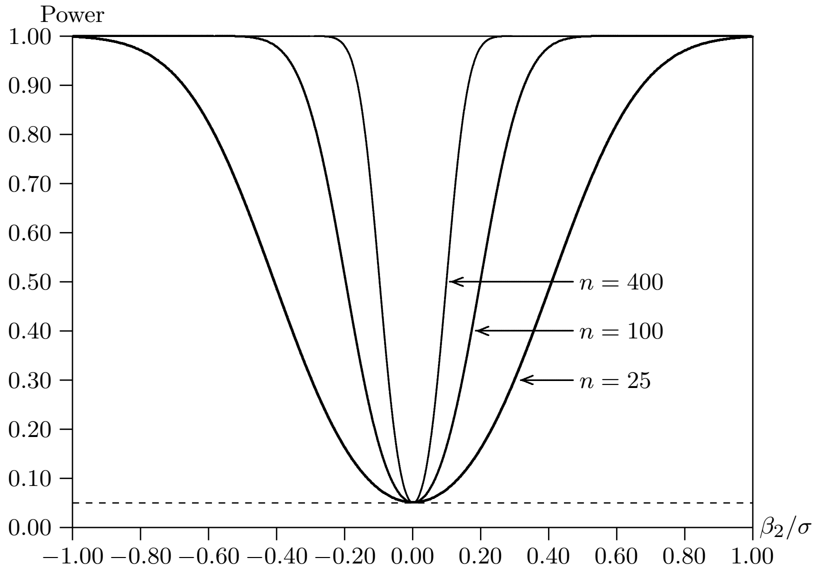

Figure 4.10 Power functions for t tests at the .05 level

Figure 4.10 shows power functions for a very simple model, in which $\boldsymbol{x}_2$, the only regressor, is a constant. Power is plotted as a function of β_2/σ for three sample sizes: $n = 25$, $n = 100$, and $n = 400$. Since the test is exact, all the power functions are equal to .05 when $\beta_2 = 0$. Power then increases as β_2 moves away from 0. As we would expect, the power when $n = 400$ exceeds the power when $n = 100$, which in turn exceeds the power when $n = 25$, for every value of $\beta_2 \neq 0$. It is clear that, as $n \to \infty$, the power function converges to the shape of a T, with the foot of the vertical segment at .05 and the horizontal segment at 1.0. Thus, asymptotically, the test rejects the null with probability 1 whenever it is false. In finite samples, however, we can see from the figure that a false hypothesis is very unlikely to be rejected if $n^{1/2}\beta_2/\sigma$ is sufficiently small.

The Power of Bootstrap Tests

As we remarked in Section 4.6, the power of a bootstrap test depends on B, the number of bootstrap samples. The reason why it does so is illuminating. If, to any test statistic, we add random noise independent of the statistic, we inevitably reduce the power of tests based on that statistic. The bootstrap P value $\hat{p}^*(\hat{\tau})$ defined in (4.62) is simply an estimate of the **ideal bootstrap *P* value**

$$p^*(\hat{\tau}) \equiv \Pr(\tau > \hat{\tau}) = \operatorname*{plim}_{B\to\infty} \hat{p}^*(\hat{\tau}),$$

where $\Pr(\tau > \hat{\tau})$ is evaluated under the bootstrap DGP. When B is finite, $\hat{p}^*$ differs from p^* because of random variation in the bootstrap samples. This

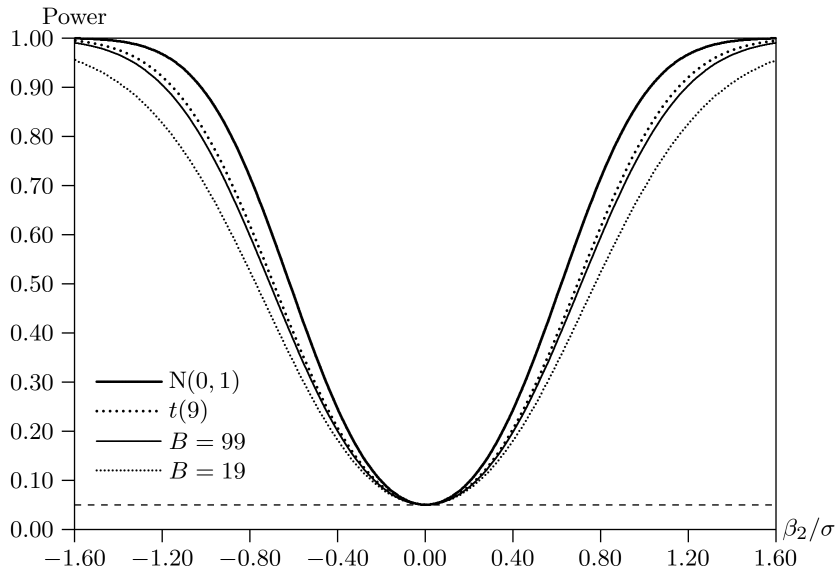

Figure 4.11 Power functions for tests at the .05 level

random variation is generated in the computer, and is therefore completely independent of the random variable τ. The bootstrap testing procedure discussed in Section 4.6 incorporates this random variation, and in so doing it reduces the power of the test.

Another example of how randomness affects test power is provided by the tests z_{β_2} and t_{β_2}, which were discussed in Section 4.4. Recall that z_{β_2} follows the $\mathrm{N}(0,1)$ distribution, because σ is known, and t_{β_2} follows the $t(n-k)$ distribution, because σ has to be estimated. As equation (4.26) shows, t_{β_2} is equal to z_{β_2} times the random variable σ/s, which has the same distribution under the null and alternative hypotheses, and is independent of z_{β_2}. Therefore, multiplying z_{β_2} by σ/s simply adds independent random noise to the test statistic. This additional randomness requires us to use a larger critical value, and that in turn causes the test based on t_{β_2} to be less powerful than the test based on z_{β_2}.

Both types of power loss are illustrated in Figure 4.11. It shows power functions for four tests at the .05 level of the null hypothesis that $\beta_2 = 0$ in the simple model used to generate Figure 4.10, but with only 10 observations. All four tests are exact, as can be seen from the fact that, in all cases, power equals .05 when $\beta_2 = 0$. For all values of $\beta_2 \neq 0$, there is a clear ordering of the four curves in Figure 4.11. The highest curve is for the test based on z_{β_2}, which uses the $\mathrm{N}(0,1)$ distribution and is available only when σ is known. The next three curves are for tests based on t_{β_2}. The loss of power from using t_{β_2} with the $t(9)$ distribution, instead of z_{β_2} with the $\mathrm{N}(0,1)$ distribution, is