

Parliamentary Enclosure, the Bootstrap, and a Red Herring or Two

By JOHN R WALTON

Abstract

Chapman finds problems in my use of standard error where none exists and fails to disclose the margin of error associated with his estimate of the extent of parliamentary enclosure.

THE invitation to comment on Chapman's response to my criticism of his enclosure estimates gives me the opportunity to clarify areas of agreement and disagreement. I have no problem with the later sections of Chapman's original paper, nor with the statistical techniques used in those sections.¹ My comment was concerned wholly with Chapman's estimates of the total acreage enclosed, as presented in section I. Had Chapman made it clear that the sample was, as he now says, 'exploratory . . . rather than confirmatory', then I might not have been moved to comment. However, the original paper did not present the resulting acreage estimates in these terms. Rather, the figures, which were said to have implications stretching beyond the question of parliamentary enclosure itself, were offered without any apparent awareness of their likely margin of error. It was merely my intent to suggest that the margin of error was likely to be substantial.

In the course of speculating on its order of magnitude, I pointed out that there was a possibility that Turner's much smaller figure would lie within conventionally-acceptable confidence limits for Chapman's own estimate. This is not to say that I regard it as likely that the extension of Chapman's method to the 90 per cent of English parliamentary enclosures which he has not

examined would produce an enclosure acreage for England as small as Turner's. Indeed, such an outcome is highly unlikely. Chapman's new comparisons of his and Turner's estimates for specific enclosures no more than suggest the same. The point I was trying to make is that the degree of variation in the population from which Chapman's sample has been drawn is so great that the confidence interval associated with his estimate is likely to be extremely wide. My hypothetical estimate of the magnitude of that confidence interval suggested that at the conventionally-acceptable 95 per cent confidence level, Turner's figure might well lie within the interval, albeit close to its limits. In other words, there is a high probability that Chapman's acreage estimate is inaccurate, and a small probability that it is so inaccurate that the true figure would turn out to be no bigger than Turner's.

A further implication of the large magnitude of the standard error is that increasing the size of the sample is unlikely to reduce the confidence interval to acceptable proportions. Of course, it is a matter of judgement as to what constitutes 'acceptable' error. I took Chapman's statement that 'for England. . . the amount of open or common land abolished by enclosure would. . . appear to fall in a range from 7.25 million . . . to 7.35 million [acres]² as

¹ J Chapman, 'The Extent and Nature of Parliamentary Enclosure', *Ag Hist Rev*, 35, 1987, pp 25-35.

² Chapman, *ibid*, p 28.

evidence of Chapman's belief that he had achieved an estimation range of 100,000 acres, and as an indication that he might therefore consider an error of this magnitude acceptable. My calculations indicated that it is not possible to be reasonably certain of having obtained a margin of error of 100,000 acres within the framework of any approach involving sampling, since the sample sizes required to produce such a margin of error, assuming a standard error as large as estimated, far exceed the size of the population.³ This is an argument against the use of sampling. I find it odd that Chapman should present it as an argument against the use of standard error.

Equally odd are Chapman's other reservations about the use of standard error and his preference for the bootstrap method. Contrary to what Chapman implies, there is no restriction (other than perhaps the non-random character of the sample) which would make standard error inappropriate as a method for setting confidence limits around his estimated mean. In large samples (and the minimum size of 'large' in this context is conventionally defined as a great deal smaller than the sample used by Chapman) the sampling distribution of the means is normally distributed about the unknown population mean irrespective of whether the population itself is or is not normally distributed. The standard deviation of the sampling distribution of the means (the standard error of the mean) may therefore serve as a basis for estimating the confidence limits associated with a particular sample mean if the sample is large, irrespective of the statistical distribution of the population from which the sample has been drawn. Only a population exhibiting

extreme skewness may require corrective measures.⁴ Chapman is in error when he suggests otherwise. He is also in error in suggesting that confidence limits for the sample mean cannot be multiplied by the size of the population to yield confidence estimates for the national enclosure acreage. After all, this is no more than an extension of the method Chapman himself uses to turn a sample mean into a national enclosure total.

Thus, recourse to the bootstrap method is utterly unnecessary when, as in this case, there is no objection to the use of standard error. Furthermore, the bootstrap has its own distinctive set of interpretative difficulties. It is perhaps necessary to explain that the technique is based on the attractive if counter-intuitive notion that we may learn more about the population from which a sample has been drawn not by increasing the size of the sample but by taking a large number of smaller samples, with replacement, from the sample itself. The technique was first proposed as an extension or elaboration of the related technique of the jack-knife in 1979.⁵ Since then it has attracted a great deal of attention among mathematical statisticians, but only limited interest among those in other disciplines who use statistical methods, perhaps because it has yet to find its way into the textbooks on which they rely: the book by Noreen referred to by Chapman is the only instance I know of to date.⁶ It is therefore true to say that the limitations of the approach have yet to be fully explored. For Noreen at least, bootstrap methods 'appear to be unreliable and should be used with caution.'⁷ Indeed, there is even confusion about precisely what sample statistics the technique may embrace. Chapman follows one of Noreen's worked

³ J R Walton, 'On Estimating the Extent of Parliamentary Enclosure', *Ag Hist Rev*, 38, 1990, p 82. In fact, required sample size exceeds the size of the population by a great deal more than my estimate indicated, given that the 100,000 acre difference between 7.25 and 7.35 million acres is an estimation range rather than a margin of error. Chapman's estimate has an effective margin of error, defined as half the confidence interval or the distance between the sample mean and one or other confidence limit, of only 50,000 acres.

⁴ See, for example, W G Cochran, *Sampling Techniques*, 3rd edition, New York, 1977, pp 39-44.

⁵ B Efron, 'Bootstrap Methods: Another Look at the Jack-knife', *Annals of Statistics*, 7, 1979, pp 1-26.

⁶ E W Noreen, *Computer-Intensive Methods for Testing Hypotheses: An Introduction*, New York, 1989.

⁷ Noreen, *ibid*, p 63.

examples in applying the bootstrap to the mean. Yet Efron, the pioneer of the bootstrap, appears to have been searching for some technique which provided for estimators other than the mean measures of statistical accuracy analogous to that already provided for the mean by the standard error of the mean. He clearly states that assessing the accuracy of the sample mean is 'a problem where [the bootstrap] is not needed.'⁸

When preparing my critique of Chapman's enclosure estimates, I presumed that in any reply Chapman would disclose the standard error of his sample. I am naturally disappointed that he has not done so, and that I have not been able to establish how close were my guesstimate and his figure. Should Chapman be allowed an additional response, then I hope that this deficiency will be remedied. Alternatively, if Chapman still shuns the standard error in favour of the bootstrap, then it would be interesting to know what is the probability of the true enclosure acreage being equal to or less than both his estimate and perhaps some hypothetical figure or figures intermediate between his own estimate and Turner's.⁹ I

still anticipate that considerable margins of error will be revealed, whichever method is favoured.

There is, however, another matter which seems to me not to have been satisfactorily resolved, and which may yet be resolved in such a way as to demonstrate that my suggested margins of error are overstated. As previously noted, random sampling methods are an essential theoretical underpinning of standard error – and, indeed, of the bootstrap. The application of such techniques to a sample which, as in this case, is not derived randomly therefore poses some problems. The only work known to me which is concerned with this issue goes little further than an assurance that the use of non-random samples does not violate the mechanics of the estimating technique involved.¹⁰ A non-random sample may be regarded as one of the possible array of random samples. And yet one would anticipate that if a non-random sample has been derived in such a way as to increase the likelihood that it reflects variation in the underlying population (which is the essence of systematic sampling procedures, as apparently used by Chapman), then logically the associated margins of error ought to be smaller than those associated with a sample derived randomly in such a way as to offer no additional protection against the possibility that an unrepresentative sample has been identified. Further exploration of sampling theory would appear to be in order.

⁸ B Efron and R Tibshirani, 'Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy', *Statistical Science*, 1, 1986, p 54.

⁹ My understanding of the bootstrap is that the technique generally yields probability estimates of this kind. Noreen's (*op cit* p 74) view of such approaches is that they work 'quite splendidly . . . where the sample is large enough to be indistinguishable from the population.' It is hard to imagine Noreen, or any other statistician, being even this generous to the rather different manifestation of the bootstrap approach which surfaces in the penultimate paragraph of Chapman's comment. Here Chapman gives some curiously precise 'bootstrap' figures for mean and median, which are in fact projected enclosure acreage totals, presumably derived from the means and medians of the bootstrap samples. I can only suppose that the method adopted corresponds to the following sequence. I have harvested a crop of 100 apples and would like to know the mean circumference of the entire crop but am only able to measure a sample of ten apples, which I do. I then rearrange those ten apples into as many different samples of six as are possible (i.e. 10C6, or 210 different samples) calculating the mean circumference of each sample of six. Unsurprisingly, it turns out that the mean of those 210 means is almost identical with the mean of the sample of 10 from which they have all been drawn. From this I conclude that the mean of the sample of ten is almost identical with the mean of the entire crop of 100 apples. Moral: easy access to computing power may addle the brain.

¹⁰ Contrast T M F Smith, 'On the Validity of Inferences from Non-Random Samples', *JRSS, A* 146, 1983, pp 394–403 and Cochran, *op cit*, p 229.