would be to have only one column in $\boldsymbol{Z}$, that column being proportional to $\boldsymbol{a}$. In practice, we are rarely in that happy position. There are normally a number of things that we suspect might be wrong with our model and hence a large number of regression directions in which to test. Faced with this situation, there are at least two ways to proceed.

One approach is to test against each type of potential misspecification separately, with each test having only one or a few degrees of freedom. If the model is in fact wrong in one or a few of the regression directions in which these tests are carried out, such a procedure is as likely as any to inform us of that fact. However, the investigator must be careful to control the overall size of the test, since when one does, say, 10 different tests each at the .05 level, the overall size could be as high as .40; see Savin (1980). Moreover, one should avoid jumping to the conclusion that the model is wrong in a particular way just because a certain test statistic is significant. One must remember that $\cos^2\phi$ will often be well above zero for *many* tests, even if only one thing is wrong with the model.

Alternatively, it is possible to test for a great many types of misspecification at once by putting all the regression directions we want to test against into one big $\boldsymbol{Z}$ matrix. This maximizes $\cos^2\phi$ and hence maximizes the chance that the test is consistent, and it also makes it easy to control the size of the test. But because such a test will have many degrees of freedom, power may be poor except when the sample size is large. Moreover, if such a test rejects the null, that rejection gives us very little information as to what may be wrong with the model. Of course, the coefficients on the individual columns of $\boldsymbol{Z}$ in the test regression may well be informative.

This raises the question of what to do when one or more tests reject the null hypothesis. That is a very difficult question, and we will discuss it in Section 12.7.

12.6 Asymptotic Relative Efficiency

Since all consistent tests reject with probability one as the sample size tends to infinity, it is not obvious how to compare the power of tests of which the distributions are known only asymptotically. Various approaches have been proposed in the statistical literature, of which the best known is probably the concept of **asymptotic relative efficiency**, or **ARE**. This concept, which is closely related to the idea of local alternatives, is due to Pitman (1949), and has since been developed by many other authors; see Kendall and Stuart (1979, Chapter 25). Suppose that we have two test statistics, say τ_1 and τ_2, both of which have the same asymptotic distribution under the null and both of which, like all the test statistics we have discussed in this chapter, are root-n consistent. This means that, for the test to have a nondegenerate asymptotic distribution, the drifting DGP must approach the simple null hypothesis at a rate proportional to $n^{-1/2}$. In this case, the asymptotic efficiency of τ_2 relative to τ_1 is defined as

$$\mathrm{ARE}_{21} = \lim_{n\to\infty}\left(\frac{n_1}{n_2}\right),$$

where n_1 and n_2 are sample sizes such that τ_1 and τ_2 have the same power, and the limit is taken as both n_1 and n_2 tend to infinity. If, for example, ARE_{21} were 0.25, τ_2 would asymptotically require 4 times as large a sample as τ_1 to achieve the same power.

For tests with the same number of degrees of freedom, it is easy to see that

$$\mathrm{ARE}_{21} = \frac{\cos^2\phi_2}{\cos^2\phi_1}.$$

Recall from expression (12.23) that the NCP is proportional to $\cos^2\phi$. If the DGP did not drift, it would also be proportional to the sample size. If τ_1 and τ_2 are to be equally powerful in this case, they must have the same NCP. This means that n_1/n_2 must be equal to $\cos^2\phi_2/\cos^2\phi_1$. Suppose, for example, that $\cos^2\phi_1 = 1$ and $\cos^2\phi_2 = 0.5$. Then the implicit alternative hypothesis for τ_1 must include the DGP, while the implicit alternative for τ_2 does not. Thus the directions in which τ_1 is testing explain all of the divergence between the null hypothesis and the DGP, while the directions in which τ_2 is testing explain only half of it. But we can compensate for this reduced explanatory power by making n_2 twice as large as n_1, so as to make both tests equally powerful asymptotically. Hence ARE_{21} must be 0.5. See Davidson and MacKinnon (1987) for more on this special case.

In the more general case in which τ_1 and τ_2 have different numbers of degrees of freedom, calculating the ARE becomes more complicated. The optimal test will be one for which the implicit alternative hypothesis includes the drifting DGP (so that $\cos^2\phi = 1$) and that involves only one degree of freedom. There may of course be many asymptotically equivalent tests that satisfy these criteria, or there may in practice be none at all. Tests that involve more than one degree of freedom, or have $\cos^2\phi < 1$, will be asymptotically less efficient than the optimal test and hence will have AREs less than 1.

The consequences of using tests with $r > 1$ and/or $\cos^2\phi < 1$ are illustrated in Table 12.1. The effect of changing $\cos^2\phi$ does not depend on either the size or power of the test, but the effect of changing r depends on both of these; see Rothe (1981) and Saikkonen (1989). The table was calculated for a size of .05 and powers of .90 (the first entry in each cell) and .50 (the second entry). Each entry in the table is the ARE for the specified test relative to that for the optimal one. Thus each entry may be interpreted as the factor by which the sample size for the optimal test may be smaller than the sample size for the nonoptimal test if both are to have equal asymptotic power.

From Table 12.1, we see that the cost of using a test with a needlessly large number of degrees of freedom, or with a value of $\cos^2\phi$ less than 1,