Solution to Exercise 11.5

⋆11.5 Consider the latent variable model

$$y_t^\circ = \beta_1 + \beta_2 x_t + u_t, \quad u_t \sim \mathrm{N}(0,1),$$
$$y_t = 1 \text{ if } y_t^\circ > 0, \;\; y_t = 0 \text{ if } y_t^\circ \le 0.$$

Suppose that $x_t \sim \mathrm{N}(0,1)$. Generate 500 samples of 20 observations on (x_t, y_t) pairs, 100 assuming that $\beta_1 = 0$ and $\beta_2 = 1$, 100 assuming that $\beta_1 = 1$ and $\beta_2 = 1$, 100 assuming that $\beta_1 = -1$ and $\beta_1 = 1$, 100 assuming that $\beta_1 = 0$ and $\beta_1 = 2$, and 100 assuming that $\beta_1 = 0$ and $\beta_2 = 3$. For each of the 500 samples, attempt to estimate a probit model. In each of the five cases, what proportion of the time does the estimation fail because of perfect classifiers? Explain why there were more failures in some cases than in others.

Repeat this exercise for five sets of 100 samples of size 40, with the same parameter values. What do you conclude about the effect of sample size on the perfect classifier problem?

In order to minimize the effect of experimental randomness, we used 100,000 replications instead of 100. Table S11.1 shows the proportion of the time that perfect classifiers were encountered for each of the five cases and each of the two sample sizes.

Table S11.1 Proportion of samples with perfect classifiers

Parameters	$n = 20$	$n = 40$
$\beta_1 = 0, \beta_2 = 1$	0.0141	0.0001
$\beta_1 = 1, \beta_2 = 1$	0.0590	0.0016
$\beta_1 = -1, \beta_2 = 1$	0.0619	0.0015
$\beta_1 = 0, \beta_2 = 2$	0.1271	0.0075
$\beta_1 = 0, \beta_2 = 3$	0.2923	0.0477

The proportion of samples with perfect classifiers increases as both β_1 and β_2 increase in absolute value. When $\beta_1 = 0$, the unconditional expectation of y_t is 0.5. As β_1 increases in absolute value, this expectation becomes larger, and the proportion of 1s in the sample increases. As β_2 becomes larger in absolute value, the model fits better on average, which obviously increases the chance that it fits perfectly. The results for parameters $(1, 1)$ are almost identical to those for parameters $(-1, 1)$ because, with x_t having mean 0, the fraction of 1s in the samples with parameters $(1, 1)$ is the same, on average, as the fraction of 0s in the samples with parameters $(-1, 1)$.

Comparing the results for $n = 20$ and $n = 40$, it is clear that the probability of encountering a perfect classifier falls very rapidly as the sample size increases.