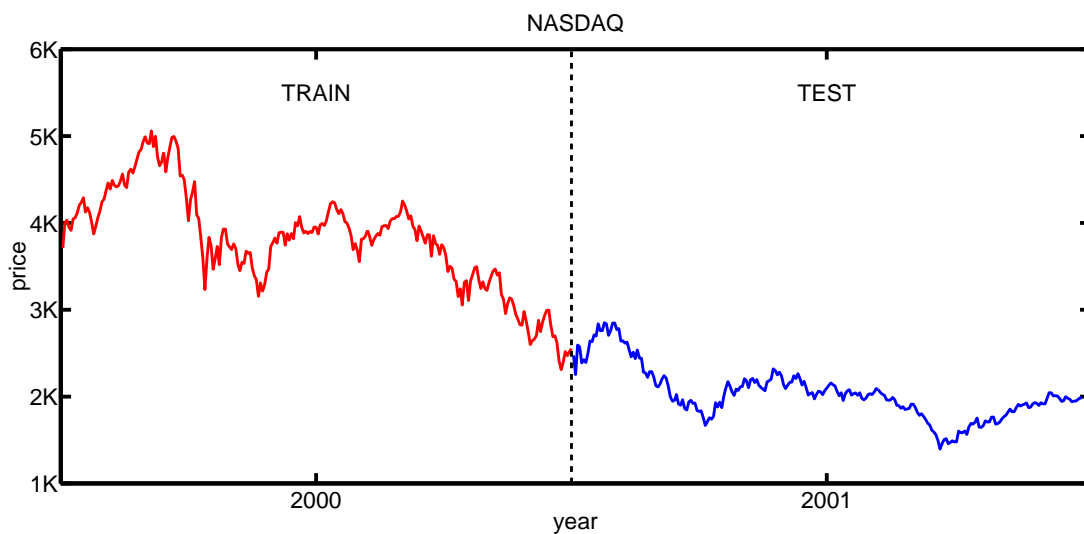## 4.4 Stock market prediction

In this problem, you will apply a simple linear model to predicting the stock market. From the course web site, download the files `nasdaq00.txt` and `nasdaq01.txt`, which contain the NASDAQ indices at the close of business days in 2000 and 2001.

NASDAQ

(a) **Linear coefficients**

How accurately can the index on one day be predicted by a linear combination of the three preceding indices? Using only data from the year 2000, compute the linear coefficients (a_1, a_2, a_3) that maximize the log-likelihood $\mathcal{L} = \sum_t \log P(x_t | x_{t-1}, x_{t-2}, x_{t-3})$, where:

$$P(x_t | x_{t-1}, x_{t-2}, x_{t-3}) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(x_t - a_1 x_{t-1} - a_2 x_{t-2} - a_3 x_{t-3} \right)^2 \right],$$

and the sum is over business days in the year 2000 (starting from the fourth day).

(b) **Root mean squared prediction error**

For the coefficients estimated in part (a), compare the model's performance (in terms of root mean squared error) on the data from the years 2000 and 2001. (A rhetorical question: does a lower prediction error in 2001 indicate that the model worked better that year?)

(c) **Source code**

Turn in a print-out of your source code. You may program in the language of your choice, and you may solve the required system of linear equations either by hand or by using built-in routines (e.g., in Matlab, NumPy).