and that these are estimated by IV using the instrument matrix $\boldsymbol{W}$. Now suppose that the estimates are actually obtained by two-stage least squares. It is easy to see that the sum of squared residuals from the second-stage regression for (7.43), in which $\boldsymbol{X}_1$ is replaced by $\boldsymbol{P}_W\boldsymbol{X}_1$, will be

$$\text{RSSR}^* \equiv \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y}, \tag{7.45}$$

where $\boldsymbol{M}_1$ denotes the matrix that projects orthogonally onto $\mathcal{S}^\perp(\boldsymbol{P}_W\boldsymbol{X}_1)$. Similarly, it can be shown (doing so is a good exercise) that the sum of squared residuals from the second-stage regression for (7.44) will be

$$\text{USSR}^* \equiv \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y} - \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{P}_W\boldsymbol{X}_2\big(\boldsymbol{X}_2^\top\boldsymbol{P}_W\boldsymbol{M}_1\boldsymbol{P}_W\boldsymbol{X}_2\big)^{-1}\boldsymbol{X}_2^\top\boldsymbol{P}_W\boldsymbol{M}_1\boldsymbol{y}. \tag{7.46}$$

The difference between (7.45) and (7.46) is

$$\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{P}_W\boldsymbol{X}_2\big(\boldsymbol{X}_2^\top\boldsymbol{P}_W\boldsymbol{M}_1\boldsymbol{P}_W\boldsymbol{X}_2\big)^{-1}\boldsymbol{X}_2^\top\boldsymbol{P}_W\boldsymbol{M}_1\boldsymbol{y}, \tag{7.47}$$

which bears a striking and by no means coincidental resemblance to expression (7.41). Under the null hypothesis (7.43), $\boldsymbol{y}$ is equal to $\boldsymbol{X}_1\boldsymbol{\beta}_1+\boldsymbol{u}$. Since $\boldsymbol{P}_W\boldsymbol{M}_1$ annihilates $\boldsymbol{X}_1$, (7.47) reduces to

$$\boldsymbol{u}^\top\boldsymbol{M}_1\boldsymbol{P}_W\boldsymbol{X}_2\big(\boldsymbol{X}_2^\top\boldsymbol{P}_W\boldsymbol{M}_1\boldsymbol{P}_W\boldsymbol{X}_2\big)^{-1}\boldsymbol{X}_2^\top\boldsymbol{P}_W\boldsymbol{M}_1\boldsymbol{u}$$

under the null. It should be easy to see that, under reasonable assumptions, this quantity, divided by anything which estimates σ^2 consistently, will be asymptotically distributed as $\chi^2(r)$. The needed assumptions are essentially (7.18a)–(7.18c), plus assumptions sufficient for a central limit theorem to apply to $n^{-1/2}\boldsymbol{W}^\top\boldsymbol{u}$.

The problem, then, is to estimate σ^2. Notice that $\text{USSR}^*/(n-k)$ does *not* estimate σ^2 consistently, for the reasons discussed in Section 7.5. As we saw there, the residuals from the second-stage regression may be either too large or too small. Thus estimates of σ^2 must be based on the set of residuals $\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}$ rather than the set $\boldsymbol{y}-\boldsymbol{P}_W\boldsymbol{X}\tilde{\boldsymbol{\beta}}$. One valid estimate is $\text{USSR}/(n-k)$, where

$$\text{USSR} \equiv \big\|\boldsymbol{y}-\boldsymbol{X}_1\tilde{\boldsymbol{\beta}}_1-\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2\big\|^2.$$

The analog of (7.42) would then be

$$\frac{(\text{RSSR}^*-\text{USSR}^*)/r}{\text{USSR}/(n-k)} \overset{a}{\sim} F(r, n-k). \tag{7.48}$$

Notice that the numerator and denominator of this test statistic are based on different sets of residuals. The numerator is $1/r$ times the difference between the sums of squared residuals from the second-stage regressions, while the denominator is $1/(n-k)$ times the sum of squared residuals that would be printed by a program for IV estimation.