Suppose that you are estimating the relationship between Y and X and that you expect that the effect of X on Y occurs over multiple periods rather than immediately.

1. Consider the model $Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-1} + \ldots + \beta_p X_{t-p} + \varepsilon_t$.

      a. How would you decide on the appropriate lag length, *i.e.,* the appropriate value of p?

      b. Why would you expect that there might be multicollinearity? How would that affect your estimators? What might you do about it?

2. Consider as an alternative the Koyck lag: $Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + \varepsilon_t$.

      a. Why is this model less prone to multicollinearity?

      b. Discuss the number of degrees of freedom here vs. in the general distributed lag model of the previous problem.

      c. What is the effect of X_{t-5} on Y_t in this model? What happens to the effect of X_{t-s} on Y_t as lag length s gets large? Does this seem realistic?

      d. Why is it almost inevitable that ε_t is correlated with Y_{t-1}? What are the implications of this correlation for the OLS estimators?

      e. The Prais-Winsten procedure uses the OLS residuals (based on the OLS coefficient estimates) to calculate $\hat{\rho}$. Is this valid? Explain.