Question 1: Non-linear Least Squares

Let y be an outcome variable and x be a vector of explanatory variables. Assume that the conditional distribution of y given x is the Poisson distribution, i.e., that:

$$\mathbb{P}(y = k|x) = \frac{exp(kx'\theta_0)exp(-exp(x'\theta_0))}{k!}, \quad k = 0, 1, 2, ...$$

where θ_0 is the true value of the vector of parameters. This implies that the conditional mean of y given x is: $\mathbb{E}[y|x] = exp(x'\theta_0)$ and the conditional variance is: $Var(y|x) = exp(x'\theta_0)$. Suppose that you have a random sample of size n on observations (y_i, x_i).

Consider the non-linear least squares estimator of θ_0:

$$\hat{\theta}_{NLLS} = \underset{\theta \in \Theta}{\text{argmax}} -\frac{1}{n}\sum_{i=1}^{n}(y_i - exp(x_i'\theta))^2$$

(a) Is $\hat{\theta}_{NLLS}$ consistent for θ_0? (Hint: verify regularity conditions.)

(b) Derive the asymptotic distribution of $\hat{\theta}_{NLLS}$. (For (b)-(d) you can assume that the necessary regularity conditions hold.)

(c) Derive the asymptotic distribution of the maximum likelihood estimator of θ_0. Hint: the joint distribution is:

$$f(x_1, \ldots, x_n; \theta) = \prod_i \frac{e^{y_i x_i'\theta} e^{-e^{x_i'\theta}}}{y_i!}$$

The log likelihood is therefore:

$$\ln L(\theta) = \sum_i \left(y_i x_i'\theta - e^{x_i'\theta} + y_i! \right)$$

(d) Because the conditional variance of y given x depends on θ_0, we can obtain a more efficient NLLS estimator by using $1/Var(y|x)$ as a weight. Derive the asymptotic distribution of the weighted NLLS estimator:

$$\hat{\theta}_{WNLLS} = \underset{\theta \in \Theta}{\text{argmax}} -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{exp(x_i'\theta_0)}(y_i - exp(x_i'\theta))^2$$

How does it compare the to asymptotic variance of the ML estimator?

Question 2: Simultaneous Equations

Consider the following three-equation structural model:

$$y_1 = \gamma_{12} y_2 + \delta_{11} z_1 + \delta_{12} z_2 + u_1$$
$$y_2 = \gamma_{22} y_1 + \gamma_{23} y_3 + \delta_{21} z_1 + u_2$$
$$y_3 = \delta_{31} z_1 + \delta_{32} z_2 + \delta_{33} z_3 + u_3$$

where $z_1 = 1$ (to allow for an intercept) and the subscript, i, is suppressed.

(a) Derive the reduced-form system of equations for this model.

(b) Determine which of these equations is identified. First consider the order condition, and then the rank condition.

(c) Suggest a 2SLS procedure for estimating the first equation. Write out the equations for the first and second stage. Be clear on which variables are included in each stage.

(d) Explain in detail how you would estimate this system of equations by 3SLS. (Describe each of the 3 "stages", how you would estimate each stage, and write the final formula for $\hat{\gamma}_{3SLS}$.)

Question 3: Panel Data and Fixed Effects

Consider a model for new capital investment in the manufacturing industry where the cross-section observations are at the county level and there are T years of data for each country:

$$log(invest_{it}) = \delta_0 + \delta_1 tax_{it} + \delta_2 disaster_{it} + \delta_3 urban_i + c_i + u_{it}$$

The variable tax_{it} is a measure of the marginal tax rate on capital in the county, $disaster_{it}$ is a dummy variable equal to one if there was a significant natural disaster in county i in year t, and $urban_i$ is a dummy variable indicating whether the county is predominantly urban (vs. rural).

(a) What kinds of variables are captured in c_i?

(b) What sign does economic reasoning suggest for δ_1?

(c) Describe in detail how you would estimate this model. Which estimator would you choose? Why? Be clear about the relevant assumptions you are making.

(d) Which of the parameters are identified in this model?

(e) Now suppose that $T = 2$. Show that the Fixed Effects and First-Difference estimators are numerically identical.

Question 4: Stata Application

For this problem you will need to turn in a .do file **with comments** and a .log file that shows your regression output.

The data in wagepan.dta are from Vella and Verbeek (1998) for 545 men who worked every year from 1980 to 1987. Consider the wage equation:

$$log(wage_{it}) = \theta_t + \beta_1 educ_i + \beta_2 black_i + \beta_3 hisp_i + \beta_4 exper_{it} + \beta_5 exper_{it}^2 + \beta_6 married_{it} + \beta_7 union_{it} + c_i + u_{it}$$

(a) Estimate this equation by pooled OLS and report the results. Are the usual OLS standard errors reliable? Explain. Compute more appropriate standard errors.

(b) Estimate the equation by Random Effects. Compare your estimates with the pooled OLS estimates.

(c) Now estimate the equation by Fixed Effects. Why is $exper_{it}$ redundant in the model even though it changes over time? What happens to the marriage and union premiums as compared with the RE estimates?

(d) Add the interaction terms: $black_i * union_{it}$ and $hisp_i * union_{it}$. Do the union wage premiums differ by race? Obtain the usual FE statistics and those fully robust to heteroskedasticity and serial correlation.