Our objective is to calculate the NCPs and corresponding values of $\cos^2\phi$ for tests of H_0 against both H_1 and H_2 when the data are generated by (12.28). Thus we will suppose that the data are generated by a drifting DGP that is a special case of H_2. This drifting DGP can be written as

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta}_0 + \alpha_0 n^{-1/2}(\boldsymbol{X}_{t-1}\boldsymbol{\beta}_0 + u_{t-1}) + u_t,\;\; u_t \sim \mathrm{IID}(0, \sigma_0^2). \tag{12.29}$$

Note that this DGP does not involve the recursive calculation of y_t, as (12.28) seems to require, because (12.29) is locally equivalent to (12.28) in the neighborhood of $\delta = 0$ and $\alpha_0 = 0$.

When we test H_0 against H_2, we will be testing in the direction of the DGP and $\cos^2\phi$ will evidently be unity. Using expression (12.25), we see that the NCP for this test is

$$\Lambda_{22} \equiv \frac{\alpha_0^2}{\sigma_0^2} \operatorname*{plim}_{n\to\infty}\Big(\tfrac{1}{n}(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0 + \boldsymbol{u}_{-1})^{\top}\boldsymbol{M}_X(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0 + \boldsymbol{u}_{-1})\Big), \tag{12.30}$$

where $\boldsymbol{u}_{-1}$ and $\boldsymbol{X}_{-1}$ denote, respectively, the vector with typical element u_{t-1} and the matrix with typical row $\boldsymbol{X}_{t-1}$. Here $\boldsymbol{X}_{-1}\boldsymbol{\beta}_0 + \boldsymbol{u}_{-1}$ is playing the role of the vector $\boldsymbol{a}$ in expression (12.25). The notation Λ_{22} means that H_2 is the alternative against which we are testing and that the DGP belongs to H_2. Taking the probability limit, (12.30) becomes

$$\begin{aligned}\Lambda_{22} &= \frac{\alpha_0^2}{\sigma_0^2}\Big(\sigma_0^2 + \operatorname*{plim}_{n\to\infty}\tfrac{1}{n}\big\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\big\|^2\Big)\\ &= \alpha_0^2\Big(1 + \sigma_0^{-2}\operatorname*{plim}_{n\to\infty}\tfrac{1}{n}\big\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\big\|^2\Big).\end{aligned}$$

Now let us see what happens when we test H_0 against H_1. In the neighborhood of H_0, the latter is locally equivalent to

$$\boldsymbol{y} = \boldsymbol{X\beta} + \rho\boldsymbol{u}_{-1} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{IID}(\boldsymbol{0}, \sigma^2\mathbf{I}), \tag{12.31}$$

which avoids the recursive calculation that (12.27) seems to require. Because AR(1) and MA(1) processes are locally equivalent near the point where their respective parameters are zero, this looks like a model with an MA(1) error process. We see from (12.31) that $\boldsymbol{u}_{-1}$ plays the role of $\boldsymbol{Z}$. Once again, $\boldsymbol{X}_{-1}\boldsymbol{\beta}_0 + \boldsymbol{u}_{-1}$ plays the role of $\boldsymbol{a}$. Thus, from (12.18), the NCP is given by

$$\begin{aligned}\Lambda_{12} = \frac{\alpha_0^2}{\sigma_0^2} &\operatorname*{plim}_{n\to\infty}\Big(\tfrac{1}{n}(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0 + \boldsymbol{u}_{-1})^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}\Big)\operatorname*{plim}_{n\to\infty}\Big(\tfrac{1}{n}\boldsymbol{u}_{-1}{}^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}\Big)^{-1}\\ &\times \operatorname*{plim}_{n\to\infty}\Big(\tfrac{1}{n}\boldsymbol{u}_{-1}{}^{\top}\boldsymbol{M}_X(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0 + \boldsymbol{u}_{-1})\Big).\end{aligned} \tag{12.32}$$

Because

$$\begin{aligned}
&\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0+\boldsymbol{u}_{-1})^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}\Big)\\
&=\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\big(\boldsymbol{\beta}_0^{\top}\boldsymbol{X}_{-1}^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}+\boldsymbol{u}_{-1}^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}\big)\Big)=\sigma_0^2,
\end{aligned}$$

expression (12.32) simplifies to

$$\frac{\alpha_0^2}{\sigma_0^2}\sigma_0^2(\sigma_0^{-2})\sigma_0^2=\alpha_0^2.$$

Since the data were generated by a special case of H_2, $\cos^2\phi$ for the test against H_1 is simply the ratio of the NCP Λ_{12} to the NCP Λ_{22}. Thus

$$\begin{aligned}
\cos^2\phi&=\alpha_0^2\bigg(\alpha_0^2\Big(1+\sigma_0^{-2}\operatorname{plim}\frac{1}{n}\big\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\big\|^2\Big)\bigg)^{-1}\\
&=\bigg(1+\frac{\operatorname{plim}n^{-1}\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\|^2}{\sigma_0^2}\bigg)^{-1}.
\end{aligned}\tag{12.33}$$

The second line of (12.33) provides a remarkably simple expression for $\cos^2\phi$ for this special case. It depends only on the ratio of the probability limit of n^{-1} times the squared length of the vector $\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0$ to the variance of the error terms in the DGP (12.29). As this ratio tends to zero, $\cos^2\phi$ tends to unity. Conversely, as this ratio tends to infinity, $\cos^2\phi$ tends to zero. The intuition is very simple. As the ratio of $\operatorname{plim}n^{-1}\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\|^2$ to σ_0^2 tends to zero, because for instance $\boldsymbol{\beta}_0$ tends to zero, $\boldsymbol{M}_X\boldsymbol{y}_{-1}$ (where $\boldsymbol{y}_{-1}$ has typical element y_{t-1}) becomes indistinguishable from $\boldsymbol{M}_X\boldsymbol{u}_{-1}$. When that happens, a test against H_1 becomes indistinguishable from a test against H_2. On the other hand, as the ratio tends in the other direction toward infinity, the correlation between y_{t-1} and u_{t-1} tends to zero, and the directions in which H_1 and H_2 differ from H_0 tend to become mutually orthogonal.

The foregoing analysis could just as easily have been performed under the assumption that the data were generated by a special case of H_1. The drifting DGP would then be

$$y_t=\boldsymbol{X}_t\boldsymbol{\beta}_0+\rho_0 n^{-1/2}u_{t-1}+u_t,\quad u_t\sim\mathrm{IID}(0,\sigma_0^2).$$

When we test H_0 against H_1, $\cos^2\phi$ is now unity, and by an even simpler argument than the one that led to (12.32) we see that the NCP is

$$\Lambda_{11}=\frac{\rho_0^2}{\sigma_0^2}\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{u}_{-1}^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}\Big)=\rho_0^2.$$

Similarly, when we test H_0 against H_2, the NCP is

$$\begin{aligned}\Lambda_{21} = \frac{\rho_0^2}{\sigma_0^2} \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{u}_{-1}{}^{\top}\boldsymbol{M}_X(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0+\boldsymbol{u}_{-1})\Big) \\ \times \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0+\boldsymbol{u}_{-1})^{\top}\boldsymbol{M}_X(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0+\boldsymbol{u}_{-1})\Big)^{-1} \\ \times \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}(\boldsymbol{X}_{-1}\boldsymbol{\beta}_0+\boldsymbol{u}_{-1})^{\top}\boldsymbol{M}_X\boldsymbol{u}_{-1}\Big).\end{aligned}$$

This simplifies to

$$\begin{aligned}&\frac{\rho_0^2}{\sigma_0^2}\,\sigma_0^2\Big(\sigma_0^2+\operatorname{plim}\frac{1}{n}\big\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\big\|^2\Big)^{-1}\sigma_0^2 \\ &= \rho_0^2\Big(1+\sigma_0^{-2}\operatorname{plim}\frac{1}{n}\big\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\big\|^2\Big)^{-1}.\end{aligned}$$

Evidently, $\cos^2\phi$ for the test of H_0 against H_2 is the right-hand expression here divided by ρ_0^2, which is

$$\Big(1+\frac{\operatorname{plim} n^{-1}\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\|^2}{\sigma_0^2}\Big)^{-1}. \tag{12.34}$$

This last result is worth comment. We have found that $\cos^2\phi$ for the test against H_2 when the data were generated by H_1, expression (12.34), is identical to $\cos^2\phi$ for the test against H_1 when the data were generated by H_2, expression (12.33). This result is true not just for this example, but for every case in which both alternatives involve one-degree-of-freedom tests. Geometrically, this equivalence simply reflects the fact that when $\boldsymbol{z}$ is a vector, the angle between $\alpha n^{-1/2}\boldsymbol{M}_X\boldsymbol{a}$ and the projection of $\alpha n^{-1/2}\boldsymbol{M}_X\boldsymbol{a}$ onto $\mathcal{S}(\boldsymbol{X},\boldsymbol{z})$, which is

$$\alpha n^{-1/2}\boldsymbol{M}_X\boldsymbol{z}\big(\boldsymbol{z}^{\top}\boldsymbol{M}_X\boldsymbol{z}\big)^{-1}\boldsymbol{z}^{\top}\boldsymbol{M}_X\boldsymbol{a},$$

is the same as the angle between $\alpha n^{-1/2}\boldsymbol{M}_X\boldsymbol{a}$ and $\alpha n^{-1/2}\boldsymbol{M}_X\boldsymbol{z}$. The reason for this is that $(\boldsymbol{z}^{\top}\boldsymbol{M}_X\boldsymbol{z})^{-1}\boldsymbol{z}^{\top}\boldsymbol{M}_X\boldsymbol{a}$ is a scalar when $\boldsymbol{z}$ is a vector. Hence, if we reverse the roles of $\boldsymbol{a}$ and $\boldsymbol{z}$, the angle is unchanged. This geometrical fact also results in two numerical facts. First, in the regressions

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\alpha}+\gamma\boldsymbol{z} + \text{residuals} \quad \text{and} \\ \boldsymbol{z} &= \boldsymbol{X}\boldsymbol{\beta}+\delta\boldsymbol{y} + \text{residuals},\end{aligned}$$

the t statistic on $\boldsymbol{z}$ in the first is equal to that on $\boldsymbol{y}$ in the second. Second, in the regressions

$$\begin{aligned}\boldsymbol{M}_X\boldsymbol{y} &= \gamma\boldsymbol{M}_X\boldsymbol{z} + \text{residuals} \quad \text{and} \\ \boldsymbol{M}_X\boldsymbol{z} &= \delta\boldsymbol{M}_X\boldsymbol{y} + \text{residuals},\end{aligned}$$

the t statistics on γ and δ are numerically identical and so are the uncentered R^2's.

The analysis of power for this example illustrates the simplicity and generality of the idea of drifting DGPs. Although the case considered is rather simple, it is very commonly encountered in applied work. Regression models with time-series data frequently display evidence of serial correlation in the form of low Durbin-Watson statistics or other significant test statistics for AR(1) errors. We have seen that (except when $\operatorname{plim} n^{-1}\|\boldsymbol{M}_X\boldsymbol{X}_{-1}\boldsymbol{\beta}_0\|^2$ is large relative to σ_0^2) this evidence is almost as consistent with the hypothesis that the model should have included a lagged dependent variable as with the hypothesis that the error terms actually follow an AR(1) process. Thus one should be very cautious indeed when one has to interpret the results of a test against AR(1) errors that rejects the null. One would certainly want to consider several possible alternative models in addition to the alternative that the errors actually follow an AR(1) process. At the very least, before even tentatively accepting that alternative, one would want to subject it to the tests for common factor restrictions that we discussed in Section 10.9.

In the foregoing example, it was easy to evaluate analytically the values of Λ and $\cos^2\phi$ in which we were interested. This will of course not always be the case. However, it is always possible to calculate approximations to these quantities numerically. To do this one simply has to run regression (12.20), evaluating $\boldsymbol{X}(\boldsymbol{\beta})$, $\boldsymbol{a}$, and $\boldsymbol{Z}$ at assumed (or estimated) parameter values. If $\boldsymbol{a}$ and/or $\boldsymbol{Z}$ were stochastic, one would have to generate them randomly and use a very large number of generated observations (which can be obtained by repeating the actual observations as many times as necessary) so as to approximate the desired probability limits. The uncentered R^2 from the regression approximates $\cos^2\phi$ and the explained sum of squares approximates Λ.

12.8 Test Statistics that Do Not Reject the Null

For most of this chapter, we have been concerned with how to interpret test statistics that reject the null hypothesis. In many instances, of course, test statistics fail to reject. Thus it is just as important to know how to interpret a failure to reject as it is to know how to interpret a rejection. Even though we may sometimes speak about "accepting" a null hypothesis when one or more tests fail to reject it, any such acceptance should obviously be provisional and tempered with caution. Just how cautious we should be depends on the power of the test or tests that did not reject the null. We can be most confident about the validity of the null hypothesis if tests that are known to have high power against the alternatives of interest fail to reject it.

As we have seen, the power of a test depends on the way the data are actually generated. In a recent paper, Andrews (1989) has suggested that, as an aid to interpreting nonrejection of a null hypothesis by a particular test, one might consider the power the test would have under the DGPs associated with alternative hypotheses of interest. It seems reasonable that such alternatives