31 March, 2019

Analysis of DfBL Vehicle Inspection 2010–2012

Brief

_____

This brief intends to answer questions—as directed—strictly from the data set of interest: documentation of all the inspections undertaken by The Department for Big Lorries (DfBL) from 2000–2012. It is known that the department's annual inspections are subject to two constraints:

i) The sampling rate should be once every 6-12 years for every goods vehicle on the road.
ii) Vehicles in poor condition should be sampled at a higher rate, and vehicles in good condition should be sampled at a lower rate.

Over the 12 year period 30,305 vehicles were sampled, 20,509 of which were unique. The median and mode condition score was 50 on a 0-100 scale. The analysis found a decline in the mean condition of vehicles being sampled; that the number of inspections undertaken each year were sufficient to meet the constraints; but ultimately, the sampling methodology was not optimum.

I. How does the sample size vary from year-to-year?
Are there any years that the DfBL should exclude from this analysis?

Figure 1 clearly describes the change in sampling volume on a year–on-year basis. Excluding the year 2000–2001, the sampling volume and complexion are similar for each year. Along with being the first year of inspections, the sample size for 2000-2001 was much less at 545, and did not contain any vehicles produced by _Mercedes-Benz_ and only one produced by _Iveco_. However, the sample was still deemed large enough to prove useful, and provided vital information about previously sampled vehicles. This was with the proviso that care should be taken in interpretation.

II. Assuming that the data is a representative sample of the population; does the condition of road-going vehicles appear to be changing over time? If so, is it deteriorating or improving?

Figure 2 depicts the decline in the mean condition of the vehicles being sampled each year, the decline is indeed statistically significant. If the assumption in II. holds, then the condition of vehicles is deteriorating, from a condition of 51.9 ± 2.1 in 2001-2002 to 44.4 ± 1.7 in 2011–2012.

III. Are there any particular vehicle types or manufacturers that the DfBL should pay special attention to? Would you suggest any changes to the sampling methodology to obtain better information for these types?

† If the DfBL wants to achieve i) and ii), it should pay special attention to Light Goods Vehicles, and over-sample these relative to the proportion of goods vehicles that are LGVs on the road in the year. Similarly, the manufacturers *Renault* and *Volvo* should be over-sampled relative to the proportions on the road that year. These classes of vehicle all had lower mean condition scores than 50, and had been relatively under-sampled in the last 12 years. Conversely, the DfBL should under-sample Personnel Vehicles, when the class had been over-sampled.

IV. Since 2000 the vehicle inspectors have slowly improved the algorithm that they use to determine which vehicles to inspect more frequently, so that they are getting better at inspecting vehicles in poor condition more often. What effect could this have on the results?

This would result in a typical selection bias: the samples selected year-on-year would be progressively becoming less representative of the actual population of goods vehicles on the roads. There would be a higher proportion of vehicles in poor condition present in the sample than in the total population on the roads that year. This information voids the assumption in II. and could alone explain the trend seen in Figure 2. It should be noted that the bias is a necessary consequence of the DfB's sampling constraint ii).

V. Is DfBL conducting a sufficient number of inspections a year? Explain and justify your answer.

Yes. Although, more information would be needed to fully ratify this position (†). As if there was a large increase in goods vehicles over the period, then the number would be clearly insufficient; the DfBL would not be on track to achieve the prescribed sampling rates.

Assuming †, there are approximately 20,000 vehicles on the road each year. There are equal proportions of vehicles in poor condition (< 50) and good condition (> 50). Taking a baseline sampling rate of once every 9 years and bounds of 6 and 12, and due to the symmetry, one can make an estimate of the correct number of inspections needed to fulfil the criteria. Essentially half the population should be sampled twice as often, so over 12 years, 3/2 x 20000 vehicles should be sampled. An annual sampling rate of approximately 2,500 would achieve this which is what is observed.

The overall sampling methodology employed is far from optimum: 1,476 inspections were of vehicles already sampled twice in the period, which contravenes ii); 3,922 vehicles were sampled again in the period after achieving a good score (> 50); and, IV. shows how steps could be taken to more consistently sample vehicles in poor condition.

Appendix

† A key assumption for many of the conclusions drawn is that the actual population and distribution of goods vehicles did not change greatly over the period. The analysis would be aided by having access to the data of the actual population of goods vehicles on the road each year, viz. how many there are.
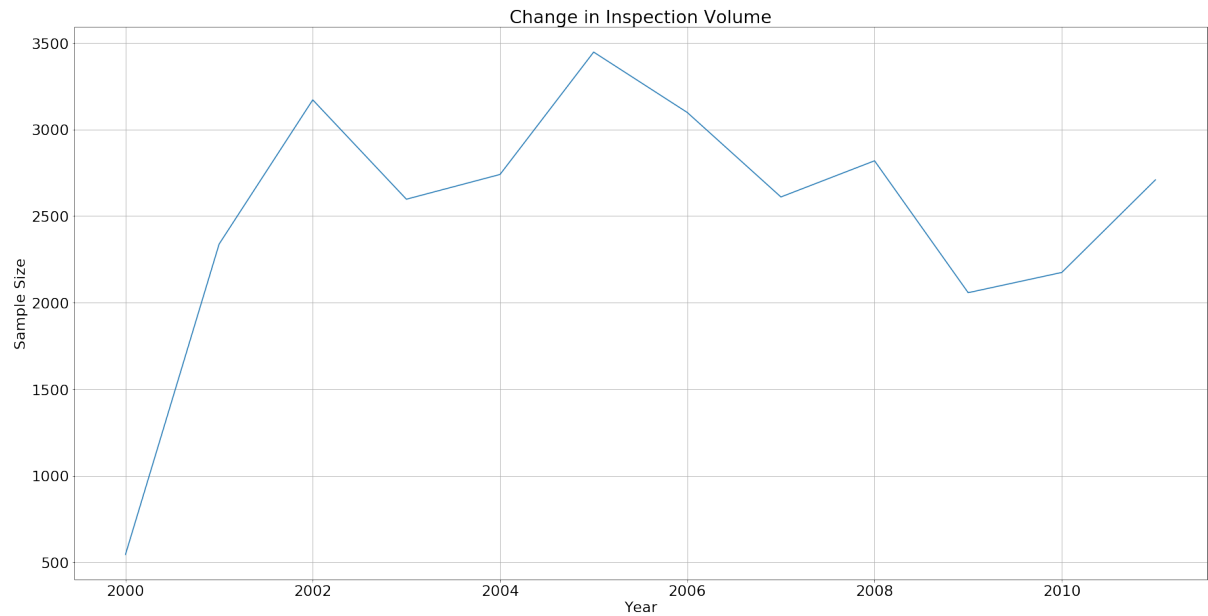

Fig. 1.—Change in sample size for the period (2000–2012). The year 2000-2001 was an outlier in terms of size.
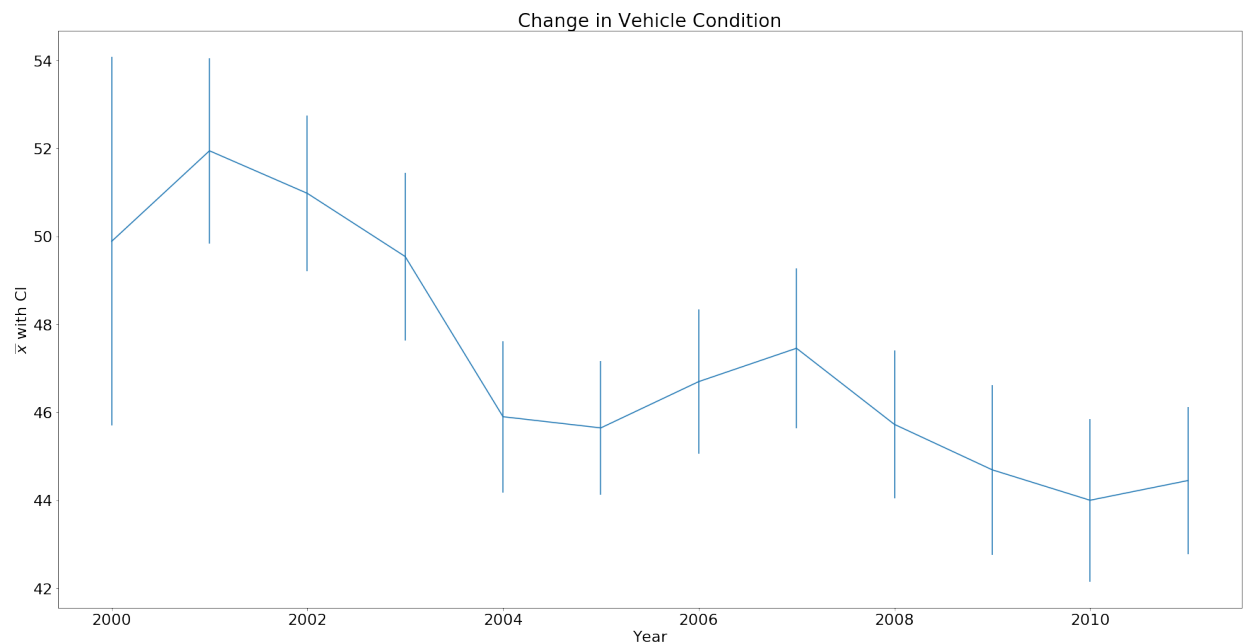

Fig. 2.—Change in sample mean condition for the period (2000–2012). 95% confidence intervals are plotted for each data point, giving an indication of the sample mean certainty.