The following is the true model:

\[ Y_i = \beta_1 + \beta_2 * X_{2i} + \beta_3 * X_{3i} + \beta_4 * X_{4i} + \epsilon_i \]
\[ X_{4i} = \psi_1 + \psi_2 * X_{2i} + \psi_3 * X_{3i} + \psi_4 * Y_i + \psi_5 * Z_{1i} + \psi_6 * Z_{2i} + \varepsilon_i \]

The parameter values are: \( \beta = \{3,2,5,3\} \), \( \psi = \{1,6,8,5,2,3\} \). Unless instructed otherwise, the variance of each of the disturbances is 1 (mean 0).

To do the following you will simulate the data in Gauss as needed. Unless instructed otherwise, draw all values from univariate normal distributions with mean 0, variance 1.

1) Compare the RMSE of \( \hat{\beta}_4 \) from OLS and 2SLS for \( N = 100 \) and \( N = 10,000 \) by running 5000 Trials of each.

2) Plot the estimates of \( \beta_4 \) from part (1) – you should have 4 separate plots: \( \hat{\beta}_{OLS}^{100} \), \( \hat{\beta}_{OLS}^{10000} \), \( \hat{\beta}_{2SLS}^{100} \), and \( \hat{\beta}_{2SLS}^{10000} \).

3) Now say the variance of the disturbance in the equation for \( X_4 (\varepsilon) \) is 100 rather than 1. Repeat (1) and (2) above.

4) Explain why things are different with the higher variance of the disturbance.