases (89%), the number of return shocks captured by the model is found to be very close to what is expected beforehand. Again at the same level of confidence, and in 16 out of 18 cases, we can conclude that the violations are spread evenly over time, i.e. they are not clustered. What is interesting in Table-2 is that the conditional empirical model results are not as satisfying as its popularity in the industry suggests. However, this might be due to insufficient length of estimation window, as it will be evidenced by the results observed in larger estimation windows (see Table-3).

As it can be inferred from the columns (1) and (2) of Table-3, working with a larger estimation window is clearly not enough to justify normality assumption for the innovations. However, it is also worth noting that both symmetric and asymmetric models with heavy-tailed disturbances now suffer from too conservative results (columns (3)-(4)). To illustrate, the observed number of model violations ranges between 0.11% and 0.87% while it is expected to be around 1%. Although this situation still prevents the position holder from incurring larger losses than its capital provisions, it is not desirable to tie up in capital more resources than needed. As the last column of Table-3 suggests, given an estimation window of nearly 4 years, conditional empirical model appears to be the best of all the risk models considered. In none of 18 cases can we reject the unconditional coverage measure at the 95% confidence level, while conditional coverage test implies clustering in a single case with a slight difference when 95% confidence level is considered.

Table-3: Backtest results (estimation window: 1000 days)

		(1) Conditional normal			(2) Asymmetric Conditional normal			(3) Conditional Student's t			(4) Asymmetric Conditional Student's t			(5) Conditional empirical		
	100*(1-α)	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc	%	p-tfcc	p-tfuc
long Crude Oil	99	1.57	0.02	0.01	1.70	0.00	0.00	0.51	0.01	0.01	0.64	0.75	0.39	1.02	86.37	87.43
	99.5	1.04	0.00	0.00	0.98	0.00	0.00	0.22	0.39	0.09	0.26	13.73	5.46	0.53	81.92	76.37
	99.8	0.64	0.00	0.00	0.66	0.00	0.00	0.07	5.29	1.54	0.09	12.91	4.33	0.26	65.55	37.93
short Crude Oil	99	1.39	1.77	0.66	1.28	8.44	4.83	0.38	0.00	0.00	0.36	0.00	0.00	1.02	86.37	87.43
	99.5	0.88	0.12	0.04	0.77	2.53	0.96	0.15	0.01	0.00	0.18	0.06	0.01	0.49	87.22	93.65
	99.8	0.53	0.00	0.00	0.47	0.05	0.01	0.07	5.29	1.54	0.11	25.75	10.03	0.29	34.65	15.46
long Natural Gas	99	1.12	58.23	53.76	1.26	26.14	18.11	0.11	0.00	0.00	0.18	0.00	0.00	1.01	75.02	95.59
	99.5	0.69	37.12	18.97	0.76	17.24	7.38	0.04	0.00	0.00	0.07	10.03	9.01	0.54	87.99	76.09
	99.8	0.40	11.86	4.10	0.40	11.86	4.10	0.04	5.88	1.73	0.04	5.88	1.73	0.22	96.90	84.76
short Natural Gas	99	1.66	0.60	0.14	1.62	0.98	0.25	0.40	0.13	0.03	0.43	8.31	10.07	0.90	69.35	59.88
	99.5	1.73	0.00	0.00	1.73	0.00	0.00	0.36	53.08	27.45	0.47	91.54	81.60	0.43	83.28	60.91
	99.8	1.19	0.00	0.00	1.23	0.00	0.00	0.11	49.40	23.61	0.11	49.40	23.61	0.18	96.42	81.47
long Heating Oil	99	1.78	0.00	0.00	0.00	0.00	1.95	0.59	0.14	0.04	0.67	0.67	0.47	1.02	91.33	89.85
	99.5	1.32	0.00	0.00	1.44	0.00	0.00	0.29	3.05	0.87	0.29	3.05	0.87	0.62	34.03	19.62
	99.8	0.89	0.00	0.00	1.02	0.00	0.00	0.17	88.22	64.50	0.17	88.22	64.50	0.25	62.85	35.73
short Heating Oil	99	2.21	0.00	0.00	2.22	0.00	0.00	0.83	7.54	15.13	0.87	2.65	30.11	1.19	19.94	13.98
	99.5	1.57	0.00	0.00	1.59	0.00	0.00	0.44	23.73	52.45	0.52	37.36	78.97	0.60	25.89	26.05
	99.8	1.17	0.00	0.00	1.22	0.00	0.00	0.19	5.51	86.50	0.21	6.55	91.02	0.30	3.43	9.33

%: proportion of observed model violations, **p-tfcc**: p-value for the test of conditional coverage, **p-tfuc**: p-value for the test of unconditional coverage

Using an estimation window of 2000 observations provides a set of similar backtest results[2] to those included in Table-3. In order to conserve space, we do not include them here.

[2] The results for a 2000-day estimation window are available upon request from authors.

The analysis so far suggests that both the conditional empirical model and the asymmetric EGARCH model in combination with a fat-tailed distribution can adequately describe the distribution of energy derivative returns. Our results confirm that, in conditional empirical case, large estimation windows (1000, 2000 days) are preferable to smaller one (250 days) for risk estimation. However, it is the reverse when asymmetric conditional Student's t model is the case due to its excessively conservative risk estimations observed in larger estimation windows.

6. Stress testing methodology

This section incorporates the three characteristics brought under the spotlight in the previous section: heavy-tailed returns assumption, volatility clustering and (asymmetric) leverage effect. The methodology adopted in this paper is proposed by Alexander and Sheedy (2008). First, a stress event based on a heavy-tailed distribution is defined below and the subsequent market response to that shock is simulated employing a conditional volatility model to allow volatility clustering and potential asymmetric volatility responses. Stress results belong to the two models mentioned above, while the poor-performing conditional normal model is also included for comparison of stress losses.

6.1. Defining a stress event

An initial shock (or a stress event) can be defined as an unanticipated but plausible event that causes a large discontinuity in prices. An alternative to recommendation of the Committee on the Global Financial System (2005) (i.e. the method of basing such kind of an event on a historical of hypothetical event by taking the experiences of management into account) is to consider extreme outcomes defined by the risk model. That is, for a long position, VaR calculated using a typically low probability region, say α_s, can be used as a stress event and the results subsequent volatility hikes can be explored. Then an initial shock at the beginning of the stress period T, say ε_T^* can be derived from the specified distribution. For example, under heavy-tails assumption, a stress event for a long portfolio is defined as:

$$\varepsilon_T^* = t_v^{-1}(\alpha_s)((v-2)/v)^{0.5}\,\overline{\sigma}_T \tag{8}$$

or similarly, under normality assumption:

$$\varepsilon_T^* = \Phi^{-1}(\alpha_s)\overline{\sigma}_T \tag{9}$$

Under the empirical approach, the initial shock is simply the upper or lower α_s percentile of the empirical distribution depending on whether the asset position is short or long.

Variance at time T, $\overline{\sigma}_T^2$, can be taken as either the volatility estimate at time T or the equally weighted historical volatility to reflect normal market conditions in the long term. Typical values for α_s are 0.0002 (0.9998) and 0.0005 (0.9995) for long (short) portfolios, implying a loss that we are respectively 99.98% and 99.95% confident that our actual loss stay beneath over one day.

Table-4 reflects the magnitudes of an initial shock for a long portfolio in corresponding contract based on two different distributional assumptions. Shocks drawn from normal density are also included for comparison. The numbers are calculated using full samples. Note that both the fitted Student's t and the empirical distribution give much more conservative results, as one could expect. To illustrate, the first day of a stress horizon under Student's t assumption is characterized by an initial decrease of 18.09% in the value of heating oil futures, if we would have chosen 99.98% confidence level. It is clear that the selection of α_s level reflects the way market risks are perceived by a firm.

Table-4: Initial shocks for the stress tests

	alpha=0.0005			alpha=0.0002		
	Gaussian	Student's t	Empirical	Gaussian	Student's t	Empirical
long Crude Oil (%)	8.01	14.43	25.58	8.62	17.12	35.16
long Natural Gas (%)	12.53	23.15	16.33	13.48	27.59	18.46
long Heating Oil (%)	7.7	15.04	21.79	8.29	18.09	24.18

6.2. Modeling the system response

The analysis in part 5 confirms that among all approaches to capture extreme events, the risk models that have heavy-tails and an asymmetric structure seem to be more suitable. Thus, we include stress tests results only for these entail the two properties above. Once the stress event is imposed on the system at time T (i.e. the innovation at time T faces an artificial shock which has a magnitude defined in section 6.1), conditional variance at time T+1 will increase. This process can be defined for simple EGARCH(1,1) process as:

$$\ln(\hat{\sigma}^2_{T+1}) = \hat{\xi}_1 + \hat{\xi}_2 \ln(\overline{\sigma}^2_T) + \hat{\gamma}\frac{\varepsilon_T}{\sqrt{\overline{\sigma}^2_T}} + \hat{\omega}\left[\frac{|\varepsilon_T|}{\sqrt{\overline{\sigma}^2_T}} - \sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right] \tag{10}$$

$$\varepsilon_T = \varepsilon^*_T \tag{11}$$

In the following day in stress horizon the evaluation proceeds with innovations drawn from selected distribution (Student's t here) and an appropriate variance:

$$\ln(\hat{\sigma}^2_{T+1+i}) = \hat{\xi}_1 + \hat{\xi}_2 \ln(\hat{\sigma}^2_{T+i}) + \hat{\gamma}\frac{\varepsilon_{T+i}}{\sqrt{\hat{\sigma}^2_{T+i}}} + \hat{\omega}\left[\frac{|\varepsilon_{T+i}|}{\sqrt{\hat{\sigma}^2_{T+i}}} - \sqrt{\frac{v-2}{\pi}}\frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})}\right] \tag{12}$$

$$\varepsilon_{T+i}(v/(v-2))^{0.5} \sim t_v \tag{13}$$

A similar process can be defined for the conditional empirical model where the initial shock and succeeding innovations are drawn from the sample of scaled returns. We do not include details for reasons of space. Both processes above are iterated until we have *s* returns, i.e. until we have an initializing stress innovation and a return for every single day in the risk horizon, and then these returns are aggregated to obtain a single *s-day* return. However, this process should be simulated for a number of potential paths to explore plausible large price discontinuities. Thus, we simulate the same path for 50.000 times using Monte Carlo simulation to get 50.000 potential *s-day* returns. Then, for a long (short) portfolio, absolute value of the lower (upper) 99% percentile of the sample returns is selected to reflect the loss that we are 99% confident that we will not lose more over *s* days.

7. Results and implications for risk capital

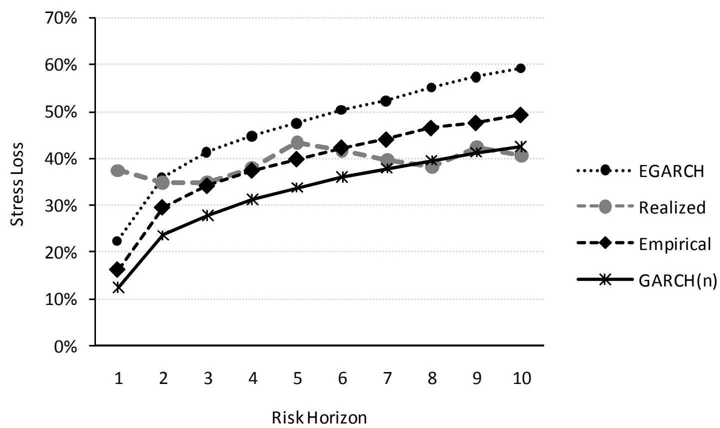
In order to be able to read outcomes more easily, below the basic parameters that characterizes a stress test are restated:

- α_s is the specified percentile for the size of the initial shock (i.e. lower or upper $100*\alpha_s$

percentile of the selected distribution),

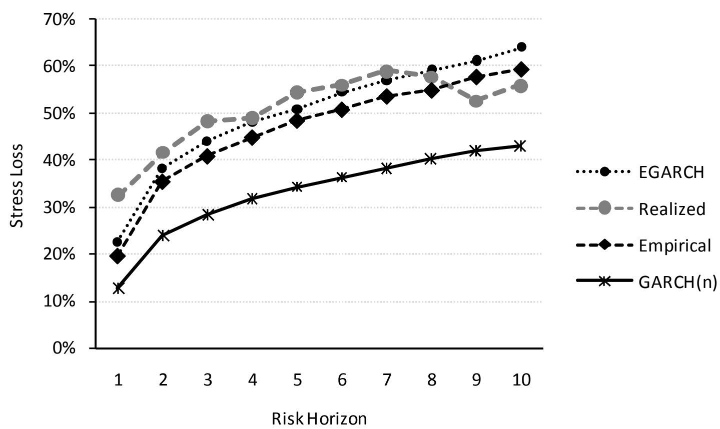
- ❖ s is the number of holding days where $s=1$ is the first day in the risk horizon. The initial stress event (plus the mean equation constant, if any) is equal to the return of that day.

Figure-1: Long natural gas, evaluation of simulated stress loss over 10 days ($\alpha_s = 0.0005$)



The results of a 10-day stress test for the long and short natural gas portfolios are shown in figures 1-4. Outcomes of the stress tests based on the two most preferable volatility models are depicted. Although the backtest results do not justify the use of the symmetric conditional normal model, the potential effects of a stress event under normality assumption are also included for comparison. The results in figures 1 and 2 for relatively smaller initial shocks are found unsatisfactory since the estimations mostly stay below the historical shocks and the best-performing two risk models appear to yield stress losses which are insufficient to compare with past shocks favorably. However, when a larger initial shock is considered (figures 3-4), stress loss estimations consistently match those observed historically. In long portfolio case, the conditional empirical model predicts a risk horizon that is apparently more close to historical realizations. The conservative stress losses estimated by the EGARCH model seem to be strongly affected by the greater initial shock defined by the model. It is also worth pointing out that, when compared to short portfolio case, the EGARCH model predicts a considerably lower stress loss in the long portfolio case. This is likely to result from the positive leverage effect estimated by the model that causes the negative initial shocks described for a long portfolio to die away in relatively shorter periods. This flexibility in defining a leverage effect obviously allows the EGARCH model to predict a much more close stress losses to those of conditional empirical model in the short portfolio case (Figure-4) which are still consistent with the past returns.

Figure-2: Short natural gas, evaluation of simulated stress loss over 10 days ($\alpha_s = 0.0005$)



The results also indicate that, given an initial shock varying from 18.46% to 27.59%, the loss for the long natural gas portfolio will 99% confidently range between 36.75% and 45.60% over 3 days, 52.73% and 63.10% over 10 days depending on the model selected. For a short natural gas, expected loss fluctuates in a narrower interval varying from 48.89% to 52.25% over 3 days and from 68.24% to 71.91 over 10 days. Similar interpretations can be put on the results for long/short crude oil and heating oil portfolios[3]. Another important point is that the poor backtest results for the symmetric conditional normal model are also reflected in the

[3] Full results for crude and heating oil are available upon request from the authors.

stress loss results. This model clearly underestimates risks and provides much lower losses during the simulated stress horizon which cannot be found reasonable when compared to past realizations.

Figure-3: Long natural gas, evaluation of simulated stress loss over 10 days ($\alpha_s = 0.0002$)

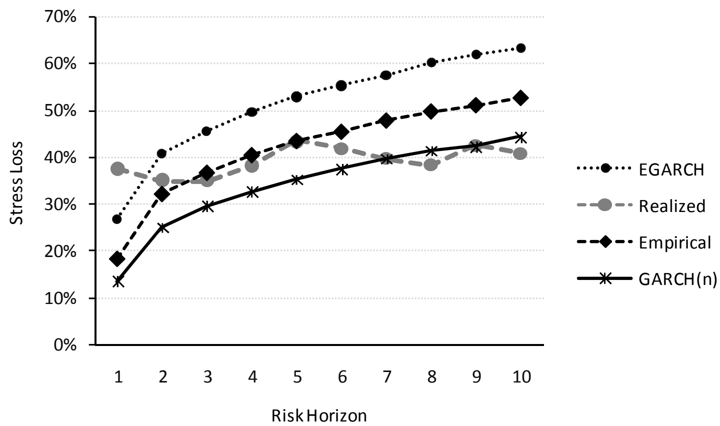
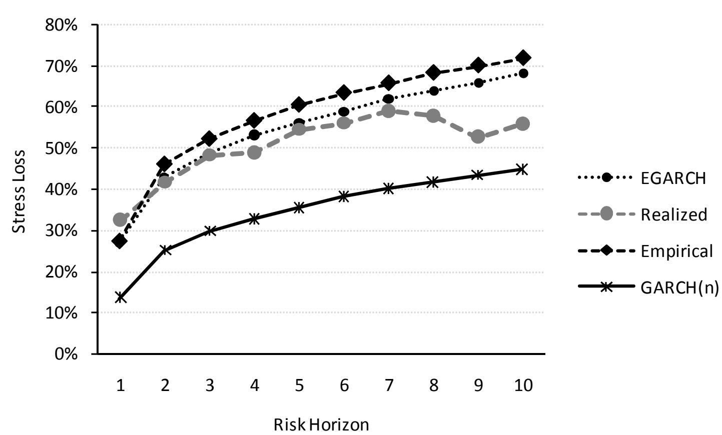


Table-5 exhibits the stress test results for all possible combinations of risk models, portfolios and characterizing parameters. The maximum historical losses occurred throughout the sample period are also depicted. To illustrate, consider short crude oil case with an initial shock reflecting the typical upper $\alpha_s = 0.0005$ or $\alpha_s = 0.0002$ percentiles of the selected distribution. According to the EGARCH model, depending on an initial stress loss ranging from a 15.34% to 18.38%, we are 99% confident that the upper bound for stress loss will vary from 28.43% to 31.99% over 3 days and from 40.61% to 44.85% over 10 days. It is also worth noting that, in some cases, models appear to be too conservative when compared to the historical realizations over the sample period. Thus, the question of whether the stress testing methodology is too sensitive to the changes in volatility of the underlying series springs to mind.

The results mark higher bounds for maximum potential losses in the most liquid energy contracts than those observed in major exchange rates. This result is particularly important for institutions that are highly engaged in energy-derivatives for a number of purposes varying from hedging physical positions to trading assets for capital gains. Besides, there are some important risk capital implications embedded in the stress test results. Here we refer to Basel II capital rules which recommend a more direct link between stress tests and risk capital. Taking account of (i) reduced market liquidity in a prolonged period of severe market conditions, (ii) the size of the portfolio and (iii) potential delays in managerial reactions, we use a 10-day risk horizon[4] to draw some conclusions for the minimum capital requirements. The 10-day stress loss estimations imply a lower bound for capital that must be set aside for single asset energy portfolios ranging between 45%-78% of the portfolio value, which means that significant capital buffers are needed.

Figure-4: Short natural gas, evaluation of simulated stress loss over 10 days ($\alpha_s = 0.0002$)



[4] Basel II Capital Accord recommends the use of an instantaneous price shock for regulatory capital calculations which is equivalent to a 10-day movement in prices (Prg. 718.LXXI, 2004)

Table-5: Stress test results

		EGARCH	Empirical
long Crude Oil	*Maximum historical loss over 3 days:* ***43.18%,*** *over 10 days=**42.82%***		
	Stress loss (99% confidence)		
	s=3,alpha=0.0005	30.33%	46.23%
	s=3,alpha=0.0002	34.02%	59.77%
	s=10,alpha=0.0005	43.97%	63.87%
	s=10,alpha=0.0002	48.47%	77.77%
short Crude Oil	*Maximum historical loss over 3 days: 32.50%, over 10 days=38.94%*		
	Stress loss (99% confidence)		
	s=3,alpha=0.0005	28.43%	40.99%
	s=3,alpha=0.0002	31.99%	43.16%
	s=10,alpha=0.0005	40.61%	56.41%
	s=10,alpha=0.0002	44.85%	59.33%
long Natural Gas	*Maximum historical loss over 3 days: 35.06%, over 10 days=40.86%*		
	Stress loss (99% confidence)		
	s=3,alpha=0.0005	41.37%	34.12%
	s=3,alpha=0.0002	45.60%	36.75%
	s=10,alpha=0.0005	59.11%	49.23%
	s=10,alpha=0.0002	63.10%	52.73%
short Natural Gas	*Maximum historical loss over 3 days: 48.31%, over 10 days=55.90%*		
	Stress loss (99% confidence)		
	s=3,alpha=0.0005	44.08%	40.86%
	s=3,alpha=0.0002	48.89%	52.25%
	s=10,alpha=0.0005	63.95%	59.38%
	s=10,alpha=0.0002	68.24%	71.91%
long Heating Oil	*Maximum historical loss over 3 days:* ***37.56%****, over 10 days=40.13%*		
	Stress loss (99% confidence)		
	s=3,alpha=0.0005	30.43%	41.83%
	s=3,alpha=0.0002	34.74%	45.47%
	s=10,alpha=0.0005	45.92%	58.70%
	s=10,alpha=0.0002	51.31%	62.56%
short Heating Oil	*Maximum historical loss over 3 days: 31.87%, over 10 days=48.99%*		
	Stress loss (99% confidence)		
	s=3,alpha=0.0005	31.80%	30.47%
	s=3,alpha=0.0002	36.53%	32.32%
	s=10,alpha=0.0005	48.73%	44.46%
	s=10,alpha=0.0002	53.96%	46.82%

Note: Maximum historical losses in bold observed at the beginning of Gulf War in January 1991.

8. Conclusions

Given the current volatile nature of energy markets, firms must indispensably perform stress tests of their portfolios and hold the necessary capital to cover their test results to avoid the undesirable results of unanticipated but still possible events. Indeed, under the new capital accord, financial institutions are required by regulators to establish a stress test framework as one of the main components of daily risk management activities. However, most of the stress testing approaches are still exposed to risk model misspecification threat and even some of them have yet to have an underlying risk model. According to Perignon and Smith (2006), unconditional historical simulation method is currently the most popular VaR methodology in the industry, which is the most likely to be misspecified and unsuited for stress testing purposes, even modeled conditionally. The studies by Berkowitz and O'Brien (2002) and Berkowitz et al. (2006) reveals the probability that firms' risk models may be misspecified since they do not seem sufficiently sensitive to the fluctuations in volatility over time.

Therefore, this study first aimed at identifying the best performing risk model together with the most suitable volatility process. We performed the backtests of both heavy-tailed and normal, as well as empirical, return assumptions under five models for symmetric and asymmetric volatility responses. Analyses were performed using daily returns for crude oil, natural gas and heating oil futures contract portfolios ranging from 16 to 27 years of observations. Stress losses are calculated for 60 possible asset position/risk model/initial shock size combinations.

We found strong evidence that the conditional empirical model outperformed other alternatives in relatively longer estimation windows (1000 and 2000 days), while the asymmetric EGARCH model was found superior in a smaller estimation window of 250 trading days. The well-performing two risk models are then used to model the consequences of a stress event via Monte Carlo simulation method. We observed that, in combination with an appropriate initial shock, the selected models provide (in some cases) conservative but still favorable results when compared to the historical realizations. Given an initial shock corresponding to a typical value of lower (upper) $\alpha_s = 0.0002$ percentile of selected distribution for the innovations, in only 2 out of 24 cases gave the selected models inconsistent results with past realizations. Both cases are characterized by the steep decline at the beginning of the Gulf War in 1991. It can also be inferred from Table-5 that an initial shock that corresponds to the lower (upper) $\alpha_s = 0.0005$ percentile of chosen distribution is inadequate to reflect potential shocks, because, in most of the cases, the calculated losses during a stress period obviously cannot cover the price shocks experienced historically. Stress losses estimated by the outperforming models yield significant risk capital requirements for energy portfolios, ranging between 48% and 77% of the portfolio value, and clearly point the market regulators towards more strict capital rules for energy traders as well as financial institutions that are intensively involved in energy markets.

Finally, our study reveals a number of superiorities that this model-based stress testing framework has over traditional methods. In traditional methods where hypothetical stress scenarios are employed, risk managers may overestimate, underestimate or even ignore the potential for some scenarios. That makes them inevitably subjective. Model-based approach apparently eliminates this problem and is much more objective. Another problem with the traditional stress tests is that the results are difficult to interpret because they give us no idea of the probabilities of the events concerned. Our model avoids this complexity by allowing the management to fix a probability for an extreme shock accordingly with the company's risk attitude and by providing the risk manager with the clear understanding of the potential risks surrounding the firm. Furthermore, the model-based framework can easily be linked to risk management process. This is not quite easy in traditional method because market risk measurement is a probabilistic approach whereas the stress scenarios are discrete events without certain probabilities each of which is likely to generate different outcomes. After all, the model-based approach is still vulnerable to risk model specification and significant structural changes in the market. An example of this type of model risk can be illustrated with reference to the first Gulf War which was broken out in 1991. At that time, on January 16th – the day before the invasion started, the market priced a highly optimistic scenario according to which the investors foresee that the war would be short enough with little or no damage to the oilfields in the region. This led to the largest historical drops in the futures market, with a daily decrease around 40% in oil prices, which our stress loss estimations could not capture favorably.

To conclude, it would be appealing to extend the discussion to modeling of tighter liquidity conditions in a more stressful market, which could be helpful in the selection of the minimum risk horizon to hedge positions. Also, the potential implications of an *s-day* stress loss for capital adequacy creates further incentive explore multi-asset portfolio cases where the correlations between assets become important and a comparison with a single appropriate VaR-based regulatory capital level would then be possible.

9. References

Alexander, C. and Sheedy, E. (2008) Developing a Stress Testing Framework Based on Conditional Market Risk Models, *Journal of Banking & Finance*, 32, pp. 2220–2236.

Aragones, J., Blanco, C. and Dowd, K. (2001) Incorporating Stress Tests into Market Risk Modeling, *Derivatives Quarterly*, Spring, pp. 44–49.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K. (1998) Don't Look Back, *Risk*, 11, pp. 100–103.

Basel Committee on Banking Supervision (1996) Amendment to the Capital Accord to Incorporate Market Risks, Bank for International Settlements, *Working Paper*.

Basel Committee on Banking Supervision (2004) International Convergence of Capital Measurement and Capital Standards (Revised Framework), Bank for International Settlements, *Working Paper*, June.

Berkowitz, J., Christoffersen, P., Pelletier, D. (2006) *Evaluating VaR Models with Desk-level Data*, McGill University Press, Canada.

Berkowitz, J. and O'Brien, J. (2002) How Accurate are Value-at-Risk Models at Commercial Banks?, *Journal of Finance*, vol. LVII 3, pp. 1093-1111.

Bollerslev, T., (1986) Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics*, 31, pp. 307–327.

Brooks, C., (2007) *Introductory Econometrics for Finance*, Cambridge University Press, England.

Christoffersen, P. (1998) Evaluating Interval Forecasts, *International Economic Review*, 39, pp. 841–862.

Committee on the Global Financial System (2005) Stress Testing at Major Financial Institutions: Survey Results and Practice, Bank for International Settlements, *Working Paper*, January.

Frey, R. and Michaud, P. (1997) The effect of GARCH-type Volatilities on Prices and Payoff Distributions of Derivative Assets—a Simulation Study, ETH Zurich, *Working Paper*.

Hansen, P. R. and Lunde, A. (2005) A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, *Journal of Applied Econometrics*, 20, pp. 873-889.

Hoppe, R. (1998) VAR and the Unreal World, *Risk*, 11, pp. 45–50.

Kupiec, P. (1995) Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivatives*, Winter, pp. 73–84.

Kupiec, P. (1998) Stress Testing in a Value-at-Risk Framework, *Journal of Derivatives*, Fall, pp. 7–24.

Nelson, D. B. (1991) Conditional Heteroscedasticity in Asset Returns: A New Approach, *Econometrica*, 59(2), pp. 347-370.

Perignon, C. and Smith, D. (2006) The Level and Quality of Value-at-Risk. Disclosure by Commercial Banks, Simon Fraser University, Vancouver.

❖