Yihui Lan (Australia), Ze Min Hu (Australia), Jackie Johnson (Australia)

Improving accuracy and precision of value-at-risk forecasts

Abstract

Value-at-risk (VaR) models are intended to measure the relationships among many uncertainties. This paper focuses on ways to improve accuracy and precision of VaR forecasts. Unlike most previous studies that are only concerned with the tail behavior of predicted returns, this paper proposes a new methodology to incorporate a number of sources of resampling uncertainty in VaR forecasts. The study illustrates the methodology using the filtered historical simulation model. This new approach employs both the bootstrap and its close alternative, the jackknife method. In particular, the delete-d jackknife is adopted as it is specifically designed for non-smooth statistics such as the quantile. The delete-d jackknife has the attraction of producing unbiased statistics since it resamples from the original distribution rather than from the empirical distribution as in the bootstrap. Applied to five return series, the proposed technique is shown to provide more accurate VaR forecasts than the other eight models in terms of statistical loss measures. In addition, they provide reasonable improvement over the other eight models in terms of statistical and regulatory tests. In particular, considerable improvement is achieved in terms of forecast precision.

Keywords: value-at-risk, forecast accuracy, forecast precision, filtered historical simulation, jackknife.
JEL Classification: C53, C63, G10.

"*Blame for the financial crisis cannot legitimately be laid on the doorstep of the risk models. Models are designed to reflect reality and prepare us for anticipated future states of reality. … Blaming a VaR model for causing the crisis demonstrates an imperfect understanding of the nature of the model and how it's meant to be used*".

"Don't blame crisis on risk models"
January 20, 2010, American Banker

Introduction

Value-at-risk (VaR) is one of the few risk metrics available and the industry standard due to regulatory requirements. It is defined as the worst loss in the market value of a portfolio expected over the target time horizon under normal market conditions for a given level of confidence. Made public in 1994 by J.P. Morgan, VaR had attracted increasing research attention and undergone continuous refinement until the 2007-8 global financial crisis, or the Great Recession. Financial crises such as the recent one are highly unusual events when they occur. When this crisis struke, VaR has been widely regarded as the major culprit. On statistical grounds, VaR bears criticisms such as: (1) violations of assumptions; (2) inadequate account for extreme losses; (3) being a non-coherent risk measure (Giannopoulos and Tunaru, 2005; Szegö, 2002). There are now two camps in the risk management community: the mainstream view is that VaR is to blame for the recent financial crisis and should be consigned to the dust bin. A minority argues that risk measures like VaR are expected not to foretell crises, but to predict the risk of losses. They claim that VaR does have virtue and model risk and human judgment should be considered when any risk model is used. Regulatory responses are to improve upon VaR by considering periods of increased volatility (BCBS, 2009).

The VaR concept became popular due to its aggregation of all risks of a portfolio into a single number. Though the concept of VaR is simple to understand, VaR modelling is a challenging task. There are three categories of VaR models: parametric, non-parametric, and semi-parametric models. Given that no VaR model has been proved adequate for all financial assets, sampling frequencies, trading positions, confidence levels and sub-periods (Angelidis et al., 2007), this paper does not intend to provide a panacea. Instead, it focuses on two important characteristics associated with VaR forecasts – accuracy and precision. They are not interchangeable terms as accuracy refers to the correctness while precision refers to the degree of uncertainty associated with the forecasts. Consider the example of weather forecasts. If it is forecast to be between 10 and 30 degrees at noon today and the actual reading turns out to be 23, then the forecast is *accurate*, but not very *precise.* The forecaster provided a true statement but without enough detail for people to make plans. If the forecast is 15 degrees and it turns out to be 23 degrees, this forecast was very *precise*, but completely *inaccurate.* This example illustrates that precision is not very useful without accuracy, and that accuracy with little precision does not tell much either. Although considerable research is devoted to producing accurate values of VaR forecasts, it is of tantamount importance that the forecasts are not only unbiased, but also accompanied by some uncertainty measure, which is usually reported in terms of variances or confidence intervals

© Yihui Lan, Ze Min Hu, Jackie Johnson, 2010.
We would like to acknowledge the financial support from the Business School at the University of Western Australia (UWA) and a UWA research grant. For helpful comments, we thank Kenneth Clements, Takayasu Ito and participants at the 2008 AFAANZ Conference. The remaining errors are our own.

of the forecasts. Providing both point estimates and uncertainty measures of VaR forecasts enables risk managers to attach a certain level of confidence to these estimates and thus make more cautious decisions, especially in a turbulent financial environment.

The contributions of the paper are twofold. First, we propose a new methodology to incorporate a number of sources of resampling uncertainty in VaR forecasting. Instead of being only concerned with the tail behavior of predicted returns, as most studies do, we identify each potential source of uncertainty during the modelling procedure and address them simultaneously. Given the vast number of VaR models proposed in the literature, we illustrate the basic idea behind our methodology using the filtered historical simulation (FHS) model. FHS is a natural candidate as there is more than one source of uncertainty inherent in this model. Our proposed methodology can be easily extended to other VaR models. Second, we resurrect the delete-d jackknife in the context of VaR forecasting. Proposed by Wu (1986) and widely known as an approximation to the bootstrap, the jackknife draws samples from the true distribution, whereas the bootstrap draws samples from the empirical distri-bution. Hence jackknife samples resemble closer to the true distribution than bootstrap samples do.

Using Australian market and industry indices data, we show that our proposed methodology produces more precise VaR forecasts. It is to be noted that our approach yields not only the point estimates but also the entire distributions of VaR forecasts. We compare the VaR forecasts derived from ten models including the two variants of our proposed approach. The results show that our two models provide more accurate VaR forecasts than the other eight models in terms of statistical loss measures. In addition, they provide reasonable improvement over the other eight models in terms of statistical and regulatory tests. The rest of the paper is structured as follows. Section 1 gives a brief review of VaR modelling and FHS, and in particular the literature on VaR accuracy and uncertainty. Section 2 discusses our proposed method of incorporating various sources of resampling uncertainty and the details of the delete-d jackknife. We show how to quantify various sources of uncertainty using the example of the FHS model. Section 3 presents the empirical results of VaR forecasts, the evaluation of our forecasts against eight other models, and further analysis based on the stress VaR. Summary and concluding remarks are provided in the last Section.

1. Literature review

As it involves challenging statistical problems, VaR modelling had received increasing research attention until the recent global financial crisis. Generally speaking, VaR models can be classified into parametric, non-parametric and semi-parametric models. Parametric models are comprised of the RiskMetrics model by J.P. Morgan, the variance-covariance approach that relies on normal distributions of returns, parametric approaches that use non-normal distributions, GARCH-type models and extreme value approaches, etc. The non-parametric approaches include historical simulation (also known as bootstrapping simulation), weighted historical simulation, some hybrid models, the use of nonparametric density estimation and neural network. Duffie and Pan (1997), Dowd (2005), Jorion (1996b; 2000; 2010) provide reviews of VaR models proposed in the last two decades.

The early VaR models mainly belong to the parametric category because of the parametric assumptions they impose on the data. The plain vanilla version of VaR is based on the assumption of normal returns, which is widely known to be unrealistic. A modified version, the Cornish-Fisher VaR (see, e.g., Campbell et al., 2001) adjusts the critical values of the standard normal distribution, but does not react to changes in the return process. There is mounting research that incorporates conditional volatility in VaR modelling. RiskMetrics by J.P. Morgan assumes that the return process follows an integrated GARCH(1,1) model without a drift. Later GARCH-type VaR models are based on a mixture of conditional mean processes, conditional variance specifications and conditional distribution of standardised or non-standardised returns. Angelidis et al. (2004), So and Yu (2006) and Lee et al. (2008) are some examples.

As a major non-parametric approach, historical simulation (HS) does not require any statistical assumption about return distributions. Its empirical performance is examined in Beder (1995) and Hendricks (1996) among others. One disadvantage of HS is that it assigns an equal probability weight to each historical return, which is equivalent to assuming returns are independently and identically distributed (i.i.d.) through time. This is unrealistic because return volatility is time-varying and tends to be clustered. The weighted historical simulation, proposed by Boudoukh et al. (1998), places more weight on recent returns and calculates VaR from the empirical distribution of the re-weighted returns.

Since the late 1990s the semi-parametric models based on a mixture of parametric and non-parametric methods have gained momentum, if not being the mainstream of VaR research. Filtered historical simulation (FHS), proposed by Barone-Adesi et al. (1998; 1999) and Barone-Adesi and Giannopoulos (2001), is one of the popular models in this category. It combines GARCH volatility forecasting and bootstrap simulation. Pristker (2006) reviews the assumptions and limitations of weighted historical simulation and FHS. Another semi-parametric approach, GARCH-EVT, is different from FHS in that it uses extreme value theory to model the tails of residual distributions. Other semi-parametric models include Fan and Gu (2003) who estimate volatility and quantiles of returns, and regression quantile techniques such as Engel and Manganelli (2004) and Chernozhukov and Umantsev (2001).

Kim and Hardy (2007), Hartz et al. (2006) and Inui et al. (2005) are among the small number of studies that examine and correct for the bias of VaR forecasts. However, the research on the uncertainty of VaR is limited. VaR uncertainty is usually obtained from Monte Carlo simulations and assumptions on the distributions of profit and loss, or VaR estimation parameters (see, e.g., Bams et al., 2005). Other approaches to estimating VaR uncertainty include analytical derivations (see, e.g., Chappell and Dowd, 1999; Jorion, 1996a), the theory of order statistics (Dowd and Blake, 2001), and neural networks. Aussenegg and Miazhynskaia (2006) accounts for estimation uncertainty by providing VaR forecast distributions. Basal and Staum (2008) propose three methods to construct confidence intervals and regions for VaR and expected shortfall, the latter of which is the coherent alternative to VaR. Pasaran et al. (2009) argues that most VaR models are concerned with the tail behaviors of predicted returns and they deal with the problem of model uncertainty by model averaging. Kerkhof et al. (2010) propose a specific procedure to take model risk into account in the computation of capital reserves. In this paper, we do not intend to deal with model risk and only concentrate on incorporating various sources of resampling uncertainty inherent in VaR modelling.

2. Methodology

The focus of the VaR forecasting is on the conditional quantile of returns $r^{(\lambda)}$ in

$$P\left(r_{t+\tau} \le r^{(\lambda)} \middle| I_t\right) = \lambda,$$

where r_t is the logarithmic return, I_t is the information available at time *t*, and the probability λ is usually specified to be one or five percent. As mentioned in the introduction to the paper, most studies in the literature are only concerned with the tail behavior of predicted returns. We are interested in not only the uncertainty associated with the conditional mean return $\mu_t = E(r_t \mid I_{t-1})$, but also other potential sources of resampling uncertainty during the modelling procedure. Specifically our new approach tries to address these sources of uncertainty simultaneously. In addition, we propose to use a resampling technique alternative to the bootstrap – the jackknife. Below we use the filtered historical simulation (FHS) model as an example to illustrate our proposed methodology[1].

The semi-parametric approach FHS is carried out in two steps. In the first step, parametric GARCH-type models are fitted for returns:

$$r_t = \mu_t + \varepsilon_t, \tag{1}$$

where conditional mean return μ_t is usually modelled by ARMA models. The conditional variance h_t of residuals ε_t is chosen empirically and has the general GARCH form

$$h_t = f\left(\varepsilon_{t-1}, \mathrm{h}_{t-1}\right). \tag{2}$$

The standardised residuals $z_t = \varepsilon_t / \sqrt{h_t}$ are i.i.d. random variables. Note that the above univariate GARCH can be easily extended to the multivariate case with $z_t = \Sigma_t^{-\frac{1}{2}} \varepsilon_t$, where $\Sigma_t = Cov\left(\varepsilon_t \middle| I_t\right)$. The second step of FHS uses the non-parametric bootstrap to resample standardised GARCH residuals. The conditional return at time $t+i$ is simulated from:

$$r_{t+1}^b = \hat{\mu}_{t+i} + z_i^b \sqrt{\hat{h}_{t+i}^b}, \tag{3}$$

$$\hat{h}_{t+i}^b = f\left(\varepsilon_{t+i-1}^b, \hat{h}_{t+i-1}^b\right), \tag{4}$$

where $i = 1,...,\tau$ is the i^{th} forecast in the forecasting period, $b = 1,...,B$ with *B* being the total number of bootstrap simulation experiments, $z^b = \{z_1^b, z_2^b, ..., z_\tau^b\}$ is the bootstrapped standardised residual vector sampled from $z_i^b \in \left\{\hat{z}_1, \hat{z}_2, ..., \hat{z}_T\right\}$, and the circumflex denotes fitted values. Note that the simulated volatility, $\hat{h}_{t+i}^b$, is obtained recursively based on $\varepsilon_{t+i-1}^b = z_{i-1}^b \sqrt{\hat{h}_{t+i-1}^b}$. An attractive feature of FHS

[1] Supporting evidence for FHS is found in Angenidis and Benos (2008), Jayasuriya and Rissiter (2008), Angenidis et al. (2007) and Pritsker (2006). Giannopoulos and Tunaru (2005) first extend FHS to the calculation of expected shortfall.

is that the simulated return r^b_{t+i} takes into account the simulated volatility forecast $\hat{h}^b_{t+i}$ conditioning on their past values. The VaR forecast at horizon $t+i$ is the λ-quantile of the distribution of simulated conditional returns, $r^{(\lambda)}_{t+i}$.

As can be seen, FHS has the limitation that it only models one source of randomness – that comes from the filtered returns, i.e. the standardised residuals. In fact, the VaR forecast obtained from the distribution of simulated returns, which is a quantile estimate, also gives rise to estimation uncertainty. Since there are two sources of uncertainty, our suggested approach incorporates both of them. We use different resampling techniques to examine two variants of our approach. Equation (3) shows that the VaR forecast is the sum of the conditional mean and conditional shock, the latter of which is the product of the simulated standardised residuals and the simulated volatility, both based on the bootstrap. In our proposed to approach, we employ the bootstrap as well as its alternative – the Jackknife, and apply the combinations of these resampling techniques to the improved FHS model.

Future returns are simulated in FHS by bootstrapping from rescaled historical shocks. A bootstrapped sample resembles the empirical sample in the same way as the empirical sample resembles its underlying distribution, especially when the sample size is large. One of the limitations of the bootstrap is that it resamples from the empirical data (Davison and Hinkly, 1997; Efron, 1979). Another limitation is associated with the total number of different combinations of bootstrapped residuals. Barone-Adesi and Giannopoulos (2001) claim that this number can be very large if the forecast horizon is ten days. This is true for the multiple holding period, e.g., ten days. However, if the forecast horizon is one day, the maximum number of distinct bootstrapped residuals is the sample size of the estimation period. Jackknifing largely avoids these potential problems. Jackknife statistics are created by systematically dropping subsets of data at a time and assessing the resulting variation in the parameter in question (Mooney and Duval, 1993). A distinctive feature of the jackknife is that its sample presents a closer representation to the data's original distribution than the typical bootstrap sample does. The jackknife method is usually used for estimation, adjustment for bias and for derivation of robust estimates of standard errors and confidence intervals (Efron and Tibshirani, 1993). The standard jackknife is known as "the method of leaving one out", i.e., the delete-1 jackknife. Suppose $X=(X_1,X_2,...,X_T)$, the k^{th} jackknife sample then consists of the dataset with the k^{th} observation removed, i.e. $X_{[k]}=(X_1,X_2,...,X_{k-1},X_{k+1},...,X_T)$, for $k=1,...,T$. The jackknife is often used as a simple and good approximation to the bootstrap. In the delete-1 jackknife, only T jackknife samples are available for investigation. But there is no loss of information compared to the bootstrap. While applicable to quantile estimates such as VaR, the standard delete-1 jackknife estimate may be inconsistent if the estimate is not smooth, i.e. a small change in the data can cause a large change in the statistic. Such non-smooth statistics include the median and percentiles. A more generalised jackknife method, the delete-d jackknife proposed by Wu (1986), is able to solve this problem of inconsistency. Rather than leaving out one observation at a time, we omit d observations, where d can be any integer satisfying $\sqrt{T}<d<T$. Unlike its standard counterpart, which has in total only T jackknife samples, the number of possible subsamples in the delete-d jackknife is $P(T,d)=T!/(d!(T-d)!)$. As the jackknife samples from the original observed sample and the delete-d jackknife is specially designed for non-smooth statistics such as quantiles, the delete-d jackknife is a natural candidate for quantile estimation. There are several alternative quantile estimators. We use the type-8 quantile estimator proposed by Hyndman and Fan (1996), which is found by Kim and Hardy (2007) to perform better than other quantile estimators.

The accuracy of VaR forecasts is of great importance not only to the financial institution itself but also to its regulators. To evaluate the performance of VaR forecasts, we consider eight evaluation criteria that are commonly used in the literature: mean absolute error (MAE); mean squared error (MSE); mean relative scaled bias (MRSB); the likelihood ratio tests of unconditional coverage, independence, conditional coverage (Christofferson, 1998; 2003); point estimator of probability of violation; and regulatory backtesting.

3. Empirical results

This Section contains the VaR forecasts of five portfolios, the comparison of our two proposed models against eight other models, and how our approach performs based on the stress VaR.

3.1. VaR forecasts. In our empirical study, we examine daily returns of five portfolios and use them as proxies for mutual funds with distinct investment perspectives. Our objective is to shed light on whether the proposed approach improves VaR forecast accuracy and precision. As different funds have quite diverse and sometimes complicated objectives, we largely ignore such diversity and complexity and confine our investigation to these return series. Portfolio 1 is a market index fund proxied by the S&P/ASX200 index, which covers approximately 80 percent of the Australian equity market by capitalization. Portfolio 2, the S&P/ASX50, is used as a proxy for a fund investing in firms with large market capitalisation and pursuing the objective of providing long term capital appreciation. Portfolio 3 uses the resources index to mimic a fund that captures the benefits of the resources industry boom in Australia. Portfolios 4 and five imitate the returns of low and high risk funds respectively[1]. The six-year trading period data from April, 3 2000 to April, 2 2006 are collected from Datastream. The sample contains 1,518 observations. Table 1 presents the summary statistics. It can be seen that all returns series are non-normal (column 7). The mean returns of portfolios one and two are insignificant. In addition, portfolios 1, 2 and 5 are leptokurtic, and all portfolios except four are significantly skewed to the left (columns 5 and 6).

Table 1. Summary statistics

Portfolio	Style	μ (×100)	σ (×100)	Skewness	Kurtosis	Bera-Jarque Statistic
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Market	0.213	6.81	-0.79**	9.88**	5,268**
2	Large cap	0.197	7.35	-0.61**	8.47**	3,854**
3	Booming	0.672*	10.98	-0.15*	3.82	773**
4	Low risk	0.348**	5.19	-0.02	3.70	721**
5	High risk	-1.315**	15.48	-2.20**	25.38**	34,917**

Notes: The 1% and 5% significance levels are indicated by ** and * respectively. The Bera-Jarque statistic is asymptotically distributed as $\chi^2_{(2)}$ under the null of normality.

We divide the whole sample into the estimation period of $T = 1{,}263$ observations spanning over five years[2], and the test period of 255 days. We use $j = 1...10{,}000$ experiments to simulate returns for the forecast period of 255 days. The model estimates are updated by using a rolling window with the same window length as the estimation sample, $T = 1{,}263$. Thus, at time $T + i$, the rolling window consists of i values of simulated returns and T most recent returns in the sample, where $i = 1,...,\tau$ is the i^{th} forecast in the forecasting period.

Regarding the specification for the mean equation (1), we determine the orders of AR and MA using ACF and PACF for each of the five portfolios. Other studies, such as Hull and White (1998) and Barone-Adesi et al. (1998), use the constant conditional mean process. Angelidis et al. (2004) argue that the mean process specification plays no important role. We thus examine both the case of a constant and a time-varying conditional mean process. To find out which GARCH-type model best fits the conditional variance and which distribution best describes the fat-tailedness of GARCH-filtered residuals, we consider four types of GARCH – GARCH, AGARCH, EGARCH, PGARCH, and three types of residual distributions – the normal, student-t, and generalised error distributions. We choose a one-day holding period for VaR forecasting ($\tau = 1$), and thus produce 255 one-day ahead forecasts N = 255. Out of the twelve GARCH models, AGARCH with the t-distribution is found to be the model with the best fit based on the AIC, BIC and log-likelihood criteria, and is thus chosen for model estimation. It takes the form of $h_1 = \omega + \alpha(\varepsilon_{t-1} + \eta)^2 + \beta h_{t-1}$. The asymmetric effects on volatility are modelled via the parameter η, the estimate of which is found to be statistically significant for all portfolios in this study.

Figure 1 presents the distributions of VaR forecasts for portfolio one over the 255-day forecasting period in the form of three-dimensional graphs[3]. Such entire distributions give us a complete picture about forecast uncertainty. The left two panels give 1%and 5% VaRs from the ARMA conditional mean model based on jackknifing (hereafter Jack_ARMA) and the right two panels presents 1% and 5% VaRs from

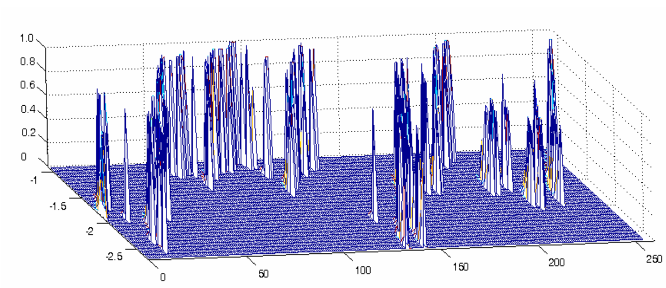
[1] To maintain simplicity, portfolios 4 and 5 contain only two industry indices. Based on the ex-post returns data of fourteen Australian industry indices, we choose two indices with the lowest variances and two indices with the highest variances to form portfolios 4 and 5 respectively. The weights of the two indices contained in each portfolio are then determined to minimise portfolio variances.

[2] The estimation periods of five to ten years are very common among the VaR literature. For example, Brooks and Persand (2003) use 5 years, and Barone-Adesi et al. (1998) use 10 years. Kuester et al. (2006) use a window size of 1,000 days, which are about four years of trading data. They also consider smaller window sizes for sensitivity analysis. It is to be noted that a large window size reduces the variability of estimated parameters, but increases modeling biases and thus approximation errors (Fan and Gijbels, 1996).

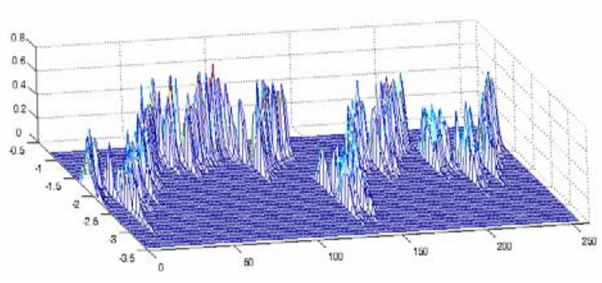
[3] Plots for other portfolios based on other models are qualitatively similar and available from the authors upon request.

the ARMA conditional mean model based on bootstrapping (hereafter Boot_ARMA). Visually speaking, graphs on the left and right reveal clearly different patterns. The distributions from Boot_ARMA spread out more, with frequency counts in continuous ranges; see Panels B and D. The VaR forecasts produced from the Jack_ARMA, on the other hand, are mostly concentrated in narrow ranges. This is shown by a smaller number of bins and a large number of frequency counts in each bin in Panels A and C. We can thus conclude that the new approach produces more precise VaR forecasts.

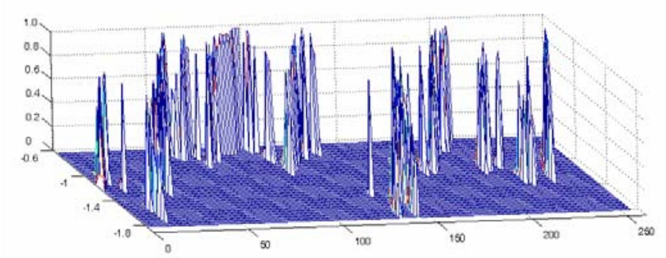
A. VaR 1% from Jack_ARMA



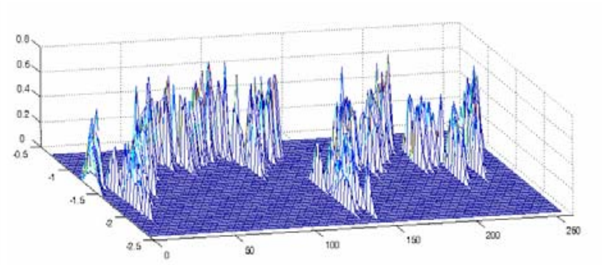
B. VaR 1% from Boot_ARMA



C. VaR 5% from Jack_ARMA



D. VaR 5% from Boot_ARMA



Note: The x-axis refers to forecasting day *t*, the y-axis gives the VaR forecasts in terms of logarithmic returns (×100), and the z-axis presents the probability.

Fig. 1. The distributions of VaR forecasts: portfolio 1

3.2. Comparison of forecast performance. In addition to Jack_ARMA, we also forecast VaR based on the constant conditional mean model based on jackknifing (hereafter Jack_Mean). To evaluate these two variants of our approach – Jack_Mean and Jack_ARMA, we use eight other competing models to make comparisons. As we are interested in how our approach performs relative to the original FHS model by Barone-Adesi et al. (1998), we consider four variants of the FHS. The first two competing models, Boot_Mean and Boot_ARMA, are assumed to have a constant and an ARMA mean respectively. Their difference from the original FHS is that rolling windows are used to update model estimates. The third competing model is simply the FHS whereby the VaR forecast is a single point estimate. Competing model four is an extension to the FHS, which does not provide a single VaR forecast but $B = 10{,}000$ bootstrapped forecast values. The basic idea is, for any forecast horizon, to repeat many times the procedure of FHS (Dowd, 2005, p.98, footnote 14, hereafter FHS_Down). The other four models are GARCH with normally-distributed residuals (GARCH_N), GARCH with student-t distributed residuals (GARCH_T), historical simulation (HS), and the variance-covariance (VCV) model.

The statistical loss measures of VaR forecasts from the ten competing models are contained in Table 2. In terms of the probability of violation, $\hat{\lambda}$, VaR forecasts from a majority of models are not too far away from the target values: $\lambda = 1\%$ in the left panel and $\lambda = 5\%$ in the right panel. Ceteris paribus, overestimation of VaR is preferred to underestimation – underestimation leaves losses uncovered, whereas a much-overestimated VaR leads to a lower return from the capital set aside. Therefore, if being right on target is difficult to achieve, slight overestimation is acceptable. As can be seen from columns 2 and 9, the two variants of jackknife FHS outperforms all other models for the high-risk portfolio, and performs as least as good as other models for the remaining four portfolios.

Table 2. Statistical loss measures of VaR models

Model	$\lambda = 1\%$							$\lambda = 5\%$							Sum of ranks
	$\hat{\lambda}$	MAE	Rank	MSE	Rank	MRSB	Rank	$\hat{\lambda}$	MAE	Rank	MSE	Rank	MRSB	Rank	
A. The market index (portfolio 1)															
1. Jack_Mean	1.18	1.62	1	3.15	3	-15.37	4	5.10	1.14	2	1.68	2	-2.84	4	16
2. Jack_ARMA	1.18	1.63	2	3.19	4	-16.07	2	5.49	1.14	1	1.68	1	-2.96	3	13
3. Boot_Mean	1.96	1.66	6	3.29	6	-15.38	3	5.88	1.18	4	1.80	5	-3.69	1	25
4. Boot_ARMA	1.96	1.63	3	3.20	5	-16.10	1	5.88	1.15	3	1.70	3	-3.24	2	17
5. FHS	0.78	2.07	9	4.78	9	7.45	7	5.10	1.28	8	2.00	9	1.49	6	48
6. FHS_Dowd	1.18	2.03	8	4.63	8	16.71	9	5.10	1.28	8	1.99	8	2.80	9	50
7. GARCH_N	1.57	1.64	4	3.13	2	13.37	8	5.49	1.22	6	1.84	7	2.92	10	37
8. GARCH_T	0.78	2.25	10	5.59	10	17.83	10	1.96	1.60	10	2.99	10	1.13	5	55
9. HS	1.18	1.78	7	3.57	7	3.37	5	5.88	1.20	5	1.75	4	1.62	7	35
10. VCV	1.96	1.64	4	3.06	1	4.18	6	5.49	1.22	6	1.81	6	2.77	8	31
B. The large-cap portfolio (portfolio 2)															
1. Jack_Mean	1.18	1.76	3	3.68	3	-10.99	4	5.49	1.21	1	1.89	2	-2.00	4	17
2. Jack_ARMA	1.18	1.76	4	3.70	4	-11.05	3	5.49	1.21	1	1.88	1	-2.13	3	16
3. Boot_Mean	1.18	1.78	5	3.78	5	-11.21	1	5.49	1.24	3	1.97	3	-3.29	1	18
4. Boot_ARMA	1.18	1.78	5	3.82	6	-11.20	2	5.49	1.25	4	1.99	5	-3.15	2	24
5. FHS	0.78	2.07	9	4.84	9	3.70	7	5.10	1.33	9	2.16	9	2.15	9	52
6. FHS_Dowd	0.78	2.04	8	4.72	8	2.30	6	5.10	1.32	8	2.14	8	1.49	8	46
7. GARCH_N	1.96	1.70	1	3.37	1	-5.13	5	5.49	1.26	5	1.98	4	0.42	7	23
8. GARCH_T	0.78	2.25	10	5.61	10	13.58	10	2.75	1.61	10	3.05	10	2.83	10	60
9. HS	1.18	1.90	7	4.02	7	3.84	8	5.10	1.28	6	1.99	5	-1.50	5	38
10. VCV	1.18	1.75	2	3.49	2	3.92	9	4.71	1.30	7	2.04	7	-0.10	6	33
C. The booming sector portfolio (portfolio 3)															
1. Jack_Mean	1.57	2.91	4	9.91	3	-10.74	2	5.49	2.01	2	5.23	2	-9.57	2	15
2. Jack_ARMA	1.57	2.93	5	10.06	5	-10.78	1	6.27	2.01	1	5.23	1	-9.62	1	14
3. Boot_Mean	1.18	2.96	6	10.38	6	-1.95	3	4.71	2.14	6	5.95	7	-1.76	4	32
4. Boot_ARMA	1.18	2.89	3	9.95	4	-1.45	4	4.71	2.06	3	5.59	4	-2.54	3	21
5. FHS	1.18	3.59	10	14.62	10	5.37	8	4.71	2.27	9	6.45	9	3.73	8	54
6. FHS_Dowd	1.96	3.52	9	14.11	9	2.55	6	4.71	2.26	8	6.43	8	3.39	6	46
7. GARCH_N	2.35	3.07	7	10.99	7	2.02	5	4.71	2.28	10	6.55	10	4.65	9	48
8. GARCH_T	3.14	2.60	1	8.16	1	3.28	7	6.28	2.07	4	5.53	3	3.71	7	23
9. HS	2.75	3.15	8	11.47	8	5.86	10	5.88	2.09	5	5.62	5	1.28	5	41
10. VCV	2.75	2.87	2	9.70	2	5.85	9	5.49	2.15	7	5.92	6	6.74	10	36
D. The low-risk portfolio (portfolio 4)															
1. Jack_Mean	1.96	1.27	1	1.98	2	-6.50	1	5.49	0.93	3	1.15	3	-1.79	2	12
2. Jack_ARMA	1.96	1.28	3	2.01	3	-5.54	3	5.88	0.93	2	1.15	2	-1.84	1	14
3. Boot_Mean	2.35	1.35	6	2.19	7	-4.97	4	6.67	0.97	6	1.23	6	-0.74	5	34
4. Boot_ARMA	2.35	1.33	5	2.13	5	-5.54	2	6.67	0.94	4	1.18	5	-1.23	3	24
5. FHS	0.78	1.59	9	2.90	9	8.14	10	6.28	1.02	8	1.33	8	-0.21	7	51
6. FHS_Dowd	0.78	1.57	8	2.85	8	7.08	8	6.28	1.02	8	1.33	8	0.50	8	48
7. GARCH_N	3.14	1.32	4	2.08	4	7.65	9	6.67	0.98	7	1.23	7	3.59	10	41
8. GARCH_T	0.39	1.73	10	3.38	10	3.29	7	2.75	1.29	10	2.00	10	3.14	9	56
9. HS	1.57	1.36	7	2.16	6	-1.59	6	6.28	0.93	1	1.11	1	-0.98	4	25
10. VCV	1.57	1.28	2	1.95	1	-2.03	5	6.28	0.95	5	1.16	4	-0.44	6	23
E. The high-risk portfolio (portfolio 5)															
1. Jack_Mean	0.57	3.62	3	15.10	4	3.79	9	4.08	2.24	1	6.62	1	-4.23	3	21
2. Jack_ARMA	0.58	3.71	6	15.82	6	4.36	10	4.10	2.30	4	6.92	4	-4.11	4	34
3. Boot_Mean	1.19	3.62	4	14.98	3	0.94	7	4.71	2.28	3	6.74	3	-7.09	2	22
4. Boot_ARMA	1.20	3.66	5	15.31	5	1.12	8	4.71	2.28	2	6.73	2	-7.13	1	23
5. FHS	1.39	5.28	10	30.34	10	-3.43	1	4.71	2.53	9	8.05	9	4.24	9	48
6. FHS_Dowd	1.39	5.16	9	28.95	9	-1.21	5	4.71	2.52	8	8.01	8	3.80	7	46

Table 2 (cont.). Statistical loss measures of VaR models

Model	$\lambda = 1\%$							$\lambda = 5\%$							Sum of ranks
	$\hat{\lambda}$	MAE	Rank	MSE	Rank	MRSB	Rank	$\hat{\lambda}$	MAE	Rank	MSE	Rank	MRSB	Rank	
7. GARCH_N	1.18	3.40	1	13.52	1	-2.19	2	4.71	2.47	6	7.77	6	1.39	5	21
8. GARCH_T	0.39	4.89	8	26.02	8	-0.57	6	2.35	3.05	10	11.16	10	5.36	10	52
9. HS	0.39	4.25	7	19.91	7	-1.44	3	5.49	2.39	5	7.34	5	3.57	6	33
10. VCV	1.18	3.44	2	13.69	2	-1.36	4	5.10	2.50	7	7.91	7	4.20	8	30

Notes: The entries in columns 2, 3, 9 and 10 are to be divided by 10^2, and those in columns 5 and 12 are to be divided by 10^4. The ARMA specifications in the conditional mean equation for each portfolio are: ARAM(2,2) for portfolio 1, AR(1) for portfolio 2, AR(1) for portfolio 3, ARMA(4,4) for portfolio 4 and ARMA(1,1) for portfolio 5.

The relatively smaller values of MAE and MSE for our two jackknife FHS models presented in columns 3-6 and 10-13 of Table 2 suggest that the proposed approach is able to track actual returns more closely than other models. Comparing columns 11 and 13 for 5% VaR, we can see that the rankings based on the two criteria are more or less the same. The two jackkinfe variants rank at the top for almost all portfolios. At the 1% level, their rankings slip down but are still above average in general (columns 4 and 6).

The essence of the MRSB is to consider the deviation from the mean of VaR forecasts of different models. Therefore, results depend on the choice number of models used in the comparison and would be different when adding or deleting models from comparison. A value of 0.10 implies that a given VaR forecast is 10 percent larger, on average, than the average of all VaR models under consideration. The results in columns 7-8 and 14-15 in Table 2 show that our two models are top performers for portfolios one and two, and are above average for portfolios three and four. For the high-risk portfolio, both jackknife variants are above average at the 5% level, but do not fare well at the 1% level.

The statistical and regulatory test results are presented in Table 3. The conditional coverage test (LR_{cc}) offers the most comprehensive test as it combines the unconditional coverage (LR_{uc}) and independence tests (LR_{ind})[1]. The results in columns 2-4 and 7-9 suggest that in general, the bootstrap FHS models are top performers, passing all LR tests for all portfolios. The jackknife FHS models perform satisfactorily: the two jackknife variants pass all LR tests except in the case of 1% VaR for portfolio one. For portfolios 3 and 4, Boot_Mean and Boot_ARMA are clear winners. For portfolio 5, all models pass the LR tests, suggesting that the LR tests are not useful in picking up the winning model when the returns are highly volatile.

Table 3. Statistical and regulatory tests of VaR models

Model	$\lambda = 1\%$					$\lambda = 5\%$			
	LR_{uc}	LR_{ind}	LR_{cc}	Is λ in the estimated range of $\hat{\lambda}$?	Regulatory colour zone	LR_{uc}	LR_{ind}	LR_{cc}	Is λ in the estimated range of $\hat{\lambda}$?
A. The market index (portfolio 1)									
1. Jack_Mean	78.29	1.93	6.24	Yes	Green	94.29	15.20	35.75	Yes
2. Jack_ARMA	78.29	1.93	6.24	Yes	Green	72.35	20.51	42.09	Yes
3. Boot_Mean	17.29	7.37	7.98	No	Yellow	52.89	26.83	44.45	Yes
4. Boot_ARMA	17.29	7.37	7.98	No	Yellow	52.89	26.83	44.45	Yes
5. FHS	71.90	0.60	2.16	Yes	Green	94.29	2.06	6.85	Yes
6. FHS_Dowd	78.29	1.93	6.24	Yes	Green	94.29	2.06	6.85	Yes
7. GARCH_N	39.95	4.16	8.79	Yes	Green	94.29	2.06	6.85	Yes
8. GARCH_T	71.90	0.60	2.16	Yes	Green	1.15	7.37	0.83	No
9. HS	78.29	1.93	6.24	Yes	Green	52.89	5.13	12.29	Yes
10. VCV	17.29	7.37	7.98	No	Yellow	72.35	3.33	9.76	Yes
B. The large-cap portfolio (portfolio 2)									
1. Jack_Mean	78.29	78.93	92.90	Yes	Green	72.35	20.51	42.09	Yes
2. Jack_ARMA	78.29	78.93	92.90	Yes	Green	72.35	20.51	42.09	Yes
3. Boot_Mean	78.29	78.93	92.90	Yes	Green	72.35	20.51	42.09	Yes

[1] A general guide to examine all LR test results is as follows: (1) models with p-values greater than five are desirable; (2) models having higher p-values from the LR tests are preferred; and (3) models estimates are unreliable if they pass the unconditional coverage test but fails either or both the independence and conditional coverage tests (Christofferson, 1998; 2003).

Table 3 (cont.). Statistical and regulatory tests of VaR models

	$\lambda = 1\%$					$\lambda = 5\%$			
Model	LR_{uc}	LR_{ind}	LR_{cc}	Is λ in the estimated range of $\hat{\lambda}$?	Regulatory colour zone	LR_{uc}	LR_{ind}	LR_{cc}	Is λ in the estimated range of $\hat{\lambda}$?
4. Boot_ARMA	78.29	78.93	92.90	Yes	Green	72.35	20.51	42.09	Yes
5. FHS	71.90	0.60	2.16	Yes	Green	94.29	2.06	6.85	Yes
6. FHS_Dowd	71.90	0.60	2.16	Yes	Green	94.29	2.06	6.85	Yes
7. GARCH_N	17.29	7.37	7.98	No	Yellow	72.35	3.33	9.76	Yes
8. GARCH_T	71.90	0.60	2.16	Yes	Green	7.18	16.97	7.71	No
9. HS	78.29	1.93	6.24	Yes	Green	94.29	2.06	6.85	Yes
10. VCV	78.29	1.93	6.24	Yes	Green	82.78	1.21	4.20	Yes
C. The booming sector portfolio (portfolio 3)									
1. Jack_Mean	39.95	72.10	65.79	Yes	Green	72.35	3.33	9.76	Yes
2. Jack_ARMA	39.95	72.10	65.79	Yes	Green	36.82	7.59	13.81	Yes
3. Boot_Mean	78.29	78.93	92.90	Yes	Green	82.78	10.87	26.98	Yes
4. Boot_ARMA	78.29	78.93	92.90	Yes	Green	82.78	10.87	26.98	Yes
5. FHS	78.29	1.93	6.24	Yes	Green	82.78	1.21	4.20	Yes
6. FHS_Dowd	17.29	0.16	0.27	No	Yellow	82.78	1.21	4.20	Yes
7. GARCH_N	6.46	0.41	0.30	No	Yellow	82.78	1.21	4.20	Yes
8. GARCH_T	0.61	1.72	0.14	No	Yellow	36.82	1.15	2.73	Yes
9. HS	2.11	0.90	0.23	No	Yellow	52.89	5.13	12.29	Yes
10. VCV	2.11	0.90	0.23	No	Yellow	72.35	3.33	9.76	Yes
D. The low-risk portfolio (portfolio 4)									
1. Jack_Mean	17.29	7.37	7.98	No	Yellow	72.35	3.33	9.76	Yes
2. Jack_ARMA	17.29	7.37	7.98	No	Yellow	52.89	5.13	12.29	Yes
3. Boot_Mean	6.46	11.63	5.28	No	Yellow	24.42	10.82	13.97	No
4. Boot_ARMA	6.46	11.63	5.28	No	Yellow	24.42	10.82	13.97	No
5. FHS	71.90	0.60	2.16	Yes	Green	36.82	1.15	2.73	Yes
6. FHS_Dowd	71.90	0.60	2.16	Yes	Green	36.82	1.15	2.73	Yes
7. GARCH_N	0.61	1.72	0.14	No	Yellow	24.42	1.90	3.25	No
8. GARCH_T	26.60	92.93	53.66	No	Green	7.18	0.90	0.65	No
9. HS	39.95	4.16	8.79	Yes	Green	36.82	7.59	13.81	Yes
10. VCV	39.95	4.16	8.79	Yes	Green	36.82	7.59	13.81	Yes
E. The high-risk portfolio (portfolio 5)									
1. Jack_Mean	39.95	72.10	65.79	Yes	Green	52.89	17.08	32.10	Yes
2. Jack_ARMA	71.90	85.89	92.26	Yes	Green	94.29	23.72	49.60	Yes
3. Boot_Mean	26.60	92.93	53.66	No	Green	82.78	27.62	53.98	Yes
4. Boot_ARMA	26.60	92.93	53.66	No	Green	82.78	27.62	53.98	Yes
5. FHS	26.60	92.93	53.66	No	Green	82.78	27.62	53.98	Yes
6. FHS_Dowd	26.60	92.93	53.66	No	Green	82.78	27.62	53.98	Yes
7. GARCH_N	78.29	78.93	92.90	Yes	Green	82.78	27.62	53.98	Yes
8. GARCH_T	26.60	92.93	53.66	No	Green	3.12	59.07	8.50	No
9. HS	26.60	92.93	53.66	No	Green	72.35	78.86	90.61	Yes
10. VCV	78.29	78.93	92.90	Yes	Green	94.29	68.25	91.74	Yes

Notes: The entries in columns 2-4 and 7-9 are the significance levels (p-values) of the respective likelihood ratio tests. A model is classified into the green zone if there are 4 or less violations in 250 trading days. The numbers of violations corresponding to the yellow and red zones are 5 to 9, and 10 or more respectively. The ARMA specifications in the conditional mean equation for each portfolio are: ARAM(2,2) for portfolio 1, AR(1) for portfolio 2, AR(1) for portfolio 3, ARMA(4,4) for portfolio 4 and ARMA(1,1) for portfolio 5.

As the regulatory backtesting specifies $\lambda = 0.01$, column 6 reports the colour zone that each model falls into. The VaR forecasts from the original FHS are in the green zone for all portfolios, following by our approach for 4 out of 5 portfolios. The results based on the estimated range of probability of violation are given in columns 5 and 10 of Table 3. It can be seen that our two models outperform all other eight models. In the ten combinations (i.e. 1% and 5% VaRs for five portfolios) considered, the realised value of VaR, $\hat{\lambda}$, is inside the estimated confidence range from the jackknife FHS forecasts except for the case of 1% VaR for portfolio 4. Counting the number of "no" for the ten combinations, we can rank the models according to this test: the first place is taken by the two jackknife variants and FHS, and the second place by FHS_Dowd, HS and VCV. The regulatory backtesting results, shown in column 6 of Table 3, arrive at a similar conclusion.

Finally, we summarise the results in this subsection. The proposed approach based on the delete-d jackknife clearly produces much more accurate and more precise VaR forecasts. Precision can be seen in Figure 1, where the distribution of VaR forecasts is highly concentrated. Accuracy can be seen in column 16 of Table 2, which reports the sum of the ranks of each model at both the 1% and 5% levels based on MAE, MSE and MRSB. The smaller the sum, the higher the accuracy of a model. The conclusion is that the delete-d jackknife, though an approximation to the bootstrap, possesses considerable advantages over bootstrap to forecast VaR, which is a quantile estimate. Table 3 shows that the two variants of the proposed approach outperform all other models in terms of the range and regulatory tests and provide reasonable improvement over their bootstrap counterpart in terms of the LR tests for unconditional and conditional coverage.

3.3. Further analysis. VaR is widely critised as being reflecting only normal market conditions, but not what happens during crises. A variety of stress tests are proposed in the literature to address the shortcomings of VaR. Though a standard methodology for stress testing has yet to emerge, the consensus is that when a portfolio is under stress, the assumptions underpinning VaR becomes invalid (Tan and Chan, 2003). Therefore, the selection of appropriate and economically plausible scenarios is vital to stress testing. In July 2009, the Bank for International Settlements finalized revisions to Basel II, which require the stress VaR (sVaR) added to the market risk VaR in the calculation of capital requirements (CR), i.e:

$$CR = \max\{VaR, M_c \times VaR_{ave}\} + \max\{sVaR, M_s \times sVaR_{ave}\},$$

where $M_c, M_s \geq 3$ are multiplication factors, and the subscript "ave" denotes the average over 60 days. Regarding the calculation of sVaR, there are three types of stress scenarios: (1) recent history; (2) user-defined scenarios; and (3) mechanical-search stress tests (Aragones et al., 2001).

The Basel Committee on Banking Supervision considers the 2007-8 period adequate for reflecting unusual stress (BCBS, 2009). Staley (2010), among others, suggests that the inclusion of sVaR may double capital requirements under normal market conditions. In our analysis above, our data end in 2006. It would be useful to assess the performance of our proposed methodology under extreme market conditions. We therefore carry out further analysis by examining VaR forecasts and capital requirements for the forecasting period of 2008-2009. It can be seen from Table 4 that our approach leads to underestimation of loss during extreme market conditions, based purely on Market VaR. Using the Basel II capital requirements, our approach can adequately account for losses at the 1% VaR level but not at the 5% level. The last two columns show that the violations are zero once the additional sVaR requirement is taken into consideration, suggesting that our proposed approach makes an institutional or trading portfolio stay solvent under the revised Basel II framework. Finally, it is to be noted that the conclusions are qualitatively the same across all of the five portfolios.

Table 4. Violations of VaR and capital requirements in 2008-2009: JACK_ARMA

Portfolio	VaR violation $\hat{\lambda}$		Violation of CR based on market VaR		Violation of CR based on market VaR plus sVaR	
	1%	5%	1%	5%	1%	5%
1	1.18	6.69	0	0.20	0	0
2	1.38	6.69	0	0.20	0	0
3	2.36	7.68	0	0.20	0	0
4	1.77	9.06	0	0.39	0	0
5	1.77	7.09	0	0.00	0	0

Notes: The violations are expressed in terms of percentage. As argued in the text, when sVaR is added on top of market VaR, the CR is sometimes doubled. For simplicity, we thus calculate CR under the revisions to Basil II framework as twice of market VaR. We used both five-year and one-year rolling windows but only report the results from one-year rolling window here. The two sets of results are qualitatively similar. Summary and conclusions

Since VaR models are intended to measure the relationships among many uncertainties, it is of tantamount importance that VaR forecasts are both accurate and precise. Accuracy relates to unbiasness while the precision relates to the uncertainty associated with the forecasts. Unlike most previous studies that are only concerned with the tail behavior of predicted returns, our approach incorporates a number of sources of resampling uncertainty.

As VaR forecasts are quantile estimates, bootstrapping is widely used. Our proposed new approach is based on its alternative – the jackknife. In particular, we employed the delete-d jackknife, which is specifically designed for non-smooth statistics such as the quantile and has the attraction of producing unbiased statistics because it resamples from the original distribution rather than from the empirical distribution as in the bootstrap.

We applied the proposed approach to five return series. It was found that the distributions of VaR forecasts are narrower than those from bootstrap, indicating forecast precision. To evaluate forecast performance, we compared the two variants of the proposed approach to eight other models. The results from statistical loss measures showed that our proposed approach provided more accurate VaR forecasts. In addition, our two models achieved reasonable improvement over the other eight models in terms of statistical and regulatory tests. Finally, our approach was able to track actual returns more closely, reducing the chances of both VaR underestimation and large overestimation.

Given that no VaR model has been proved adequate for all financial assets, sampling frequencies, trading positions, confidence levels and sub-periods, this paper does not intend to provide a cure all. Instead, by proposing a new approach, we tried to shed light on how to incorporate possible sources of resampling uncertainty in VaR modelling. We showed that the delete-d jackknife, an alternative to the bootstrap is very useful in improving accuracy and precision of VaR forecasts.

References

1. Angelidis, T. and Benos, A.V. (2008). Value-at-Risk for Greek Stocks, *Multinational Finance Journal*, 12, 67-105.
2. Angelidis, T., Benos, A.V. and Degiannakis, S. (2004). The Use of Garch Models in VaR Estimation, *Statistical Methodology*, 1, 105-128.
3. Angelidis, T., Benos, A.V. and Degiannakis, S. (2007). A Robust VaR Model under Different Time Periods and Weighting Schemes, *Review of Quantitative Finance and Accounting*, 28, 187-201.
4. Aragones, J.R., Blanco, C. and Dowd, K. (2001). Incorporating Stress Tests into Market Risk Modeling, *Derivatives Quarterly*, 7, 44.
5. Aussenegg, W., Miazhynskaia, T. (2006). Uncertainty in Value-at-Risk Estimates under Parametric and Non-Parametric Modeling, *Financial Markets and Portfolio Management*, 20, 243-264.
6. Bams, D., Lehnert, T. and Wolff, C. (2005). An Evaluation Framework for Alternative VaR Models, *Journal of International Money and Finance*, 24, 944-958.
7. Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K. (1998). Don't Look Back, *Risk*, 11, 100-103.
8. Barone-Adesi, G. and Giannopoulos, K. (2001). Non-Parametric VaR Techniques, Myths and Realities, *Economics Notes*, 30, 167-181.
9. Barone-Adesi, G., Giannopoulos, K. and Vosper, L. (1999). VaR without Correlations for Portfolios of Derivative Securities, *Journal of Futures Markets*, 19, 583-602.
10. Baysal, R. and Staum, J. (2008). Empirical Likelihood for Value-at-Risk and Expected Shortfall, *The journal of risk*, 11, 3-32.
11. BCBS (2009). Revisions to the Basel II Market Risk Framework. Basel Committee on Banking Supervision.
12. Beder, T.S. (1995). Var: Seductive but Dangerous, *Financial Analysts Journal*, September.
13. Boudoukh, J., Richardson, M. and Whitelaw, R. (1998). The Best of Both Worlds, *Risk*, 11, 64-67.
14. Brooks, C. and Persand, G. (2003). Volatility Forecasting for Risk Management, *Journal of Forecasting*, 22, 1-22.
15. Campbell, R., Huisman, R. and Koedijk, K. (2001). Optimal Portfolio Selection in a Value at Risk Framework, *Jornal of Banking and Finance*, 25, 1789-1804.
16. Chappell, D. and Dowd, K. (1999). Confidence Intervals for VaR, *Financial Engineering News*, 9, 1-2.
17. Chernozhukov, V.L. and Umantsev, L. (2001). Conditional Value-at-Risk: Aspects of Modeling and Estimation, *Empirical Economics*, 26, 272-292.
18. Christofferson, P.F. (1998). Evaluating Interval Forecasts, *International Economic Review*, 39, 841-861.
19. Christofferson, P.F. (2003). Elements of Financial Risk Management, Academic Press, London.
20. Davison, A.C. and Hinkly, D. (1997). Bootstrap Applications and Their Applications, Cambridge University Press, Cambridge.
21. Dowd, K. (2005). Measuring Market Risk, John Wiley and Sons, Chichester.
22. Dowd, K. and Blake, D. (2001). Estimating VaR with Order Statistics, *Journal of Derivatives*, 8, 23-30.
23. Duffie, D. and Pan, J. (1997). An Overview of Value at Risk, *Journal of Derivatives*, 4, 7-49.
24. Efron, B. (1979). Bootstrap Method: Another Look at the Jackknife, *Annals of Statistics*, 7, 1-26.
25. Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap, Chapman and Hall, New York.
26. Engel, R.F. and Manganelli, S. (2004). Caviar: Conditional Value at Risk by Quantile Regression, *Journal of Business & Economic Statistics*, 22, 367-381.
27. Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications, Chapman and Hall, London.
28. Fan, J. and Gu, J. (2003). Semiparametric Estimation of Value at Risk, *The Econometrics Journal*, 6, 261-290.
29. Giannopoulos, K. and Tunaru, R. (2005). Coherent Risk Measurement under Filtered Historical Simulation, *Journal of Banking and Finance*, 29, 979-996.
30. Hartz, C., Mittnik, S. and Paolella, M.S. (2006). Accurate Value-at-Risk Forecasting Based on the Normal-Garch Model, *Computational Statistics and Data Analysis*, 51, 2295-2312.

31. Hendricks, D. (1996). Evaluation of Value-at-Risk Models Using Historical Data, *Federal Reserve Bank of New York Economic Policy Review*, April, 39-69.
32. Hyndman, R.J. and Fan, Y. (1996). Sample Quantities in Statistical Packeges. The American Statistion, 50, pp. 361-365.
33. Inui, K., Kijima, M. and Kitano, A. (2005). VaR Is Subject to a Significant Positive Bias, *Statistics and Probability Letters*, 72, 299-311.
34. Jayasuriya, S.A. and Rossiter, R. (2008). Asymmetric Volatility, Risk and Filtered Historical Simulation, *Journal of Academy of Business and Economics*, 8, 56-67.
35. Jorion, P. (1996a). Risk2: Measuring the Risk in Value at Risk, *Financial Analysts Journal*, 52, 47-56.
36. Jorion, P. (1996b). Value-at-Risk: the New Benchmark for Managing Financial Risk, 1st Edition, McGraw-Hill.
37. Jorion, P. (2000). Value-at-Risk: the New Benchmark for Managing Financial Risk, 2nd Edition, McGraw-Hill.
38. Jorion, P. (2010). Value-at-Risk: the New Benchmark for Managing Financial Risk, 3rd Edition, McGraw-Hill.
39. Kerkhof, J., Melenberg, B. and Schumacher, H. (2010). Model Risk and Capital Reserves, *Journal of Banking and Finance*, 34, 267-279.
40. Kim, H.T. and Hardy, M.R. (2007). Quantifying and Correcting the Bias in Estimated Risk Measures, *ASTIN Bulletin*, 37, 365-386.
41. Kuester, K., Mittnik, S. and Paolella, M. (2006). Value-at-Risk Prediction: a Comparison of Alternative Strategies, *Journal of Financial Econometrics*, 4, 53-89.
42. Lee, M.C., Su, J.B. and Liu, H.C. (2008). Value-at-Risk in US Stock Indices with Skewed Generalized Error Distribution, *Applied Financial Economics Letters*, 4, 425-431.
43. Mooney, C.Z. and Duval, R.D. (1993). Bootstrapping: a Nonparametric Approach to Statistical Inference, Sage Publications, Newbury Park, CA.
44. Pesaran, M.H., Schleicher, C. and Zaffaroni, P. (2009). Model Averaging in Risk Management with an Application to Futures Markets, *Journal of Empirical Finance*, 16, 280-305.
45. Pristker, M. (2006). The Hidden Dangers of Historical Simulation, *Journal of Banking and Finance*, 30, 561-582.
46. So, M.K.P. and Yu, P.L.H. (2006). Empirical Analysis of Garch Models in Value-at-Risk Estimation, *International Financial Markets*, *Institutions and Money*, 16, 180-197.
47. Staley, M. (2010). Building on Our Momentum: The Incremental Risk Charge in Basel II. TD Bank.
48. Szegö, G. (2002). Measures of Risk, *Journal of Banking and Finance*, 26, 1253-1272.
49. Tan, K.H. and Chan, I.L. (2003). Stress Testing Using VaR Approach – a Case for Asian Currencies, *Journal of international financial markets, institutions and money*, 13, 39-55.
50. Wu, C.F.J. (1986). Jackknife, Bootstrap and Other Resampling Plans in Regression Analysis (with Discussion), *Annals of Statistics*, 14, 1261-1350.