where $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ are the projections complementary to $\boldsymbol{M}_0$ and $\boldsymbol{M}_1$. By the result of Exercise 2.18, $\boldsymbol{P}_1 - \boldsymbol{P}_0$ is an orthogonal projection matrix, which projects on to a space of dimension $k - k_1 = k_2$. Thus the numerator of (6.70) is asymptotically σ_0^2 times a χ^2 variable with k_2 degrees of freedom, divided by $r = k_2$; recall Exercise 4.13. The denominator of (6.70) is just a consistent estimate of σ_0^2, and so, under H_0, the statistic (6.70) itself is asymptotically distributed as $F(k_2, \infty) = \chi^2(k_2)/k_2$.

For linear models, we saw in Section 5.4 that the F statistic could be written as (5.26), which is a special case of the more general form (5.23). Not surprisingly, it is also possible to calculate test statistics of the form (5.23) to test the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$ in the nonlinear model (6.69). This type of test statistic is often called a **Wald statistic**, because the approach was suggested by Wald (1943). It can be written as

$$W_{\boldsymbol{\beta}_2} \equiv \hat{\boldsymbol{\beta}}_2^{\top}\big(\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)\big)^{-1}\hat{\boldsymbol{\beta}}_2, \tag{6.71}$$

where $\hat{\boldsymbol{\beta}}_2$ is a vector of NLS estimates from the unrestricted model (6.69), and $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)$ is the NLS estimate of its covariance matrix. This is just a quadratic form in the vector $\hat{\boldsymbol{\beta}}_2$ and the inverse of an estimate of its covariance matrix. When $k_2 = 1$, the signed square root of (6.71) is equivalent to a t statistic. We will see below that the Wald statistic (6.71) is asymptotically equivalent to the F statistic (6.70), except for the factor of $1/k_2$.

Tests Based on the Gauss-Newton Regression

Since the GNR provides a one-step estimator asymptotically equivalent to the NLS estimator, and it also provides the NLS estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}_2$, a statistic asymptotically equivalent to (6.71) can be computed by means of a GNR. This statistic also turns out to be asymptotically equivalent to the F statistic (6.70), except for the factor of $1/k_2$.

The Gauss-Newton regression corresponding to the model (6.69) is

$$\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \boldsymbol{X}_1(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\boldsymbol{b}_1 + \boldsymbol{X}_2(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\boldsymbol{b}_2 + \text{residuals}, \tag{6.72}$$

where the vector of artificial parameters $\boldsymbol{b}$ has been partitioned as $[\boldsymbol{b}_1 \,\vdots\, \boldsymbol{b}_2]$, conformably with the partition of $\boldsymbol{X}(\boldsymbol{\beta})$. If the GNR is to be used to test the null hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$, the regressand and regressors must be evaluated at parameter estimates which satisfy the null. We will suppose that they are evaluated at the point $\acute{\boldsymbol{\beta}} \equiv [\acute{\boldsymbol{\beta}}_1, \boldsymbol{0}]$, where $\acute{\boldsymbol{\beta}}_1$ may be any root-n consistent estimator of $\boldsymbol{\beta}_1$. Then the one-step estimator of $\boldsymbol{\beta}$ can be written as

$$\acute{\boldsymbol{\beta}} + \acute{\boldsymbol{b}} = \begin{bmatrix} \acute{\boldsymbol{\beta}}_1 + \acute{\boldsymbol{b}}_1 \\ \acute{\boldsymbol{b}}_2 \end{bmatrix}. \tag{6.73}$$

By the results of Section 6.6, $n^{1/2}\acute{\boldsymbol{b}}_2$ is asymptotically equivalent to $n^{1/2}\hat{\boldsymbol{\beta}}_2$ under the null, where $\hat{\boldsymbol{\beta}}_2$ is the NLS estimator of $\boldsymbol{\beta}_2$ from (6.69).

In practice, the two estimators that are most likely to be used for $\acute{\boldsymbol{\beta}}_1$ are $\tilde{\boldsymbol{\beta}}_1$, the restricted NLS estimator, and $\hat{\boldsymbol{\beta}}_1$, a subvector of the unrestricted NLS estimator. Here we are once more adopting the convention, previously used in Chapter 4, whereby a tilde denotes restricted estimates and a hat denotes unrestricted ones. Both these estimators are root-n consistent under the null hypothesis, but $\tilde{\boldsymbol{\beta}}_1$ is generally more efficient than $\hat{\boldsymbol{\beta}}_1$. Whether we want to use $\tilde{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_1$, or some other root-$n$ consistent estimator when performing GNR-based tests depends on how difficult the various estimators are to compute and on the finite-sample properties of the test statistics that result from the various choices.

Now consider the vector of residuals $\acute{\boldsymbol{u}}$ from OLS estimation of the GNR (6.72) evaluated at $\acute{\boldsymbol{\beta}}$, when the true DGP is characterized by the parameter vector $\boldsymbol{\beta}_0 \equiv [\boldsymbol{\beta}_1^0 \,\vdots\, \mathbf{0}]$. Under the null, we have

$$
\begin{aligned}
\acute{\boldsymbol{u}} &= \boldsymbol{y} - \boldsymbol{x}(\acute{\boldsymbol{\beta}}_1, \mathbf{0}) - \acute{\boldsymbol{X}}_1 \acute{\boldsymbol{b}}_1 - \acute{\boldsymbol{X}}_2 \acute{\boldsymbol{b}}_2 \\
&= \boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}_1^0, \mathbf{0}) - \boldsymbol{X}_1(\bar{\boldsymbol{\beta}})(\acute{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) - \acute{\boldsymbol{X}}_1 \acute{\boldsymbol{b}}_1 - \acute{\boldsymbol{X}}_2 \acute{\boldsymbol{b}}_2 \\
&\overset{a}{=} \boldsymbol{u} - \acute{\boldsymbol{X}}_1(\acute{\boldsymbol{\beta}}_1 + \acute{\boldsymbol{b}}_1 - \boldsymbol{\beta}_1^0) - \acute{\boldsymbol{X}}_2 \acute{\boldsymbol{b}}_2.
\end{aligned}
\tag{6.74}
$$

Here, $\bar{\boldsymbol{\beta}}$ is a parameter vector between $\boldsymbol{\beta}_0$ and $\acute{\boldsymbol{\beta}}$. To obtain the asymptotic equality in the last line, we have used the fact that $\boldsymbol{X}_1(\bar{\boldsymbol{\beta}}) \overset{a}{=} \acute{\boldsymbol{X}}_1$. The one-step estimator (6.73) is consistent, and so the last two terms in (6.74) tend to zero as $n \to \infty$. Thus the residuals $\acute{u}_t$ are asymptotically equal to the error terms u_t, and so $n^{-1}\acute{\boldsymbol{u}}^\top\acute{\boldsymbol{u}}$ is asymptotically equal to σ_0^2, the true error variance. In fact, because of the asymptotic equivalence of the one-step estimator $\grave{\boldsymbol{\beta}}$ and the NLS estimator $\hat{\boldsymbol{\beta}}$, (6.74) tells us that $\acute{\boldsymbol{u}} \overset{a}{=} \boldsymbol{u} - \acute{\boldsymbol{X}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. An argument like that of (6.40) then shows that $\acute{\boldsymbol{u}}$ is asymptotically equivalent to $\boldsymbol{M}_{\boldsymbol{X}_0}\boldsymbol{u}$. For the moment, however, we do not need this more refined result.

The GNR (6.72) evaluated at $\acute{\boldsymbol{\beta}}$ is

$$
\boldsymbol{y} - \acute{\boldsymbol{x}} = \acute{\boldsymbol{X}}_1 \boldsymbol{b}_1 + \acute{\boldsymbol{X}}_2 \boldsymbol{b}_2 + \text{residuals}. \tag{6.75}
$$

Since this is a linear regression, we can apply the FWL Theorem to it. Writing $\boldsymbol{M}_{\acute{\boldsymbol{X}}_1}$ for the projection on to $\mathcal{S}^\perp(\acute{\boldsymbol{X}}_1)$, we see that the FWL regression can be written as

$$
\boldsymbol{M}_{\acute{\boldsymbol{X}}_1}(\boldsymbol{y} - \acute{\boldsymbol{x}}) = \boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2 \boldsymbol{b}_2 + \text{residuals}.
$$

This FWL regression yields the same estimates $\acute{\boldsymbol{b}}_2$ as does (6.75). Thus, inserting the factors of powers of n that are needed for asymptotic analysis, we find that

$$
n^{1/2}\acute{\boldsymbol{b}}_2 = (n^{-1}\acute{\boldsymbol{X}}_2^\top \boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2)^{-1} n^{-1/2}\acute{\boldsymbol{X}}_2^\top \boldsymbol{M}_{\acute{\boldsymbol{X}}_1}(\boldsymbol{y} - \acute{\boldsymbol{x}}). \tag{6.76}
$$

In addition to yielding the same parameter estimates $\acute{\boldsymbol{b}}_2$, the FWL regression has the same residuals as regression (6.75) and the same estimated covariance matrix for $\acute{\boldsymbol{b}}_2$. The latter is $\acute{\sigma}^2(\acute{\boldsymbol{X}}_2^\top \boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2)^{-1}$, where $\acute{\sigma}^2$ is the error variance estimator from (6.75), which, as we just saw, is asymptotically equal to σ_0^2. If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ denote $\boldsymbol{X}_1(\boldsymbol{\beta}_0)$ and $\boldsymbol{X}_2(\boldsymbol{\beta}_0)$, respectively, we see that

$$
\begin{aligned}
n^{-1}\acute{\boldsymbol{X}}_2^\top \boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2 &= n^{-1}\acute{\boldsymbol{X}}_2^\top\acute{\boldsymbol{X}}_2 - n^{-1}\acute{\boldsymbol{X}}_2^\top\acute{\boldsymbol{X}}_1(n^{-1}\acute{\boldsymbol{X}}_1^\top\acute{\boldsymbol{X}}_1)^{-1}n^{-1}\acute{\boldsymbol{X}}_1^\top\acute{\boldsymbol{X}}_2 \\
&\overset{a}{=} n^{-1}\boldsymbol{X}_2^\top\boldsymbol{X}_2 - n^{-1}\boldsymbol{X}_2^\top\boldsymbol{X}_1(n^{-1}\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}n^{-1}\boldsymbol{X}_1^\top\boldsymbol{X}_2 \\
&= n^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_{\boldsymbol{X}_1}\boldsymbol{X}_2,
\end{aligned}
$$

where the asymptotic equality follows, as usual, from the consistency of $\acute{\boldsymbol{\beta}}$. Thus n times the covariance matrix estimator for $\acute{\boldsymbol{b}}_2$ given by the GNR (6.75) provides a consistent estimate of the asymptotic covariance matrix of the vector $n^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)$, as would be given by the lower right block of (6.31) if that matrix were partitioned appropriately.

The Wald test statistic (6.71) can be rewritten as

$$
n^{1/2}\hat{\boldsymbol{\beta}}_2^\top\big(n\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)\big)^{-1}n^{1/2}\hat{\boldsymbol{\beta}}_2. \tag{6.77}
$$

Under the null, this is asymptotically equivalent to the statistic

$$
\frac{1}{\acute{\sigma}^2}\,n^{1/2}\acute{\boldsymbol{b}}_2^\top(n^{-1}\acute{\boldsymbol{X}}_2^\top\boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2)\,n^{1/2}\acute{\boldsymbol{b}}_2, \tag{6.78}
$$

which is based entirely on quantities from the GNR (6.75). That (6.77) and (6.78) are asymptotically equal relies on (6.76) and the fact, which we have just shown, that the covariance matrix estimator for $\acute{\boldsymbol{b}}_2$ is also valid for $\hat{\boldsymbol{\beta}}_2$.

By equation (6.76), the GNR-based statistic (6.78) can also be expressed as

$$
\frac{1}{\acute{\sigma}^2}\,n^{-1/2}(\boldsymbol{y}-\acute{\boldsymbol{x}})^\top\boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2(n^{-1}\acute{\boldsymbol{X}}_2^\top\boldsymbol{M}_{\acute{\boldsymbol{X}}_1}\acute{\boldsymbol{X}}_2)^{-1}n^{-1/2}\acute{\boldsymbol{X}}_2^\top\boldsymbol{M}_{\acute{\boldsymbol{X}}_1}(\boldsymbol{y}-\acute{\boldsymbol{x}}). \tag{6.79}
$$

When this statistic is divided by $r = k_2$, we can see by comparison with (4.33) that it is precisely the F statistic for a test of the artificial hypothesis that $\boldsymbol{b}_2 = \mathbf{0}$ in the GNR (6.75). In particular, $\acute{\sigma}^2$ is just the sum of squared residuals from equation (6.75), divided by $n-k$. Thus a valid test statistic can be computed as an ordinary F statistic using the sums of squared residuals from the "restricted" and "unrestricted" GNRs,

$$
\mathrm{GNR}_0:\ \ \boldsymbol{y}-\acute{\boldsymbol{x}} = \acute{\boldsymbol{X}}_1\boldsymbol{b}_1 + \text{residuals, and} \tag{6.80}
$$

$$
\mathrm{GNR}_1:\ \ \boldsymbol{y}-\acute{\boldsymbol{x}} = \acute{\boldsymbol{X}}_1\boldsymbol{b}_1 + \acute{\boldsymbol{X}}_2\boldsymbol{b}_2 + \text{residuals.} \tag{6.81}
$$

In Exercise 6.9, readers are invited to show that such an F statistic is asymptotically equivalent to the F statistic computed from the sums of squared residuals from the two nonlinear regressions (6.68) and (6.69).