Report on Lance Armstrong's Cycling Efficiency

by Dr Ken Sharpe, Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, AUSTRALIA.

17 November, 2005

Introduction

The purpose of this report is to comment, from a purely statistical point of view, on the claim that the following data, from Coyle[1], demonstrate that Lance Armstrong increased his 'cycling efficiency' appreciably over the seven year period November 1992 to November 1999.

Date	Gross Efficiency (GE)	Delta Efficiency (DE)
November 1992	21.18	21.37
January 1993	21.61	21.75
August 1997	22.66	22.69
November 1999	23.05	23.12
mean	22.125	22.233
standard deviation	0.876	0.811

While the recorded values of both measures of efficiency (GE and DE) have increased over time, it is reasonable to ask whether the increases in the recorded values are greater than could reasonably be due to sampling error.

All measures, and especially those like 'cycling efficiency' which can be influenced by an individual's physical state at the time of measurement, are subject to varying degrees of variation and even measurements taken close together, but far enough apart to avoid carry-over effects, would be expected to differ. One way to think of this is that, at any given time an individual will have a true, underlying 'efficiency', but that the recorded measure of the efficiency will vary around this true value by an amount that depends upon the accuracy with which efficiency can be measured. For example, if you are timing a race using the latest electronic equipment then you would expect the recorded time to be extremely accurate, to the nearest hundredth or even thousandth of a second, but if you were using a simple watch with an analogue second-hand then you would do well to measure the time to the nearest second. The accuracy of a measure is referred to as its 'reproducibility', and special studies conducted to quantify the reproducibility of a measure provide useful insight into the accuracy which can be achieved.

The only study that I am aware of that has investigated the reproducibility of cycling efficiency is the one by Moseley and Jeukendrup[2].

RESPONDENTS 78

In their study, Moseley and Jeukendrup took measurements on 17 subjects on three separate occasions separated by 5-7 days. Using the raw data provided in the paper, the following summary statistics were obtained.

	Gross Efficiency	Delta Efficiency
mean	19.75	25.82
typical error	0.786	1.693

'Typical Error', as defined by Hopkins[3], is the (estimated) standard deviation of the measurement error, after allowing for subject and trial effects. In this situation, with three trials for each subject, the typical errors were obtained using a standard two-way, Analysis of Variance model.

Evaluation of the Armstrong data

There are a number of ways to assess the claim that Armstrong's efficiency did increase over the seven-year period, three of which are detailed below.

1. The overall variation in Armstrong's data

If Armstrong did improve his efficiency over the period, as claimed, then the variation between his (4) values would be expected to be large relative to the typical error. The variation between his values would be expected to have a component due to his improved efficiency, in addition to one due to solely to measurement error. Assuming normality of the measures of efficiency, a formal test is possible using an F-test (with 3 and 32 degrees of freedom).

The null hypothesis is that the variation in Armstrong's data is no greater than measurement error versus the alternative that the variation is greater, due to real changes in his efficiency over the period.

For Gross Efficiency [$F = (0.876/0.786)^2 = 1.242$], the P-value of the test is 0.311 whereas for Delta Efficiency [$F = (0.811/1.693)^2 = 0.229$] the P-value is 0.875; neither test comes even close to achieving statistical significance. (The P-value is the probability of observing variation, in Armstrong's values, greater than or equal to that observed, if there had been no change in his (true) value.) So, unless it can be demonstrated that Moseley and Jeukendrup's data are inappropriate for making the comparison with Armstrong's data, the observed variation in Armstrong's data is consistent with it being due solely to measurement error and hence casts doubt on whether there has been any real change in Armstrong's underlying values.

2. The trend in Armstrong's data

The one aspect of Armstrong's data which does lend some support to the claim that his values have increased is the fact that both the GE and DE values are monotonically increasing (see Figure 1). Assuming that Armstrong's true efficiency had not changed, so that each set of four values is just a random sample from some distribution, then the probability that the four values are in strictly increasing order is 1 in 24.

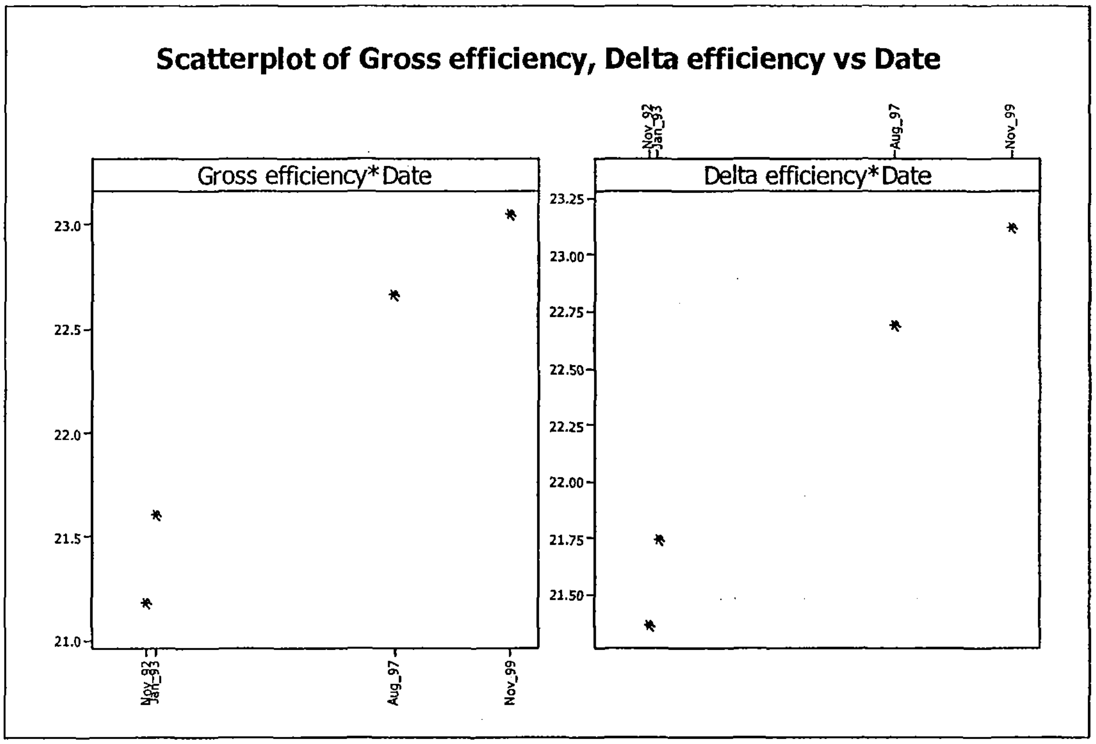

Figure 1. Plot of efficiency versus date for Armstrong's data.

3. Fitting a regression line to Armstrong's data

Another way to assess whether values are increasing (or decreasing) over time is to fit a straight line (a simple regression model) to the data (see Figure 2). For Armstrong's data the regressions are statistically significant with P-values of 0.020 and 0.016 for GE and DE, respectively, which means that the slopes of the lines differ significantly from zero, or that the lines differ significantly from being parallel to the x-axis.

However, the estimated residual standard deviations, a measure of how close the points are to the fitted line, are 0.212 and 0.177 for GE and DE respectively. Now these values should be at least comparable to the typical errors from Moseley and Jeukendrup's data and it is possible to carry out formal comparisons (using 2-sided F tests with 2 and 32 degrees of freedom) which gave P-values of 0.140 and 0.022 for GE and DE respectively. These small P-values, and especially the one for DE, suggest that the precision achieved by Coyle was either very much greater than that achieved by Moseley and Jeukendrup, or that the values for Armstrong lie considerably closer to straight lines than could reasonably be expected by chance alone.

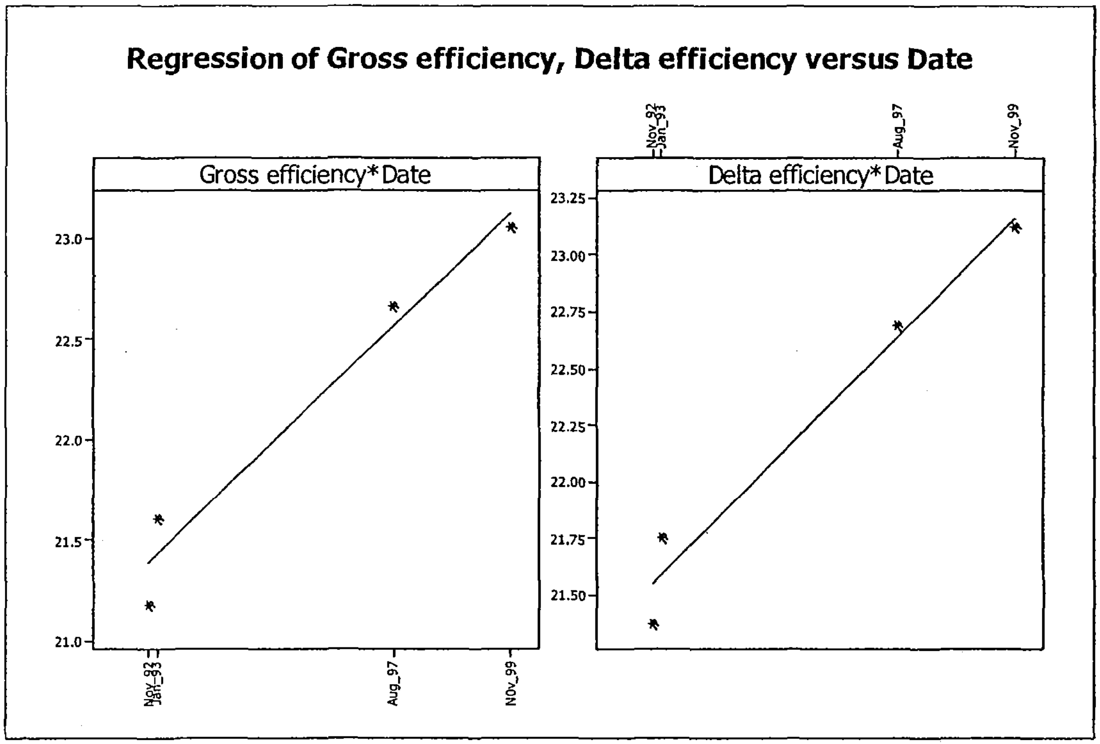

Figure 2. Armstrong's data with fitted regression lines.

Other aspects of Armstrong's data which struck me as being worthy of comment are:

- compared with Moseley and Jeukendrup's data, Armstrong's GE values are on the high side whereas his DE values are on the low side (see Figure 3);
- in all cases, for each pair of values for each individual, the DE value is greater than the corresponding GE value, but whereas the values of the differences (DE-GE) range from 0.8 to 12.0 for Moseley and Jeukendrup's data, they are much, much smaller for Armstrong's data, ranging from 0.03 to 0.19 (see Figure 4). The probability of all of the DE-GE differences for Armstrong being less (or greater) than all of the differences from Moseley and Jeukendrup's 17 subjects is 0.001 which makes it an extremely unexpected outcome;
- the correlation between Armstrong's GE and DE values is extremely high, r = 0.999, compared with a value of r = -0.186 for the Moseley and Jeukendrup data (see Figure 5), another very surprising outcome.

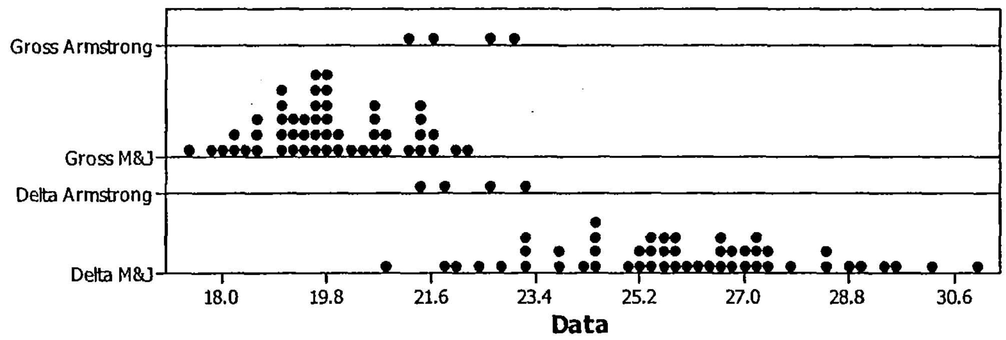

Figure 3. Comparison of Gross and Delta efficiency values from Moseley & Jeukendrup and Armstrong.

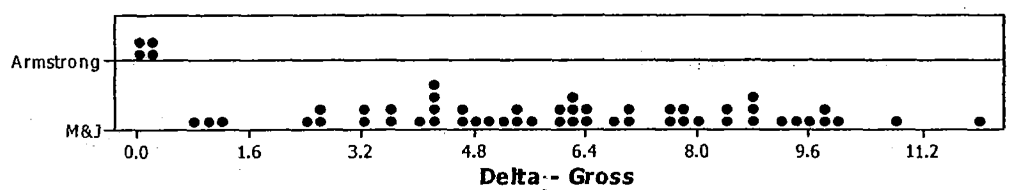

Figure 4. Dotplots of differences between Delta and Gross efficiencies for Armstrong and Moseley & Jeukendrup.

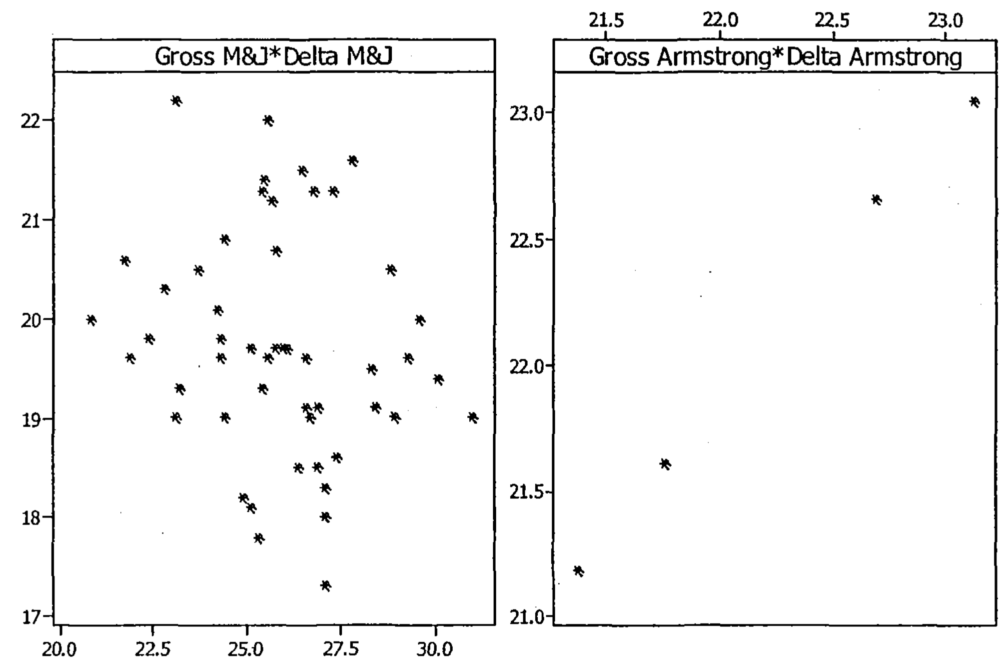

Figure 5. Gross efficiency versus delta efficiency for Moseley & Jeukendrup (left panel) and Armstrong (right panel).

Conclusions

While some aspects of Armstrong's data do suggest that his cycling efficiency has improved over the seven years, the actual change in values is no more than could reasonably be expected due to sampling error for GE and is actually considerably less than could be expected due to sampling error for DE. Further, regression lines fitted to Armstrong's data versus time, resulted in residual standard deviations that are considerably smaller than expected from the Moseley and Jeukendrup evaluation of the reproducibility of the efficiency measures. In other words, if the reproducibility obtained by Moseley and Jeukenprup is applicable, then the fit of the regression lines to Armstrong's data is "too good to be true", especially for Delta efficiency. Other aspects of Armstrong's efficiency data, as described above, also strike me as being extremely unexpected in the light of Moseley and Jeukendrup's, more extensive, data.

Hence, based solely on the data considered here, I am not convinced that Armstrong's efficiency has really increased over the period.

Finally, I was surprised by the small amount of data that is available for Armstrong. Given the amount of testing that elite athletes are subjected to, I would have expected there to be many more values available (for Armstrong) over such a long period. Had such values been available, and had they shown a significant increase, but with a residual standard deviation comparable to that obtained by Moseley and Jeukendrup, then I would have found the claim of an increase in efficiency to have been much more convincing.

References

1. Coyle, E. Improved muscular efficiency displayed as Tour de France champion matures. J. Appl. Physiol. 2005; 98: 2191-2196.
2. Moseley, L and Jeukendrup, A. The reliability of cycling efficiency. Med. Sci. Sports Exerc. 2001; 33: 621-627.
3. Hopkins, W. Measures of reliability in sports medicine and science. Sports Med. 2000; 30(1): 1-15.