In practice, we are interested in the covariance matrix of $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ rather than that of $n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, and we will not know σ_0. We may estimate σ^2 by

$$\tilde{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}).$$

It would of course be possible to divide by $n-k$ rather than n here, but that is not necessarily a good idea. Since the SSR is *not* the value of the objective function for IV estimation (in contrast to the situation for least squares), its expectation is not necessarily smaller than $n\sigma_0^2$ and certainly is not equal to $(n-k)\sigma_0^2$. Asymptotically, of course, it makes no difference whether we divide by n or $n-k$. However we define $\tilde{\sigma}$, we will estimate the covariance matrix of $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ by

$$\tilde{\boldsymbol{V}}(\tilde{\boldsymbol{\beta}}) = \tilde{\sigma}^2\big(\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{X}\big)^{-1}. \tag{7.24}$$

The IV estimator $\tilde{\boldsymbol{\beta}}$ that we have been discussing is actually a **generalized IV estimator**. It may be contrasted with the **simple IV estimator** that is discussed in many elementary statistics and econometrics texts and that was developed first. For the simple IV estimator, each explanatory variable has associated with it a single instrument, which may be the variable itself if it is assumed uncorrelated with $\boldsymbol{u}$. Thus the matrix $\boldsymbol{W}$ has the same dimensions, $n \times k$, as the matrix $\boldsymbol{X}$. In this special case, the generalized IV estimator (7.17) simplifies substantially:

$$\begin{aligned}
\tilde{\boldsymbol{\beta}} &= \big(\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{P}_W\boldsymbol{y} \\
&= \big(\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{y} \\
&= \big(\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{W}^\top\boldsymbol{W}\big(\boldsymbol{X}^\top\boldsymbol{W}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{W}\big(\boldsymbol{W}^\top\boldsymbol{W}\big)^{-1}\boldsymbol{W}^\top\boldsymbol{y} \\
&= \big(\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{W}^\top\boldsymbol{W}\big(\boldsymbol{W}^\top\boldsymbol{W}\big)^{-1}\boldsymbol{W}^\top\boldsymbol{y} \\
&= \big(\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{W}^\top\boldsymbol{y}.
\end{aligned} \tag{7.25}$$

The key result here is that

$$\big(\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1} = \big(\boldsymbol{W}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{W}^\top\boldsymbol{W}\big(\boldsymbol{X}^\top\boldsymbol{W}\big)^{-1},$$

which depends on the facts that the matrix $\boldsymbol{W}^\top\boldsymbol{X}$ is square and has full rank. The last line of (7.25) is the formula for the simple IV estimator that appears in many textbooks. We will not discuss this estimator further in this chapter but will encounter it again when we discuss the generalized method of moments in Chapter 17.

The biggest problem with using IV procedures in practice is choosing the matrix of instruments $\boldsymbol{W}$. Even though every valid set of instruments will yield consistent estimates, different choices will yield different estimates in any finite sample. When using time-series data, it is natural to use lagged