Sample Questions

(1) Consider the following wage equation for n individuals observed over T periods

$$Y_{it} = X_{it}\beta + \varepsilon_{it} \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T \tag{1}$$

where $Y_{it} = \log(wage_{it})$, X_{it} is a $1 \times k$ vector of observable characteristics that affect the wage of individual i at time period t, and β is a $k \times 1$ vector of coefficients.

(a) If you estimate equation (1) by OLS, what are the assumptions you would be making including assumptions about the properties of ε_{it}.

(b) Suppose you know that $Var(\varepsilon_{it}) = \sigma_i^2$, $i = 1, \ldots, n$.

(i) Describe an *approximate* maximum likelihood procedure to estimate β.

(ii) Derive the lagrange multiplier test statistic to test the hypothesis

$$\text{H}_0 : \sigma_1^2 = \ldots = \sigma_n^2 \quad \text{v.s.} \quad \text{H}_1 : \sim \text{H}_0$$

(c) Suppose you believe that a constant conditional mean function for each individual i is not a realistic assumption.

(i) How would you modify equation (1) if you believe that individual specific constant terms α_i would be enough to capture cross-sectional heterogeneity. Also explain in detail how you would estimate these individual constant terms as well as the β vector.

(ii) How would you modify equation (1) if you believe that unobserved heterogeneity (in the form of individual productivity) should be controlled for in the wage equation by setting α_i from part (i) to $\alpha_i = \alpha + prod_i$ where $prod_i$ is individual productivity. Explain the assumptions that one needs to make about $prod_i$. Also describe in detail an *efficient* way of estimating β.

(iii) How would you determine whether the model in part (i) or (ii) is a better model given the data.

(2) Consider the model specified by

$$Y_{t1} = \gamma_{10} + \beta_{12}Y_{t2} + \beta_{13}Y_{t3} + \gamma_{11}X_{t1} + u_{t1} \tag{1}$$

$$Y_{t2} = \gamma_{20} + \beta_{23}Y_{t3} + \gamma_{24}Y_{t-1,1} + \gamma_{22}X_{t2} + u_{t2} \tag{2}$$

$$Y_{t3} = \gamma_{30} + \beta_{31}Y_{t1} + \beta_{32}Y_{t2} + \gamma_{33}X_{t3} + u_{t3} \tag{3}$$

where the X's are exogenous variables. Assume that the error terms have zero means, constant variances, and are not serially correlated.

(a) Determine the identification status of equation (1) according to the order and rank conditions.

(b) Suppose someone suggests using the seemingly unrelated regressions (SUR) technique to estimate the unrestricted reduced form equations for the above model. Comment on this suggestion.

(3) Consider the following model for individual data to test whether nutrition affects productivity

$$\log(produc) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + \beta_4 calories + u$$

where *produc* is some measure of worker productivity, *exper* is labor market experience, *educ* is years of schooling, and *calories* is caloric intake per day. Assume that *exper*, $exper^2$, and *educ* are all exogenous. The variable *calories* is possibly correlated with u. Possible instrumental variables (IVs) for *calories* are prices of various goods such as grains, meats, breads, dairy products, etc.

(a) Determine whether the coefficient vector $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^\top$ is consistently estimated by OLS if *calories* variable is indeed correlated with u.

(b) Under what circumstances do prices make good IVs for *calories*?

(c) How many prices are needed to identify the equation?

(d) Suppose we have m prices, $p_1, \ldots, p_m$. Explain how to test the null hypothesis that *calories* is exogenous.

(4) Consider the classical linear regression model $Y = X\beta + u$ where Y is $n \times 1$, X is $n \times k$, β is $k \times 1$, and $u \sim \mathrm{N}(0, \sigma_u^2 \mathrm{I}_n)$. Derive the lagrange multiplier test statistic to test the null hypothesis $\mathrm{H}_0 : A\beta = a$ for q linear restrictions where A is $q \times k$ and a is $q \times 1$.