

Due Thursday, November 3
(20 points)

This assignment covers panel data methods. The necessary data are contained in the Excel file WAGEPAN308.xls available from the website <http://u.arizona.edu/~rlo>. Be sure to attach the supporting computer print out to the completed assignment and make clear where your answers are shown.

The data for this exercise are a national sample of $i = 1, 2, \dots, 545$ male workers over the 8 year period $t = 1980, \dots, 1987$.

Five models of wage determination are given by

$$lwage_{it} = \alpha + \beta_1 exper_{it} + \beta_2 expersq_{it} + \beta_3 married_{it} + \beta_4 union_{it} + \varepsilon_{it} \quad (1)$$

$$lwage_{it} = \alpha + \beta_1 exper_{it} + \beta_2 expersq_{it} + \beta_3 married_{it} + \beta_4 union_{it} + \gamma_1 educ_i + \gamma_2 black_i + \gamma_3 hisp_i + \varepsilon_{it}$$

$$lwage_{it} = \alpha_i + \beta_1 exper_{it} + \beta_2 expersq_{it} + \beta_3 married_{it} + \beta_4 union_{it} + \varepsilon_{it} \quad (3)$$

$$lwage_{it} = \alpha + \beta_1 exper_{it} + \beta_2 expersq_{it} + \beta_3 married_{it} + \beta_4 union_{it} + u_i + \varepsilon_{it} \quad (4)$$

$$lwage_{it} = \alpha + \beta_1 exper_{it} + \beta_2 expersq_{it} + \beta_3 married_{it} + \beta_4 union_{it} + \gamma_1 educ_i + \gamma_2 black_i + \gamma_3 hisp_i + u_i + \varepsilon_{it}, \quad (5)$$

The time varying covariates are $lwage_{it}$ (log of the hourly wage), $exper_{it}$ (years of work experience), $expersq_{it}$ (square of years of work experience), $married_{it}$ (indicator for currently married), and $union_{it}$ (indicator for current union membership). The time invariant covariates are $educ_i$ (highest years of schooling as of 1980), and $black_i$ and $hisp_i$ are indicators for whether or not the worker is black or hispanic. Each individual is identified by a unique number contained in the variable nr . You may assume $\begin{pmatrix} u_i \\ \varepsilon_{it} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \right]$. You may also assume that u_i and ε_{it} are distributed independently of the regressors except where otherwise indicated.

1. Estimate all five models using the most appropriate estimator in each case and defend your choice of estimator.

2. All of the following statistical tests should be conducted at the 5% level of significance. Be sure to explicitly calculate the test statistics and explain exactly what is being tested.
 - a. LM test of model 1 against model 4.
 - b. χ^2 test of model 4 against model 3.
 - c. F test of model 1 against model 3.
 - d. F test of model 2 against model 3.

3. In model 5 suppose it were the case that $cov(union_{it}, u_i) \neq 0$ and $cov(educ_i, u_i) \neq 0$.
 - a. Use a consistent and asymptotically efficient estimator to estimate model 5.
 - b. Develop and carry out a test at the 5% level of significance that might be indicative of whether or not it is the case that $cov(union_{it}, u_i) \neq 0$ and $cov(educ_i, u_i) \neq 0$. Hint: compare model 3 with model 5.