Direct determination of Operational Value-at-Risk using Descriptive Statistics

Peter Mitic[1][0000−0002−9845−4435]

[1] Dept. Computer Science, UCL, Gower Street, London WC1E 6BT
[2] Santander UK, 2 Triton Square, Regent's Place, London NW1 3AN
p.mitic@ucl.ac.uk
https://www.ucl.ac.uk/computer-science/ucl-computer-science

Abstract. Regression and machine learning methods are applied to the problem of *Value-at-Risk* determination in the context of financial Operational Risk, in order to determine an optimal technique that agrees sufficiently well with established Monte Carlo analyses. The annualised sum of operational losses is identified as the most significant statistical influence on *Value-at-Risk*, and a technique using it as a proxy for measured *Value-at-Risk* in a Test environment is formalised. The optimal stand-alone model is Generalized Additive, with approximately 61% success. The success rate can be enhanced to approximately 65% using a stacked model.

Keywords: Operational Risk · Pickands-Balkema-deHaan · Descriptive Statistics · Value-at-Risk · Generalized Additive Model · Loss Distribution

1 Introduction and Motivation

Intuitively, descriptive statistics of the financial data should provide a broad indication of the financial risk associated with the data. Surprisingly, little attention has been paid to using data properties directly to measure financial risk. In this paper we attempt to relate the descriptive statistics of *Operational* [3] losses ("OpRisk") to *Value-at-Risk* (VaR): a financial risk metric originating from the *J.P. Morgan/Reuters* "RiskMetrics" measure from the 1990s. [11].

The most generally applicable way to determine VaR in the context of OpRisk is the Monte-Carlo-based *Loss Distribution Approach* (LDA [7]). Using the LDA has been a persistent problem because it is nearly always possible to find, for any single data set, multiple disparate 'solutions'. Then, which to choose is unclear. In this paper we explore a range of regression and machine learning (ML) techniques in order to find an optimal method to estimate VaR using descriptive statistics of the data. Any selected technique should be quick to use, interpretable, and agree tolerably well with the LDA-calculated value.

[3] The risk of financial loss due to flawed or failed processes, policies, systems or events that disrupt business operations.

2 Literature Review

We concentrate on the few attempts to relate OpRisk VaR to properties of the underlying data. Curti and Migueis [5] and [6] test whether metrics derived from past VaR calculations can be used to predict future VaR calculations, using quantile regression. They use explanatory features such as loss frequency and mean total loss, with balance sheet co-variates such as Market Capitalisation. Some characteristics of our analysis are proposed by Chavez-Demoulin et al [3]. They implement a bespoke *GAM*, using economic co-variates as distribution hyper-parameters. The statistical properties used are *Mean, Standard Deviation, Median, Third Quartile, Maximum, Skewness, Number of Losses Above 1€m.* Consequently, *Tail* properties [4] were missing. No methodological comparisons are given. Very recently, Mitic [10] established an upper bound for VaR as $7\frac{1}{3} \times S/Y$, where S is the loss sum and Y is the number of years spanned by those losses.

ML methods have been applied in the context of OpRisk, but not directly to predict VaR. Chen and Wen [4] used simulated data to predict OpRisk losses with operational control explanatory variables. Pakhchanyan et al [12] used SVM and naive Bayes algorithms to assign data to specific risk classes. Aziz and Dowling [1], and Carrivick and Westphal [2], note the use of ML methods in OpRisk control (e.g. fraud prevention). Pena et al [13] is a rare example of a ML method for calculating OpRisk capital. Diverse data sources are integrated using a convolutional neural network empowered by fuzzy cognitive maps. In common with the LDA, the inputs to their model are OpRisk frequency and severities.

3 Optimal methodology determination for VaR

Our approach is to apply *Linear Regression* (LR) and ML methods to a common data set that would be appropriate for a mid-to-large western European bank, optimised using a *Proxy* (described in Section 3.4).

3.1 Data and pre-processing

Approximately 1100 random samples of sizes between 200 and 1000 were generated using appropriate 'fat-tailed' distributions (LogNormal, Weibull, Generalised Pareto etc.). The time span was a nominal 5 years. A single 66.7% Training set was determined by sampling and optimising *Mean Absolute Error* (MAE) using (LR). LR is thereby set as a stringent base from which to judge other results.

The Training and Test set features were standardised, separately, to mean 0 and standard deviation 1. We have found that results were impaired by either normalising data to [0,1], or by removing significantly correlated features. For each data set, the following principal statistics were calculated.

[4] The largest $p\%$ of losses, with, typically, $p \in (1, 10)$

- *Mean, SD, Skewness, Kurtosis, Maximum, Sum* (applied to all data)
- *MeanTail SDTail, SkewnessTail, KurtosisTail* (applied to the data Tail)
- Quantiles *Q10-Q90* in steps of 10, Quantiles *Q91-Q99* in steps of 1

The *Capital* (VaR at 99.9%) for each data set was calculated by fitting all distributions, and selecting the optimal distribution with respect to the *TNA* 'best fit' statistic [8], which is robust with respect to data set size.

3.2 Candidate VaR assessment models

A range of regression, ML, and other models were applied to the common data set (Section 3.1). The names below are used in the results tables in Section 4.

- LR, acting as a base reference method
- Other regressions: *BAYES*, *RIDGE*, *LASSO*, *LOESS*
- ML: Neural Network (*NN*), Random Forest (*RF*), Support Vector Machine (*SVM*), Bagging (*BAG*)
- Boost methods: *XG*, Adaptive Gradient (*ADA*), Gradient (*GRAD*)
- Others: *GLM*, *GAM*, k-Nearest Neighbours (*KNN*)
- *STACK*: the optimised (Section 3.4) mean of a subset of preceding models

3.3 Success Metrics

Established metrics (R^2, *MAE*) do not provide a sufficiently precise comparison of predicted and actual results. We use them mainly in parameter optimisations. Instead, a multi-part metric is used to measure prediction "success". The percentage of instances for which the predicted value differs from the LDA-calculated value by 10%, 25%, 50%, 75% and 90% provides a *Progressive Success* (abbreviated to *PS*) metric. In particular, the 10%, 25% figures correspond approximately to tolerable error bounds in the context of OpRisk VaR calculations.

3.4 Optimisation using a Proxy

We have found that routine application of the methods of Section 3.2 does not provide sufficient success when the multi-part metric in Section 3.3 is applied. The primary problem is for small ($<$ €20m) VaR. The prediction error can the very large in comparison with the 'actual' value, and the 10% component of the multi-part success metric is particularly sensitive in that range. The *Proxy*, in which a known feature replaces an unknown feature, improves results considerably. It works by scaling predicted results in the Test environment using scale factors calculated in the Training environment, and using a *known* feature of the Test data in place of "actual" (but unknown) Test VaR values.

Proxy Correction
Recent work [9] has shown that feature *Annualised Sum* (i.e. the sum of all data divided by the number of years spanned by the data) is highly correlated with OpRisk VaR, so we define and use a *Proxy* for "actual" Test VaR values using *Annualised Sum*. We call it *AnnSum* for short, and stress that *AnnSum* is always known for Test data, whereas "actual" Test VaR is not. The *Proxy* is used to modify Test predictions in a way that is beneficial in Training. The graphic in Figure 1 shows the principal stages in *Proxy* optimisation. Processes are applied in the order indicated by the *STEP* numbers. Indexing is done for over- and under-predictions (relative to the *Proxy*). In Figure 1, the resulting indices are denoted by I_U and I_L, and scale factors derived using them are denoted by K_U and K_L.

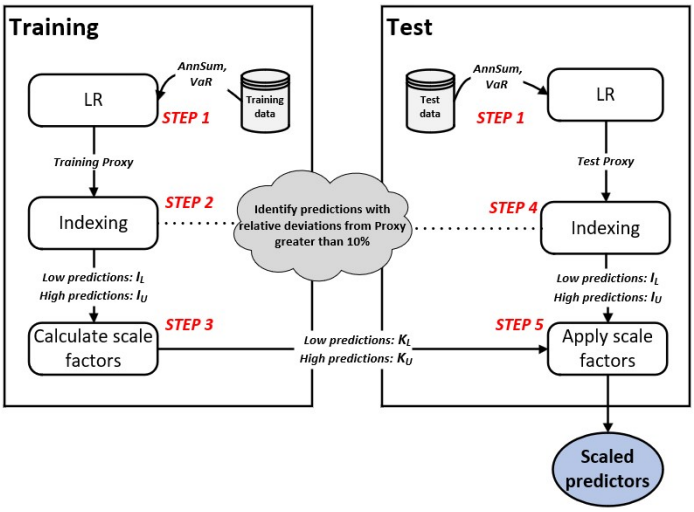

Fig. 1. The principal stages in *Proxy* optimisation. *STEP* labels refer to the detailed steps in Section 3.4

.

Proxy Correction: Details
The formal explanation below follows the same steps as in Figure 1, and uses notation that makes a distinction between the Training and Test environments: superscript **Tr** for the Training environment, superscript **Te** for the Test environment, **P** for *Proxy*, **Y** for actual VaR, **Z** for predicted VaR, and subscript **opt** for optimised variates.

STEP 1. Determine a linear fit of the *Proxy* feature to the LDA-calculated VaR in the Training and Test environments separately. The intercepts +1 represent *de minimis* VaR for zero loss.

$$\mathbf{P^{(Tr)}} = a^{(Tr)}\mathbf{Y^{(Tr)}} + 1; \qquad \mathbf{P^{(Te)}} = a^{(Te)}\mathbf{Y^{(Te)}} + 1; \tag{1}$$

STEP 2. In the main optimisation stage, the instances that have a 10% difference for the Training Proxy relative to the predicted VaR are identified, and indexed by indices $I_l^{(Tr)}$ and $I_u^{(Tr)}$.

$$I_l^{(Tr)} = \frac{-\mathbf{Z^{(Tr)}} + \mathbf{P^{(Tr)}}}{\mathbf{P^{(Tr)}}}; \qquad I_u^{(Tr)} = \frac{\mathbf{Z^{(Tr)}} - \mathbf{P^{(Tr)}}}{\mathbf{P^{(Tr)}}} \tag{2}$$

STEP 3. Scaling constants K_l and K_u are found by optimising MAE for the indexed predicted ("Z") and Proxy ("P").

$$\mathbf{Z_{opt}^{(Tr)}} = \min_{(0,1)} \mathbb{E}\Big[\big|K_l\mathbf{Z^{(Tr)}} - \mathbf{P^{(Tr)}}\big|\mathbb{I}(I_l^{(Tr)}) + \big|K_u\mathbf{Z^{(Tr)}} - \mathbf{P^{(Tr)}}\big|\mathbb{I}(I_u^{(Tr)})\Big] \tag{3}$$

STEP 4. Indices marking 10% relative deviations of Test predictions from the Test *Proxy* values are defined in a similar way to *STEP 2.*

$$I_l^{(Te)} = \frac{-\mathbf{Z^{(Te)}} + \mathbf{P^{(Te)}}}{\mathbf{P^{(Te)}}}; \quad I_u^{(Te)} = \frac{\mathbf{Z^{(Te)}} - \mathbf{P^{(Te)}}}{\mathbf{P^{(Te)}}} \tag{4}$$

STEP 5. The constants K_l and K_u feed through to the Test environment. The final Test predictor is the model-derived predictor (index m in Equation 4), with the Proxy optimisations applied on the sets indexed u or l. No transformation is applied to Test predictions that correspond to the complement of the indices from *STEP 4* (these are the "acceptable" predictions).

$$\begin{aligned}
\text{Under predictions:} \quad & \mathbf{Z_{opt}^{(Te)}}\mathbb{I}(I_l^{(Te)}) = 1 + K_l\mathbf{Z^{(Te)}}\mathbb{I}(I_l^{(Te)}) \\
\text{Over predictions:} \quad & \mathbf{Z_{opt}^{(Te)}}\mathbb{I}(I_u^{(Te)}) = 1 + K_u\mathbf{Z^{(Te)}}\mathbb{I}(I_u^{(Te)}) \\
\text{Acceptable predictions index:} \quad & I_m^{(Te)} = \mathbb{I}(\sim I_u^{(Te)} \;\&\; \sim I_l^{(Te)}) \\
\text{Acceptable predictions unscaled:} \quad & \mathbf{Z_{opt}^{(Te)}}\mathbb{I}(I_m^{(Te)}) = \mathbf{Z^{(Te)}}\mathbb{I}(I_m^{(Te)})
\end{aligned} \tag{5}$$

Section 4.2 shows empirical results for the Proxy/Actual relationship.

4 Results

Overall results are presented first, and are followed by results specific to application of the Proxy correction. We concentrate on results using all available explanatory features in standardised form, since that configuration usually admits superior performance over others.

4.1 Model Prediction Results

Table 1 shows ordinates of the *Progressive Success* metric per model, in the Test environment, in decreasing order of the 10% and 25% components of the metric. Their ranges (22-35% and 56-63% respectively) indicate reasonable success, given the stringency of the metric and Training set selection method. The LR model using normalised uncorrelated features out-performed the LR model using standardised correlated features. Both LR results are shown.

Table 1. Test environment *Progressive Success* metric ordinates per model, with *Proxy*: retaining correlated standardised features in all cases except for LR Normalised Uncorrelated.

Metric	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**MAE**	R^2
STACK	0.35	0.61	0.84	0.96	0.97	36.57	0.84
LR Norm, Uncorr	0.31	0.63	0.90	0.97	0.97	29.85	0.90
GAM	0.31	0.63	0.88	0.97	0.97	33.95	0.84
BAYES	0.29	0.63	0.87	0.96	0.97	34.63	0.84
XG	0.29	0.59	0.85	0.95	0.96	36.51	0.81
LR Std, Corr	0.28	0.62	0.87	0.95	0.97	35.93	0.83
NN	0.28	0.61	0.86	0.96	0.96	35.26	0.83
RIDGE	0.28	0.62	0.88	0.96	0.97	35.14	0.84
LOESS	0.28	0.61	0.89	0.98	0.99	34.16	0.86
RF	0.26	0.62	0.86	0.95	0.97	34.85	0.85
KNN	0.26	0.60	0.84	0.96	0.96	41.67	0.79
SVM	0.26	0.61	0.88	0.98	0.98	35.06	0.85
GRAD	0.26	0.54	0.88	0.96	0.96	33.28	0.87
LASSO	0.25	0.62	0.86	0.96	0.97	38.30	0.80
GLM	0.24	0.59	0.86	0.99	0.99	40.73	0.78
ADA	0.23	0.60	0.88	0.97	0.97	34.72	0.85
BAG	0.22	0.56	0.88	0.97	0.99	34.65	0.85

A selection of scatter plots is shown in Figure 2. Each node is an LDA-calculated/ prediction pair. The four plots illustrate optimal cases. There are two contrasting 'simplest' LR models, one with standardised features and the other without. Normalising features usually results in more accurate predictions for LDA-calculated VaR > €500m, at the expense of impaired overall success, and this is apparent in the upper right scatter. The *GAM* produced the highest individual 10% and 25% components of the *Progressive Success* metric using all features, standardised. The *Stack Model* produced the highest overall 10% and 25% components, again using all features, standardised.

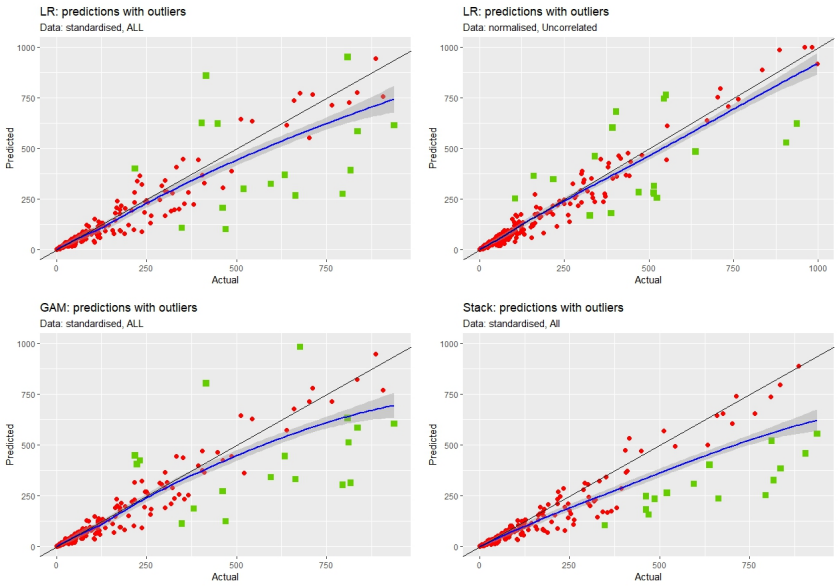

Fig. 2. Result scatters: Test environment. Upper Left: LR All features/Standardised. Upper Right: LR Uncorrelated features/Normalised. Lower Left: *GAM* all features/Standardised. Lower Right: STACK all features/Standardised. Filled circles are nodes for which VaR predictions agree with actual VaR with 95% 2-tail confidence. Filled squares are outliers.

4.2 The effect of the Proxy

Feature importance may be assessed by the *Feature Difference* method, noting the effect on MAE of successively excluding each feature using LR. When applied to the Training set, *Sum* is the most significant feature (54.1% of the variance). Table 2 shows the effect in percentage terms of the Proxy in boosting components of the *PS* metric. The large percentage improvements are notable in nearly all cases. They come at the expense of reduced MAE and R^2 scores in most cases.

Table 2. Percentage boosting of components of the *PS* metric by applying the Proxy, expressed as the *Annual Sum* variate. Negatives indicate performance impairment.

Metric	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**MAE**	R^2
GAM	38.57	56.8	33.66	24.79	18.75	-12.66	0.93
BAYES	100	79.82	49.73	30.13	25.83	-18.01	1.4
XG	36.92	39.69	15.72	10.86	9.49	2.67	-5.18
LR	81.25	81.13	41.67	27.71	25.21	-17.27	1.4
NN	44.25	30.16	10.67	8.71	5.68	-2.95	-3.24
RIDGE	55.23	72.44	41.3	27.07	21.48	-13.79	0.42
LOESS	34.38	33.1	27.52	18.99	14.5	-16.11	2.26
RF	47.5	43.25	32.59	23.52	17.69	-16.02	1.66
KNN	95.24	74.07	32.32	27	22.86	-19.23	-3.16
SVM	38.98	29.93	15.48	11.31	9.29	-7.69	0.02
GRAD	9.59	21.43	10	7.19	3.83	-5.99	-0.31
LASSO	182.14	128.24	68.55	44.02	36.65	-15.25	-1.71
GLM	0	0.55	0.75	0.65	0.65	-6.5	4.92
ADA	40.39	61.34	29.38	21.97	18.06	-8.19	-0.48
BAG	20.9	18.18	9.62	6.92	4.22	-6.46	-0.24

Figure 3 shows how much difference the *Proxy* can make. Nodes with low value (< €100m) VaR are shown with and without the *Proxy*. In the former case, nodes are concentrated on the diagonal ("predicted = actual") line, but some under-prediction is evident. In the latter case, there are multiple disparate predictions for individual nodes with low-valued actual VaR.

5 Discussion

The motivation for this study was to determine if OpRisk VaR can be estimated directly from the statistical properties of the underlying data. The principal model for doing that was to be multivariate linear regression. The results using that method, measured using the *PS* metric, proved to be lower than expected. That prompted alternative models to be used. It also raises several issues.

1. The values obtained using the LDA may not be optimal as far as VaR assessment is concerned, even though their fitted distributions are optimal.

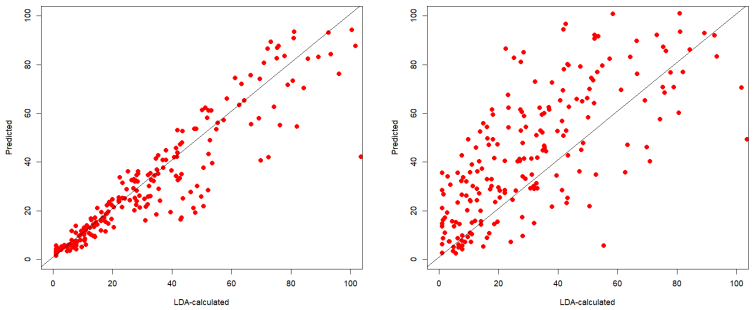

Fig. 3. Linear Model actual/predicted scatter. Left: with *Proxy*. Right: without.

Considerations such as shape of the model QQ curve and consistency with results from previous years are also important.
2. The variability of LDA-derived results suggests that a simple model (such as LR) could be used for validation, or even as an LDA replacement. In the former case, LDA-derived distributions could be rejected based on an alternative assessment.
3. Possibly the descriptive statistics of the data are not adequate for VaR determination. Suitable alternative are not apparent.
4. The Proxy correction can only be applied if a suitable proxy can be found.

6 Conclusion

Any of the methods discussed in this paper should be used to validate, or even replace, an LDA-derived VaR, provided that we can be confident that whichever is selected is a correct reflection of the underlying data. Given the multiple alternatives, which, if any, should be selected? There is compelling pressure to select a simple method that is explainable to practitioners in terms of data features. For this reason, LR is a clear choice.

References

1. Aziz,S. and M. Dowling,M.: Machine Learning and AI for Risk Management. In: Editor, Lynn,T., Mooney,G. et al (eds.)) Disrupting Finance: FinTech and Strategy in the 21st Century, pp. 33–50. (2019) doi:10.1007/978-3-030-02330-0_3
2. Carrivick,L. and Westphal,A.: Machine learning in operational risk. ORX White Paper. (2019) https://managingrisktogether.orx.org/free-resources
3. Chavez-Demoulin, V., Embrechts, P., and Hofert, M. An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates. Jnl. Risk and Insurance, **83**(3), pp. 735–776 (2016) http://www.jstor.org/stable/43998282
4. Chen,Q. and Wen,Y.: A BP-Neural Network Predictor Model for Operational Risk Losses of Commercial Bank. Third International Symposium on Information Processing, Qingdao, China, pp. 291–295 (2010) doi:10.1109/ISIP.2010.43.

5. Curti, F. and Migueis,M.: Predicting Operational Loss Exposure Using Past Losses. Finance and Economics Discussion Series 2016-002. Washington: Board of Governors of the Federal Reserve System. (2016) doi:10.17016/FEDS.2016.002
6. Curti, F. and Migueis,M.: The Information Value of Past Losses in Operational Risk. Finance and Economics Discussion Series 2023-003. Washington: Board of Governors of the Federal Reserve System. (2023) doi:10.17016/FEDS.2023.003
7. Frachot, A., Georges, P., and Roncalli, T.: Loss distribution approach for operational risk. Working paper, Groupe de Recherche Operationnelle, Credit Lyonnais, France. (2001) http://ssrn.com/abstract=1032523
8. Mitic,P.: Improved Goodness-of-Fit tests for Operational Risk. Jnl. Operational Risk, **15**(1), pp. 77–126 (2015) doi:10.21314/JOP.2015.159
9. Mitic,P.: Reasonableness and Correctness for Operational Value-at-R. Economic Analysis Letters **2**(3) pp. 35–44 (2023) doi:/10.58567/eal02030005
10. Mitic,P.: Credible Value-at-Risk. Jnl. Operational Risk (2023) To appear
11. J.P.Morgan/Reuters: RiskMetrics—Technical Document 4th edition (1996) https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a
12. Pakhchanyan,S., Fieberg,C., Metko,D. and Kaspereit,T.: Machine learning for categorization of operational risk events using textual description. Jnl. Operational Risk **17**(4), pp. 37–65 (2022) doi:10.21314/JOP.2022.026
13. Pena,A., Patino, A. et al: Fuzzy convolutional deep-learning model to estimate the operational risk capital using multi-source risk events. Applied Soft Computing **107**(107381) (2021) https://www.sciencedirect.com/science/article/pii/S1568494621003045