to be uncorrelated with y_i^* and u_i^*. Therefore, we run the regression

$$c_i = \beta_1 + \beta_2 y_i + u_i.$$

Under the plausible assumption that the true value β_{20} is positive, show that y_i is negatively correlated with u_i. Using this result, evaluate the plim of the OLS estimator $\hat{\beta}_2$, and show that this plim is less than β_{20}.

8.2 Consider the simple IV estimator (8.12), computed first with an $n \times k$ matrix $\boldsymbol{W}$ of instrumental variables, and then with another $n \times k$ matrix $\boldsymbol{WJ}$, where $\boldsymbol{J}$ is a $k \times k$ nonsingular matrix. Show that the two estimators coincide. Why does this fact show that (8.12) depends on $\boldsymbol{W}$ only through the orthogonal projection matrix $\boldsymbol{P_W}$?

8.3 Show that, if the matrix of instrumental variables $\boldsymbol{W}$ is $n \times k$, with the same dimensions as the matrix $\boldsymbol{X}$ of explanatory variables, then the generalized IV estimator (8.29) is identical to the simple IV estimator (8.12).

8.4 Show that minimizing the criterion function (8.30) with respect to $\boldsymbol{\beta}$ yields the generalized IV estimator (8.29).

⋆8.5 Under the usual assumptions of this chapter, including (8.16), show that the plim of

$$\frac{1}{n}Q(\boldsymbol{\beta}_0, \boldsymbol{y}) = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)^\top \boldsymbol{P_W}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)$$

is zero if $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u}$. Under the same assumptions, along with the asymptotic identification condition that $\boldsymbol{S}_{\boldsymbol{X}^\top\boldsymbol{W}}(\boldsymbol{S}_{\boldsymbol{W}^\top\boldsymbol{W}})^{-1}\boldsymbol{S}_{\boldsymbol{W}^\top\boldsymbol{X}}$ has full rank, show further that $\operatorname{plim} n^{-1}Q(\boldsymbol{\beta}, \boldsymbol{y})$ is strictly positive for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$.

8.6 Under assumption (8.16) and the asymptotic identification condition that $\boldsymbol{S}_{\boldsymbol{X}^\top\boldsymbol{W}}(\boldsymbol{S}_{\boldsymbol{W}^\top\boldsymbol{W}})^{-1}\boldsymbol{S}_{\boldsymbol{W}^\top\boldsymbol{X}}$ has full rank, show that the GIV estimator $\hat{\boldsymbol{\beta}}_{\text{IV}}$ is consistent by explicitly computing the probability limit of the estimator for a DGP such that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u}$.

8.7 Suppose that you can apply a central limit theorem to the vector $n^{-1/2}\boldsymbol{W}^\top\boldsymbol{u}$, with the result that it is asymptotically multivariate normal, with mean $\mathbf{0}$ and covariance matrix (8.33). Use equation (8.32) to demonstrate explicitly that, if $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u}$, then $n^{1/2}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0)$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix (8.17).

8.8 Suppose that $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are, respectively, $n \times l_1$ and $n \times l_2$ matrices of instruments, and that $\boldsymbol{W}_2$ consists of $\boldsymbol{W}_1$ plus $l_2 - l_1$ additional columns. Prove that the generalized IV estimator using $\boldsymbol{W}_2$ is asymptotically more efficient than the generalized IV estimator using $\boldsymbol{W}_1$. To do this, you need to show that the matrix $(\boldsymbol{X}^\top\boldsymbol{P}_{\boldsymbol{W}_1}\boldsymbol{X})^{-1} - (\boldsymbol{X}^\top\boldsymbol{P}_{\boldsymbol{W}_2}\boldsymbol{X})^{-1}$ is positive semidefinite. **Hint:** see Exercise 3.8.

8.9 Show that the simple IV estimator defined in (8.41) is unbiased when the data are generated by (8.40) with $\sigma_v = 0$. Interpret this result.

8.10 Use the DGP (8.40) to generate at least 1000 sets of simulated data for $\boldsymbol{x}$ and $\boldsymbol{y}$ with sample size $n = 10$, using normally distributed error terms and parameter values $\sigma_u = \sigma_v = 1$, $\pi_0 = 1$, $\beta_0 = 0$, and $\rho = 0.5$. For the exogenous instrument $\boldsymbol{w}$, use independent drawings from the standard normal distribution, and then rescale $\boldsymbol{w}$ so that $\boldsymbol{w}^\top\boldsymbol{w}$ is equal to n, rather than 1 as in Section 8.4.