Deparment for Big Lorries Data Analysis

Brief created by Nickolas Theodoulou
Amey Strategic Consulting

October 18, 2018

Abstract

The DfBL data set was analysed. The biggest variation in sample size was found between the Financial Year 2000-01 to 2001-02 with a 329% increase. The overall safety condition score dropped from 2000-01 to 2011-12 with the upper quartile decreasing from 70% to 65% and the lower quartile decreasing from 30% to 20% likely due to the vehicle selecting algorithm improving. Iveco and Volvo were found to be the worst safety scoring manufacturers overall. Improvements to the algorithm used to select vehicles for inspection such as knowing the number of fatal UK HGV and LGV accidents is suggested. Stratified Sampling is suggested to improve the sampling methodology. Whether the number of inspections conducted each year is sufficient is found to be inconclusive. Methods of obtaining more data to draw a conclusion are discussed.

1 Year-To-Tear Variation Of The Sample Size

The number of inspections the DfBL conducted had the biggest variation between the Financial Year 2000-01 2001-02 being an increase from 545 to 2338 (329%) as seen in Fig.1. This is likely due to 2000-01 being the first year DfBL conducted inspections. The number of inspections after the Financial Year 2000-01 fluctuate between increasing by a maximum of 36% (between 2001-02 and 2002-03) and decreasing by a maximum of 27% (between 2008-09 and 2009-10): a percentage change much lower than 329%. From 2001-02 to 2008-09 the number of inspections appear the most consistent with a mean: of $\mu = 2853$ and standard deviation: $\sigma = 362$ compared to $\mu = 2525$ and $\sigma = 743$ when all the Financial Years are included. This is likely due to the DfBL being satisfied with the number of inspections conducted between 2001-02 to 2008-09.

One should be extremely hesitant when discarding any data. Especially within the DfBL data set as there isn't enough information present to investigate the cause of outliers. The data obtained between 2000-01 could be considered an outlier due to the smaller sample size. However, a reasonable assumption would be that the DfBL would have spent more time testing and validating the procedure used to quantify the Condition Score of each vehicle. This should have been prioritised over inspecting as many vehicles as possible as it would be essential to use reliable and repeatable testing on a variety of vehicles. Thus, rather than discarding the 2000-01 data this report will analyse results including and excluding the 2000-01 data when necessary. If a model were to be made of the data set, instead of discarding the 2000-01 data, it should instead be weighted lower than the other years.

1.1 The Effects Of Improving The Algorithm

The largest percentage decrease between 2008-09 and 2009-10 at 27% may have occurred due to budget cuts. However, this decrease in the sample size could have been compensated by improving the algorithm that determines which vehicles to inspect. A possible solution of the algorithm would have been to inspect the same amount of Heavy Goods Vehicle (HGV) and inspect less Light Goods Vehicles (LGV) and Personnel Vehicles (PV) from the year 2008-09 to 2009-10. This would be due to HGV's being considered more 'high risk' as the condition score is generally the lowest for them in comparison to the other vehicle types as seen in Fig.3.

A negative effect of improving the algorithm by inspecting previously found poor condition vehicles more often and good condition less often is that it will create a bias to certain manufacturers. This is further discussed in the final paragraph of Sec.3.

2 Condition of Road Vehicles Changing Over Time

The condition score of all the Vehicles and Manufacturer in each consecutive Financial Year is plotted in Fig.2. For each year, the median score is 50 as expected. For the first four years: 2000-01 to 2003-04 there appears to be no change in the overall condition of the inspected vehicles. From 2004-05 to 2007-08, the upper quartile (Q_3) decreases from 70% to 65% and the lower quartile (Q_1) remains the same aside from in the year 2005-06 where it decreases from the previous values 30% to 20%. Thus the overall condition score decreases from 2000-01 to 2007-08. From 2008-09 until 2011-12, (Q_1) decreases from 30% 20%. Thus, the condition score once again lowers for these four years. Overall, since the data was first obtained in 2000-01 up until the last year 2011-12, the overall condition score for all the data decreases.

Assuming that the data is a representative sample of the population, the clear trend is that the conditions of road-going vehicles is deteriorating over time. However, this is a naive conclusion that this does not portray the entire picture as it assumes that a random mixture of both poor and good condition vehicles are inspected each year. The purpose of the DfBL is to ensure that goods vehicles do not pose a danger to Britain's road users thus, it is more efficient for the organisation to inspect vehicles that are more likely to be in poorer condition. A more valid conclusion would be that as time has progressed, the DfBL inspects more high risk vehicles as opposed to the condition of road-going vehicles deteriorating.

3 Comparison Of The Manufacturers

The 'VehicleType' and 'Manufacturer' attributes were combined and a box plot was produced in Fig.3 as opposed to just comparing manufacturers. The reason for this is that it could be the case that the lower scoring manufacturers may happen to have a higher proportion of HGV's and higher scoring manufacturers happen to have a higher proportion of PV's inspected by the DfBL creating a bias.

In Fig.3 it can be seen that DAF, Ford and MAN have the best condition score for for HGV's with $Q_3 = 65\%$, $Q_2 = 50\%$ and $Q_1 = 30\%$. Iveco, Mercedes Benz, Renault, Scania and Volvo have the worst scores with $Q_3 = 60\%$ and $Q_1 = 15\%$ however, the worst performing manufacturer is Mercedes Benz with $Q_2 = 30\%$. For LGV's, the lowest scoring manufacturers are Fiat ($Q_3 = 50\%$), Iveco ($Q_3 = 55\%$) and Volvo ($Q_3 = 55\%$) all with $Q_2 = 40\%$ and $Q_1 = 20\%$. Overall, the condition scores for Personnel Vehicles are much higher for all manufacturers. The worst scoring manufacturers in this category were Volvo with: $Q_1 = 55\%$, $Q_2 = 70\%$, $Q_3 = 80\%$, Fiat and Iveco with: $Q_1 = 60\%$, $Q_2 = 70\%$, $Q_3 = 80\%$. From this analysis, Iveco and Volvo are the two manufacturers that the DfBL should pay special attention to.

As opposed to choosing a sample of the vehicles on the road, an alternative methodology would be using Stratified Sampling. With this method, all the road vehicles should be categorised by manufacturer along with their actual representation in the number of vehicles on the road. Random sampling is then used to select a sufficient number of vehicles from each manufacturer. This method would ensure that all manufacturers are inspected fairly as opposed to inspecting previously found vehicles in poor condition more often. The reason for this is that the manufactures are aware that vehicles previously found in good condition are less likely to be inspected and thus may be less inclined to focus on the safety of their new vehicle if their previous vehicle scored highly.

4 Changing The Number Of Inspections Conducted Each Year

To quantify whether the DfBL is conducting a sufficient number of inspections each year, more data is required. The Transport Statistics Great Britain (TSGB) publishes an annual report on a variety of statistics such as road accident casualties for HGV's and LGV's from 2000 to 2011 (in the 2012 publication) [1]. This along with knowing the percentage of vehicles the DfBL inspect and possibly other information would make it possible to quantify whether the inspections are making an impact. Thus, with this data, it will be possible to conclude if a sufficient number of inspections are conducted each year. The TSGB also publishes information on the average length of haul by type of vehicle each and the type of goods moved by commodity. This information could also be used to improve the algorithm the DfBL use to decide which vehicles to inspect more frequently.

References

[1] P. McEvoy, "Transport statistics great britain 2012," *Department of Transport*, 2012.

A Appendix: Figures

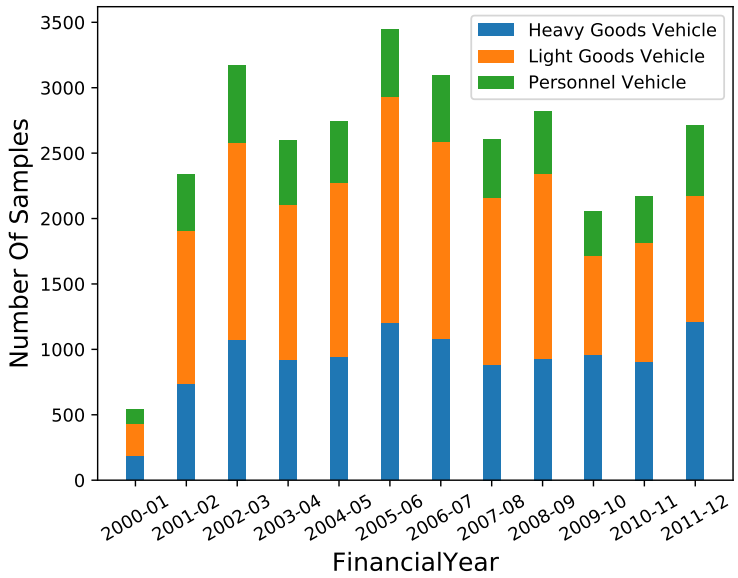

Figure 1: Number of inspections each consecutive Financial Year from 2000-01 to 2011-12 for the three type of vehicles.

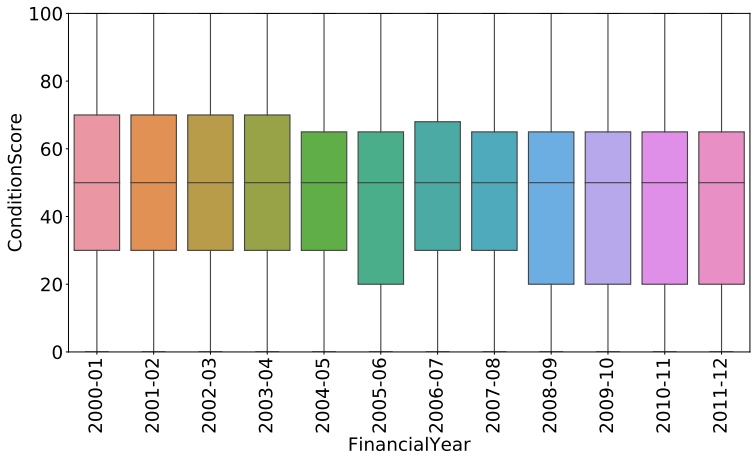

Figure 2: The condition Score of all road-vehicles and manufacturers each consecutive Financial Year from 2000-01 to 2011-12.

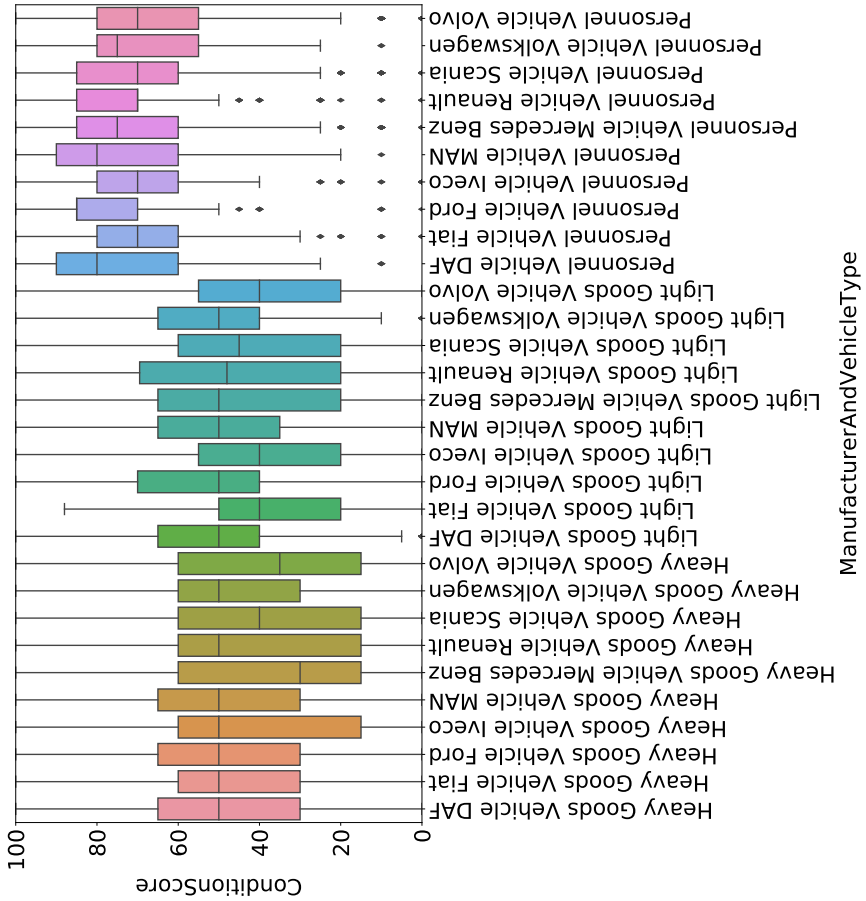

Figure 3: Box plot of the condition Score for Heavy goods vehicles, light goods vehicle and Personnel vehicles categorised by vehicle Manufacturer.