Problem 5.4

(a)

(i) The problem is that there is data snooping involved when we manually selects the S&P 500 companies. The hypothesis, which is the companies in this case, is already narrowed down before looking at the data. Therefore the Hoeffding inequality does not apply.

(ii) A better estimate should evaluate all hypothesis (all companies ever traded/is trading), which we have . We then have the estimate:

(b)

(i) There is sampling bias involved. Therefore the results are biased because only currently trading companies are evaluated (S&P 500). That is, all the companies that bankrupted or do not make it to the top five hundred is not considered in the evaluation and yet we do not know if our current trading companies will survive or make to five hundred in the future.

Hence, our training data is not representative of the test data and the corresponding conclusion does not work out for general stock trading.

(ii) Based on the information given, buy and hold is a bad strategy because the calculated company survive rate is approximately, which is low.