This looks just like expression (7.59), with $\boldsymbol{A}$ replacing $\boldsymbol{P}_W$, and may be derived in exactly the same way. The first factor in (11.33), $(\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X})^{-1}$, is simply a $k \times k$ matrix with full rank, which will have no effect on any test statistic that we might compute. Therefore, what we really want to do is test whether the vector

$$n^{-1/2}\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{M}_X\boldsymbol{y} \tag{11.34}$$

has mean zero asymptotically. This vector has k elements, but even if $\boldsymbol{A}\boldsymbol{X}$ has full rank, not all those elements may be random variables, because $\boldsymbol{M}_X$ may annihilate some columns of $\boldsymbol{A}\boldsymbol{X}$. Suppose that k^* is the number of linearly independent columns of $\boldsymbol{A}\boldsymbol{X}$ that are not annihilated by $\boldsymbol{M}_X$. Then testing (11.34) is equivalent to testing whether the vector

$$n^{-1/2}\boldsymbol{X}^{*\top}\boldsymbol{A}\boldsymbol{M}_X\boldsymbol{y} \tag{11.35}$$

has mean zero asymptotically, where $\boldsymbol{X}^*$ denotes k^* columns of $\boldsymbol{X}$ with the property that none of the columns of $\boldsymbol{A}\boldsymbol{X}^*$ is annihilated by $\boldsymbol{M}_X$.

Now consider the artificial regression

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{X}^*\boldsymbol{\delta} + \text{residuals}. \tag{11.36}$$

It is easily shown by using the FWL Theorem that the OLS estimate of $\boldsymbol{\delta}$ is

$$\acute{\boldsymbol{\delta}} = \left(\boldsymbol{X}^{*\top}\boldsymbol{A}\boldsymbol{M}_X\boldsymbol{A}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*\top}\boldsymbol{A}\boldsymbol{M}_X\boldsymbol{y},$$

and it is evident that, in general, $\operatorname{plim}(\acute{\boldsymbol{\delta}}) = \boldsymbol{0}$ if and only if (11.35) has mean zero asymptotically. The ordinary F statistic for $\boldsymbol{\delta} = \boldsymbol{0}$ in (11.36) is

$$\frac{\boldsymbol{y}^\top\boldsymbol{P}_{M_XAX^*}\boldsymbol{y}/k^*}{\boldsymbol{y}^\top\boldsymbol{M}_{X,\,M_XAX^*}\boldsymbol{y}/(n-k-k^*)}, \tag{11.37}$$

where $\boldsymbol{P}_{M_XAX^*}$ is the matrix that projects onto $\mathcal{S}(\boldsymbol{M}_X\boldsymbol{A}\boldsymbol{X}^*)$, and $\boldsymbol{M}_{X,\,M_XAX^*}$ is the matrix that projects onto $\mathcal{S}^\perp(\boldsymbol{X}, \boldsymbol{M}_X\boldsymbol{A}\boldsymbol{X}^*)$. If (11.27) actually generated the data, the statistic (11.37) will certainly be valid asymptotically, since the denominator will then consistently estimate σ^2. It will be exactly distributed as $F(k^*, n-k-k^*)$ in finite samples if the u_t's in (11.27) are normally distributed and $\boldsymbol{X}$ and $\boldsymbol{A}$ can be treated as fixed. Regression (11.36) and expression (11.37) are essentially the same as regression (7.62) and expression (7.64), respectively; the latter are special cases of the former.

The most common type of DWH test is the one we dealt with in Section 7.9, which asks whether least squares estimates are consistent when some of the regressors may be correlated with the error terms. However, there are numerous other possibilities. For example, $\check{\boldsymbol{\beta}}$ might be the OLS estimator for $\boldsymbol{\beta}$ in the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{u}, \tag{11.38}$$