Technical Assessment
Analyst, Data Science

Jack Walker

1. Introduction

This report is a prioritised list of firms to support the allocation of supervisory resources. The analysis has been conducted in Python, in an accompanying Jupyter Notebook.

2. Prioritisation

The recommended firms to be prioritised are captured in Table 1, in priority order. The values displayed in the table for each firm are for 2020 but the reasons for their inclusion span the period of the dataset. The list could be extended to a large number of additional firms, as evident from the charts included in this report.

Table 1. Prioritised firms

firm	year	nwp	scr-cov-ratio	gwp	gci	ncr	nwp/gwp
Firm 210	2020	60699.95	0.16	69697.93	1024.04	0.83	0.87
Firm 7	2020	6486.66	0.10	16183.57	0	0	0.40
Firm 151	2020	5554.25	0.58	15824.69	-14.26	0.05	0.35
Firm 105	2020	2612.55	1.17	5857.04	530.53	1.84	0.45
Firm 158	2020	432.78	2.97	1212.33	2214.25	0.13	0.36
Firm 234	2020	1569.10	0.05	4051.51	1231.55	0.01	0.39
Firm 127	2020	NaN	89529.28	NaN	304.50	0	NaN

Firms to be prioritised are dependent on a number of factors, including the amount of resource available and the timeframes for this analysis. Where charts have been plotted, these have been included in this report to support any further prioritisation, or expansion of this work.

2.1. Gross written premium

Gross written premium (GWP) - total revenue written by an insurer (equivalent of turnover for a non-insurance firm) was used as the primary measure for firm size. Figure 1 shows the 30 firms by their cumulative GWP over a five, three and one year period.

Firms 210, 4 and 311 have consistently been the largest three firms in terms of GWP over the 5-year period of the data. Firm 210 is significantly larger than the other firms, 276% larger than the second largest firm. The benefit of this visualisation is that is also identifies firms 34, 7 and 151 as large firms growing at a faster rate relative to other firms around their size. They report a greater relative GWP in 2020 than they did in the cumulative 3 and 5 year preceding period.

2.2. Risk passed on to reinsurers

The risk a firm passes on to reinsurers is calculated as NWP/GWP. For this prioritisation, I was interested in firms with large or volatile values. The firms plotted in Figure 2 are the 10 largest firms identified in Figure 1 (one-year cumulative GWP).

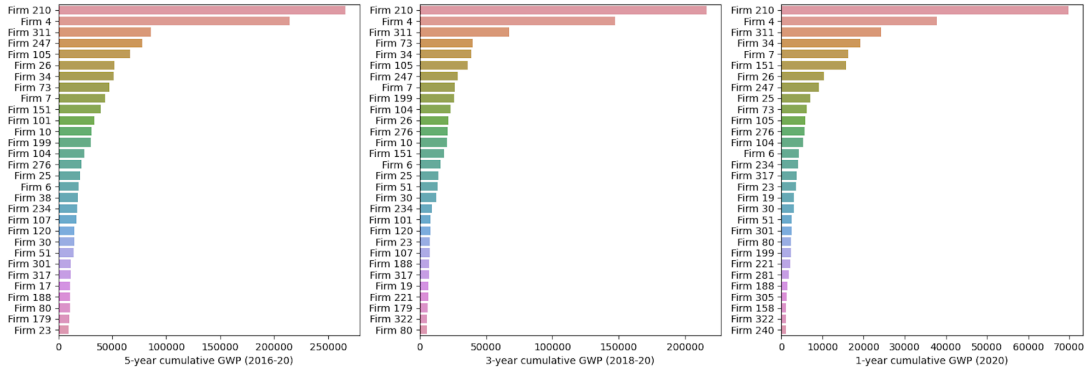

Figure 1. Firms with the greatest GWP

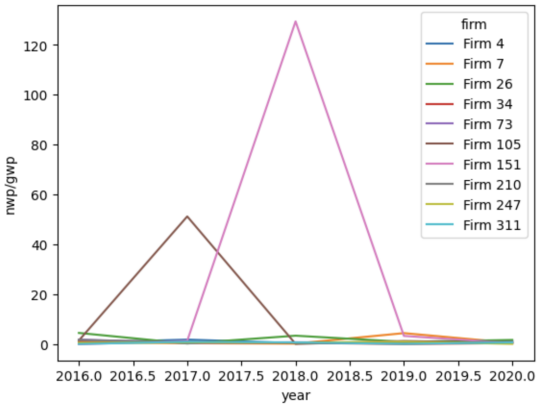

Figure 2. NWP / GWP for Firms with the Greatest GWP

Of these, Firm 151 can be seen to have reported a value of 129 for 2018, suggesting they were exposed to significant risk that wasn't passed on to reinsurers. Similarly, Firm 105 reported a value of 51 in 2017. These are the 6th and 11th firms by GWP respectively and therefore should be prioritised.

2.3. SCR coverage ratio

SCR coverage ratio is a measure of whether a firm is meeting its prudential capital requirements. Where firms report values below 1 this indicates that they are not meeting their capital requirements. Of the firms included in this data, 232 firms reported a coverage ratio of below 1. Figure 3 displays the twenty firms with the lowest ratios.

Of these, Firm 210 is identified as of particular concern given it is the largest firm and therefore it is top of the priority list. Other relatively large firms included in Figure 3 include Firms 7 and 234.

Outliers were identified in this data and removed when producing Figure 3. It was found that Firm 127 has recorded incorrect data for all years between 2017 and 2020 hence this firm has been included in the priority list due to consistent misreporting. A list of other outliers are included in the accompanying Jupyter Notebook which includes eight firms.

2.4. Gross claims incurred

Gross claims incurred represent a large cost to insurers and therefore how these change over time must be monitored. The 15 firms with the largest total gross claims incurred over the period

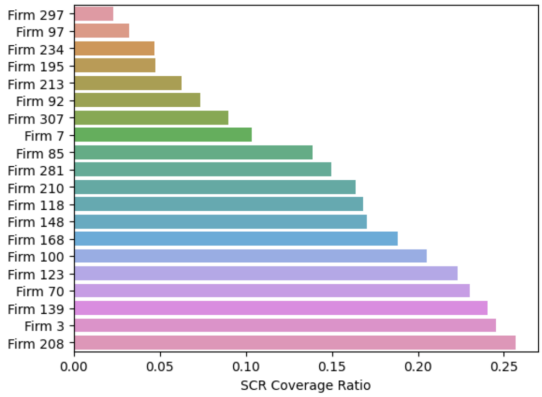

Figure 3. Firms with the lowest SCR coverage ratio

of data available were used for an in-depth analysis. For these firms, the year-on-year % change in gross claims incurred was calculated and plotted in Figure 4.

Firm 158 reported a 42.6% increase in 2020 compared to 2019 past year which represents an unusually large increase and hence the firm has been added to the priority list. It should also be noted that Firm 37 reported a 8460% increase in 2018 when compared to 2017 but this was deemed to be an outlier and removed from the plot.

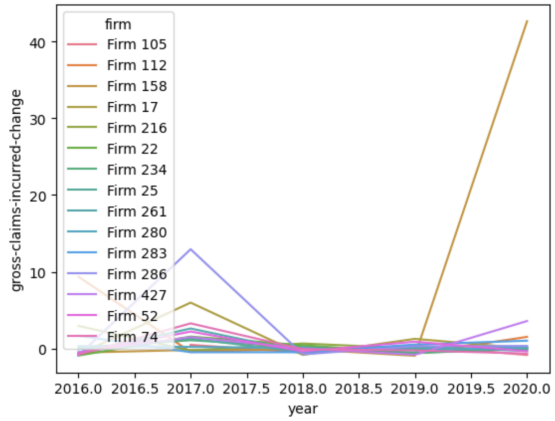

Figure 4. Year-on-year % change of gross incurred loss

2.5. Net combined ratio

Net combined ratio indicated the profitability of a firm. A value of <1 means that a firm is profitable. The ten firms with the highest profitability and the ten lowest are visualised in Figure 5. The time series indicates that Firms 331, 70, 28, 389, 228 and 166 have reported unusually large values for one or more years, creating volatility. These firms have been identified as firms to prioritise although given they are not as large as other firms on the list, are further down the prioritised list.

3. A note to outliers

Outliers are prevalent in the dataset and methods for tackling these should be utilised in future analysis to increase the efficiency and accuracy of reporting. Figure 6 gives an indication of the number of outliers included in 2020 data alone.

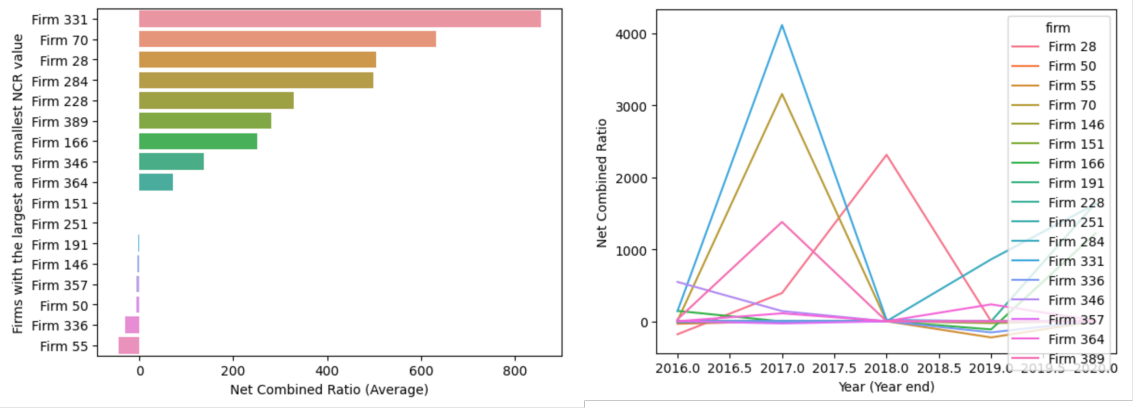

Figure 5. Net combined ratio

Appendix A - Machine Learning Applications for Future Consideration

Different ML techniques can be applied to tackle the outliers in the dataset. For firm data it is important these outliers are detected so Supervisors can decide whether the data points are of concern or likely to be erroneous.

Firstly, I applied the inter-quartile range (IQR) method to detect outliers. By using the upper and lower quartile to set the upper and lower bound for acceptable values, values that fall outside of the range can be automatically detected. Figure 7 demonstrates the distribution of NWP/GWP values before (left) and after (right) this method is applied. Whilst this process is manual, it allows for user control in setting the upper and lower bounds to what they deem to be reasonable values.

To take a more ML-driven approach, I also attempted an initial application of a Density-Based Spatial Clustering of Applications with Noise (DBSCAN). In Python this is relatively intutitve with the support of the sklearn library. This method finds core samples of high density and expands clusters from them. The result of this method is captured in Figure 8 showing the distribution of gross incurred losses before (left) and after (right) its application.

Had I had more time for this analysis, and a better familiarity with the dataset, I would have enjoyed exploring the possibility of a more developed model.

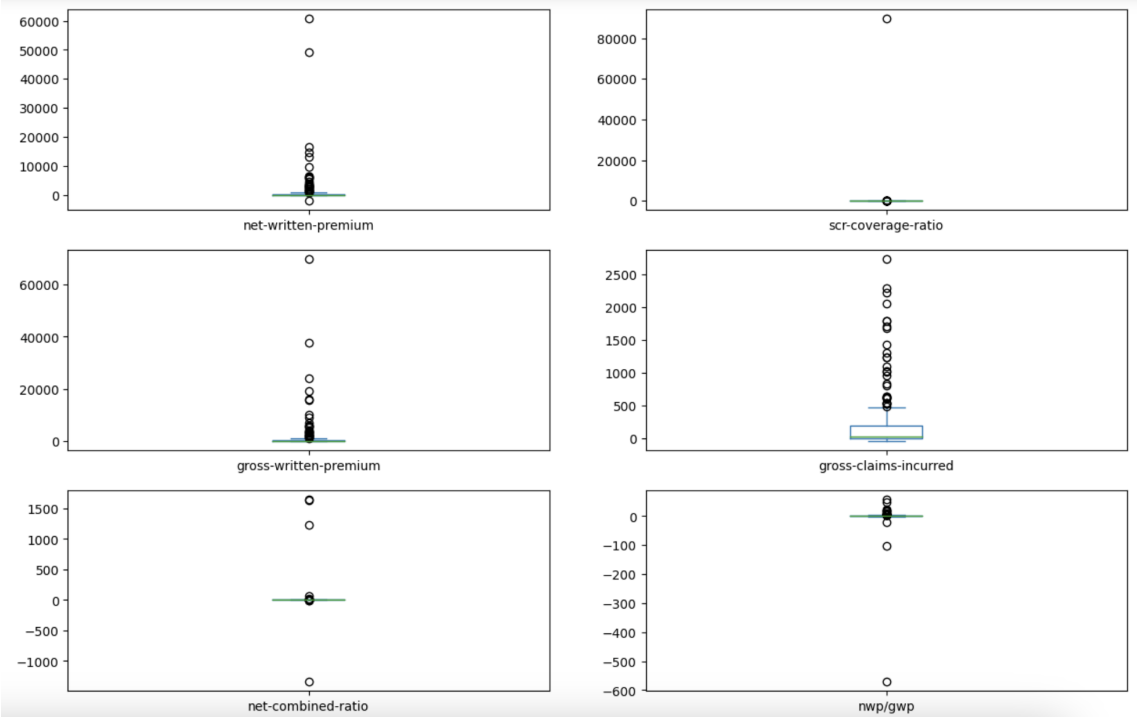

Figure 6. Prevalence of outliers in 2020 data

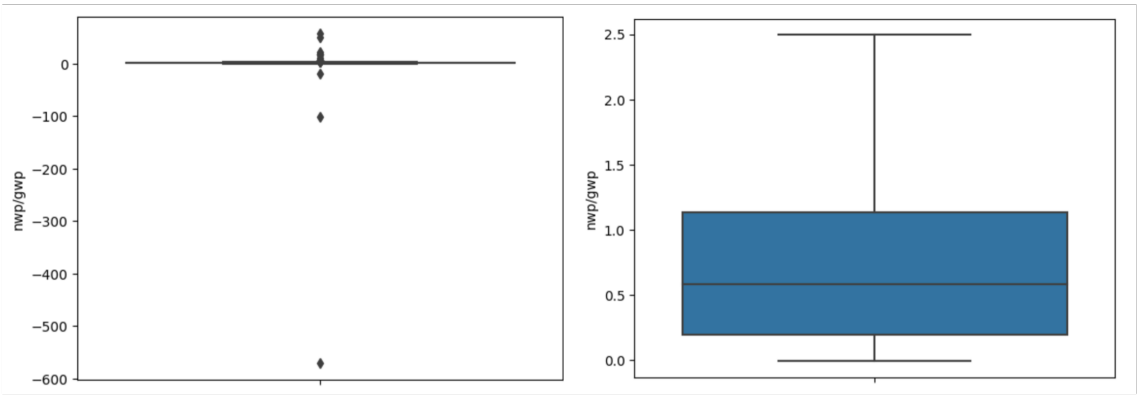

Figure 7. Outlier detection - IQR

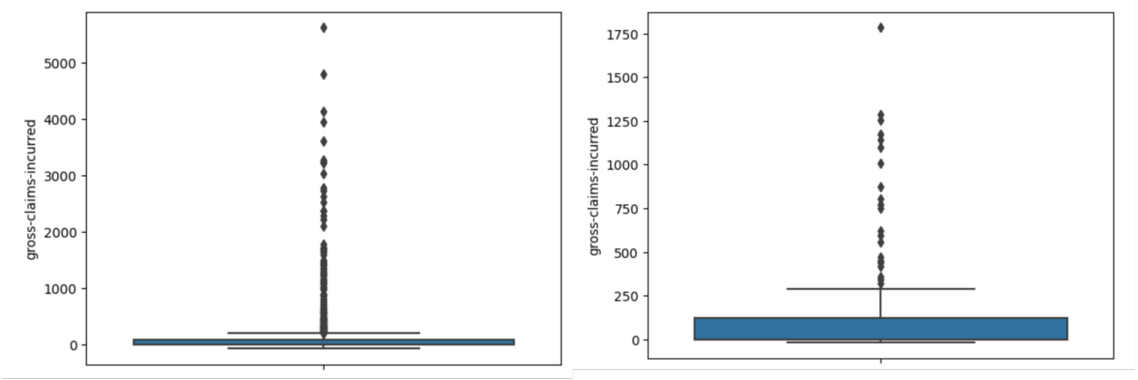

Figure 8. Outlier detection - IQR