Cornell University
Department of Economics

Econ 620 – Spring 2008
Instructor: Professor Kiefer
TA: Jae Ho Yun

Suggested Solutions for Problem Set #6

1. (Logit Model, NLS, MLE)

   (a) Recall the the following normal equation for NLS (For details, see the lecture note 18).

   $$F(\widehat{\theta})'(y - f(\widehat{\theta})) = 0$$

   Apply the above to our case. Then, we have the following.

   $$\begin{aligned} -2\sum \frac{1}{(1+\exp(-\alpha-\beta x_i))^2}\exp(-\alpha-\beta x_i)(y-\exp(-\alpha-\beta x_i)) &= 0, \\ -2\sum \frac{1}{(1+\exp(-\alpha-\beta x_i))^2} x_i\exp(-\alpha-\beta x_i)(y-\exp(-\alpha-\beta x_i)) &= 0 \end{aligned}$$

   NL estimator $\widehat{\alpha}_{NL}, \widehat{\beta}_{NL}$ solves the above system equations.

   (b) Conditional variance of d_i is $\Pr(d_i = 1 \mid x_i)[1 - \Pr(d_i = 1 \mid x_i)]$. Note that this comes from the properties of binomial random variables. Therefore, Exact formula for Conditional Variance is;

   $$\begin{aligned} Var(d_i \mid x_i) &= \frac{1}{1+\exp(-\alpha-\beta x_i)} \cdot \left(1 - \frac{1}{1+\exp(-\alpha-\beta x_i)}\right) \\ &= \frac{1}{1+\exp(-\alpha-\beta x_i)} \cdot \frac{\exp(-\alpha-\beta x_i)}{1+\exp(-\alpha-\beta x_i)} \end{aligned}$$

   (c) The Second-round estimator is better. This takes into account the information about the variance structure of the error terms. This argument is in line with the fact that GLS is better than OLS when heteroskedasticity is present.

   (d) First, we have to figure out the likelihood function of individual obaservations.

   $$P(d_i \mid x_i) = \left(\frac{1}{1+\exp(-\alpha-\beta x_i)}\right)^{d_i} \cdot \left(\frac{\exp(-\alpha-\beta x_i)}{1+\exp(-\alpha-\beta x_i)}\right)^{1-d_i}$$

Then, loglikelihood function for individual observation is;

$$\ln P(d_i \mid x_i) = d_i \ln\left(\frac{1}{1+\exp(-\alpha-\beta x_i)}\right) + (1-d_i)\ln\left(\frac{\exp(-\alpha-\beta x_i)}{1+\exp(-\alpha-\beta x_i)}\right)$$

The loglikelihood function is;

$$l(\alpha,\beta) = \sum d_i \ln\left(\frac{1}{1+\exp(-\alpha-\beta x_i)}\right) + \sum(1-d_i)\ln\left(\frac{\exp(-\alpha-\beta x_i)}{1+\exp(-\alpha-\beta x_i)}\right)$$

We can obtain the ML estimator from the first order conditions of the loglikelihood function.

$$\begin{aligned}
\frac{\partial l(\alpha,\beta)}{\partial \alpha} &= \sum d_i\left(1-\frac{1}{1+\exp(-\alpha-\beta x_i)}\right) - \sum(1-d_i)\left(\frac{1}{1+\exp(-\alpha-\beta x_i)}\right) \\
&= 0, \\
\frac{\partial l(\alpha,\beta)}{\partial \beta} &= \sum d_i\left(1-\frac{1}{1+\exp(-\alpha-\beta x_i)}\right) x_i \\
&\quad -\sum(1-d_i)\left(\frac{1}{1+\exp(-\alpha-\beta x_i)}\right) x_i \\
&= 0
\end{aligned}$$

ML estimators, $\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}$ solves the above.

Tips for obtaining FOC in the logit model:

Suppose that $F_i(x_i\beta) = \frac{1}{1+\exp(-x_i\beta)}$ (x_i can be a vector in this tip).

Then, $\exp(-x_i\beta) = \frac{1}{F_i} - 1$.

Also, $F_i' = \left(\frac{1}{1+\exp(-x_i\beta)}\right)^2 \exp(-x_i\beta)\cdot x_i = F_i^2(\frac{1}{F_i}-1)x_i = F_i(1-F_i)x_i$

Using the above tip, we can easily get the FOC.

(Using the similar method, the hessian can be obtained by the formula that $H = -\sum F_i(1-F_i)x_i x_i'$)

2. (Measurement error problem, NLS, IV, GMM)

   (a) Note that t_i is not observable, so we have to use t^* as a proxy. Then, we can run Nonlinear Least Squares (NLS).

   However, there may be a "endogeneity" problem, that is, t^* and error term might be correlated. If endogeneity is present, we cannot obtain consistent estimator from Nonlinear Least Squares.

   To avoid this problem, we're going to use instrumental variable technique (IV). Solve $z'(d-Ed)=0$, where z is $N\times K$ instrumental variable vector, $d - Ed$ is $N \times 1$ vector (you may understand how it is deifned). If this is overidentified case, we cannot have a solution. So let's apply "GLS" idea, that is, solve the following.

$$\underset{\beta,\delta}{Arg\min}(d - Ed)'z\ Var(z'(d - Ed))^{-1}\ z'(d - Ed)$$

But, what is $Var(z'(d - Ed))^{-1}$? $Var(z'(d - Ed)$ have the following form.

$$Var(z'(d - Ed)) = z' \begin{pmatrix} F_1(1 - F_1) & 0 & \cdots & 0 \\ 0 & F_2(1 - F_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_N(1 - F_N) \end{pmatrix} z,$$

where $F_i = F(x_i'\beta + \delta\widehat{t_i^*})$. Hence, in order to evaluate the above consistently, we need consistent estimator for β and δ. I suggest 2-stage method, which is analogous to feasible 3SLS. In the first stage, solve the following:

$$\underset{\beta,\delta}{Arg\min}(d - Ed)'z\ z'(d - Ed)$$

where the weighting matrix is just an identity matrix. Although it may not be efficient, we can obtain a consistent estimator. Let us call it $(\widetilde{\beta}, \widetilde{t})$. Then evaluate $Var(z'(d - Ed)$ at this consistent estimator. Let us call it $\widetilde{\Omega}$. As a second stage, solve the following:

$$\underset{\beta,\delta}{Arg\min}(d - Ed)'z\ \widetilde{\Omega}^{-1}\ z'(d - Ed) = 0$$

The resulting estimator is more efficient than that from the first stage.

(b) Note that $\widehat{\delta}_{IV}$ is consistent in any case, so long as z is exogenous. However, $\widehat{\delta}_{NLS}$ is consistent only when endogeneity is not present. Therefore, we can test measurement error problem by comparing $\widehat{\delta}_{IV}$ and $\widehat{\delta}_{NLS}$. If they are too different, we can say that there is a serious measurement error problem (endogeneity problem).

3. (Logit, MLE)

(a) Since we have N/3 iid Bernoulli observations, say $y_1, \ldots, y_{N/3}$, MLE $\widehat{F}_1 = \sum_{\{x_i=1\}} y_i/(N/3)$. Therefore, we have

$$\begin{aligned} E(\widehat{F}_1) &= \frac{\sum_{\{x_i=1\}} E(y_i)}{N/3} = \frac{\sum_{\{x_i=1\}} F_1}{N/3} = F_1, \\ Var(\widehat{F}_1) &= \frac{\sum_{\{x_i=1\}} Var(y_i)}{(N/3)^2} = \frac{F_1(1 - F_1)}{N/3}. \end{aligned}$$

(b) Log likelihood function is as follows:

$$l(\beta) = \sum \left[y_i \ln \left(\frac{1}{1+\exp(-x_i\beta)} \right) + (1-y_i) \ln \left(\frac{\exp(-x_i\beta)}{1+\exp(-x_i\beta)} \right) \right].$$

MLE is the maximizer of this likelihood function. For the asymptotic variance, we calculate (expected) hessian matrix, H. If we define $\Lambda_i = \{1/(1+\exp(-x_i\beta))\}$, we have $H = -\sum \Lambda_i(1-\Lambda_i)x_i x_i'$. Therefore, we have

$$\begin{aligned} \sqrt{N}(\widehat{\beta}-\beta) &\xrightarrow{d} N\left(0, \left[p\lim \frac{1}{N}\sum \Lambda_i(1-\Lambda_i)x_i x_i'\right]^{-1}\right) \\ &= N\left(0, \left[\frac{1}{3}F_1(1-F_1) + \frac{4}{3}F_2(1-F_2)\right]^{-1}\right). \end{aligned}$$

(c) Since $\widehat{F}_1 = 1/(1+\exp(-\widehat{\beta}))$, from the delta method and the fact that $F_1'(\beta) = F_1(1-F_1)$, we have

$$\sqrt{N}(\widehat{F}_1 - F_1) \xrightarrow{d} N\left(0, F_1^2(1-F_1)^2\left[\frac{1}{3}F_1(1-F_1) + \frac{4}{3}F_2(1-F_2)\right]^{-1}\right).$$

(d) The asymptotic variance of $\widehat{F}_1$ in (c) is smaller than or equal to the variance from (a). But (c) requires more computation and the logistic assumption should be the correct specification.

(e) If the logit specification is correct, estimators should converge to the true values F_1, F_2 and F_3. Therefore we can compute them with the estimator from (a) which are certainly consistent estimators.

4. (MLE)

(a) Note that $c_i \exp(-y_i(x_i\beta))$ is a conditional density function (likelihood function) given x_i and β and that y_i follows exponential distribution. So, we must have the following.

$$\int_0^\infty c_i \exp(-y(x_i\beta))dy = 1$$

Then, we have $c_i = x_i\beta$.

For a sample of N independent observations (y_i, x_i),

(b) First, construct loglikelihood function.

$$\begin{aligned} p_i(y_i \quad &| \quad x_i, \beta) = x_i\beta \exp(-y_i(x_i\beta)) \\ \ln p_i &= \ln(x_i\beta) - y_i(x_i\beta) \\ l(\beta) &= \sum \ln p_i = \sum \ln(x_i\beta) - \sum y_i(x_i\beta) \end{aligned}$$

Second, let's calculate a score function. This is the First Order Condition of the above loglikelihood function.

$$S = \frac{\partial l(\beta)}{\partial \beta} = \sum \frac{x_i'}{x_i\beta} - \sum y_i x_i' \text{ (This is } k \times 1 \text{ vector)}$$

Let's compute a hessian function.

$$H = \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = -\sum \frac{x_i' x_i}{(x_i\beta)^2} \text{ (This is } k \times k \text{ vector)}$$

(c) We know that $\sqrt{N}(\beta_{ML} - \beta) \overset{Asy.}{\sim} N(0, i_0^{-1})$.
And $i_0 = -p\lim(\frac{H}{N})$.
Therefore, the asymptotic distribution is as follows.

$$\sqrt{N}(\beta_{ML} - \beta) \overset{Asy.}{\sim} N\left(0, p\lim\left(\frac{\sum \frac{x_i' x_i}{(x_i\beta)^2}}{N}\right)^{-1}\right)$$

Now you want to plot some "residuals" to check the specification. Hence, you need a transformation z_i of y_i, given x_i and β, such that the distribution of z does not depend on x and β. Consider the random variable $z_i = 1 - \exp(-y_i(x_i\beta))$.

(d) Note that $1 - \exp(-y_i(x_i\beta))$ is a CDF of exponential distribution, so z_i can take values from 0 to 1. Let's call it F, that is, $z_i = F(y_i)$ given x_i, β.
Then, $P(z_i \leq \tilde{z}) = P(F(y_i) \leq \tilde{z}) = P(F^{-1}(F(y_i)) \leq F^{-1}(\tilde{z}))$ since F is strictly increasing in (0,1).
Now, we have $P(F^{-1}(F(y_i)) \leq F^{-1}(\tilde{z})) = P(y_i \leq F^{-1}(\tilde{z}))$. And we know that y_i follows exponential distribution, so $P(y_i \leq F^{-1}(\tilde{z})) = F(F^{-1}(\tilde{z})) = \tilde{z}$.
The last formula implies that $P(z_i \leq \tilde{z}) = \tilde{z}$ and that z_i follows a uniform distribution whose support is (0,1).
Suppose you calculate z using β_{ML} instead of β and you plot the empirical cdf F-hat(t)= $(\#z_i \leq t)/N$.

(e) If the model is correct and the estimate is good, our empirical cdf should look like real cdf (cdf of uniform distribution). Therefore, the plot should look like 45 degree line in (0, 1).

5. (K-variable Regression)

(a) We have

$$b^* = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i^4}{\sum x_i^2}.$$

So, conditional expectation and conditional variance of b^* are

$$\begin{aligned} E_x(b^*) &= \frac{\sum x_i^4}{\sum x_i^2}, \\ Var_x(b^*) &= 0. \end{aligned}$$

(b) Note that, when x follows standard normal, we have

$$\begin{aligned} E(x^n) &= (n-1)(n-3)\dots 1 \quad \text{when } n \text{ is odd} \\ &= 0 \quad \text{when } n \text{ is even.} \end{aligned}$$

Therefore, we have

$$b = \frac{Exy}{Ex^2} = \frac{Ex^4}{Ex^2} = \frac{3}{1} = 3.$$

(c) Note that

$$t = \frac{\widehat{b}}{S/\sqrt{\sum x_i^2}} \quad \text{and} \quad p\lim \widehat{b} = 3,$$

and standard error is $S/\sqrt{\sum x_i^2}$.

S^2 can be represented as

$$\begin{aligned} S^2 &= \frac{\sum(y_i - \widehat{b}x_i)^2}{N-1} = \frac{\sum y_i^2 - 2\widehat{b}\sum x_i y_i + \widehat{b}^2 \sum x_i^2}{N-1} \\ &= \frac{N}{N-1} \frac{\sum x_i^6 - 2\widehat{b}\sum x_i^4 + \widehat{b}^2 \sum x_i^2}{N}. \end{aligned}$$

Therefore,

$$p\lim S^2 = p\lim \frac{N}{N-1} \cdot p\lim \frac{\sum x_i^6 - 2\widehat{b}\sum x_i^4 + \widehat{b}^2 \sum x_i^2}{N} = 6.$$

So we have $p\lim S = \sqrt{6}$ (Apply the law of large numbers above). Furthermore,

$$\begin{aligned} t &= \frac{3}{\sqrt{6}}\sqrt{\sum x_i^2} = \sqrt{N} \cdot \frac{3}{\sqrt{6}} \cdot \sqrt{\frac{\sum x_i^2}{N}} \\ &\approx \frac{3}{\sqrt{6}}\sqrt{N}. \end{aligned}$$

and standard error is approximately $\sqrt{6/N}$. Next, R^2 can be approximated as follows:

$$R^2 = \frac{\widehat{b}^2 \sum x_i^2}{\sum y_i^2} = \frac{\widehat{b}^2 \sum x_i^2/N}{\sum y_i^2/N} \xrightarrow{p} \frac{9}{15} = \frac{3}{5}.$$

(d) Now we estimate the model, $y_i = b_1 x_i + b_2 x_i^2 + \varepsilon_i$. Note that x_i and x_i^2 are orthogonal since $E(x_i^3) = 0$. Hence, we have

$$\widehat{b}_2 = \frac{\sum x_i^2 y_i}{\sum x_i^4} = \frac{\sum x_i^5/\sqrt{N}}{\sum x_i^4/\sqrt{N}} \xrightarrow{p} 0.$$

For a plot of actual vs. fitted values, draw $y = x^3$ and $y = 3x$ on the same xy graph.

6. (Gauss-Markov Thorem)

(a) The OLS esimator, $\widehat{\beta}_{OLS}$ is $(X'X)^{-1}X'y$. Clearly, it is a linear estimator, where $c' = (X'X)^{-1}X'$.

(b) There can be a lot of examples for this. Suppose that $c' = 2(X'X)^{-1}X'$. Then, $T(y) = 2(X'X)^{-1}X'y$. It is easy to show that $E[T(y)] = 2\beta \neq \beta$ unless $\beta = 0$.

(c) MSE of the OLS estimator

$$\begin{aligned}
m(T(y),\beta)_{OLS} &= E[(T(y)-\beta)(T(y)-\beta)'] \\
&= E[(X'X)^{-1}X'y-\beta)(X'X)^{-1}X'y-\beta)'] \\
&= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

(d) MSE of an arbitrary linear estimator

$$\begin{aligned}
m(T(y),\beta) &= E[(T(y)-\beta)(T(y)-\beta)'] \\
&= E[(c'y-\beta)(c'y-\beta)'] \\
&= E[c'yy'c-\beta y'c-c'y\beta'+\beta\beta'] \\
&= E[c'(X\beta+\epsilon)(X\beta+\epsilon)'c-\beta(X\beta+\epsilon)'c-c'(X\beta+\epsilon)\beta'+\beta\beta'] \\
&= E[c'X\beta\beta'X'c+c'X\beta\epsilon'c+c'\epsilon\beta'X'c+c'\epsilon\epsilon'c-\beta\beta'X'c-\beta\epsilon'c \\
&\quad -c'X\beta\beta'-c'\epsilon\beta'+\beta\beta'] \\
&= c'X\beta\beta'X'c+c'X\beta E(\epsilon')c+c'E(\epsilon)\beta'X'c+c'E(\epsilon\epsilon')c-\beta\beta'X'c \\
&\quad -\beta E(\epsilon')c-c'X\beta\beta'-c'E(\epsilon)\beta'+\beta\beta' \\
&= c'X\beta\beta'X'c+\sigma^2c'c-\beta\beta'X'c-c'X\beta\beta'+\beta\beta'
\end{aligned}$$

(e) Yes, this is true. In order for boundedness to be guaranteed, unbiasedness should be satisfied. Then, we can apply the Gauss Markov Theorem.

(f) Consider the MSE of an arbitrary linear estimator, $m(T(y),\beta) = c'X\beta\beta'X'c + \sigma^2c'c - \beta\beta'X'c - c'X\beta\beta' + \beta\beta'$. This is a quadratic function of β. This function can take positive or negative infinity since $\beta \in R^k$.

Therefore, to obtain bounded MSE, there should be no term including β in the above formula. The only way of excluding β is to have that $c'X = I$. Then, only $\sigma^2 c'c$ will be left. This is nothing but a unbiasedness condition of linear estimator. Hence, we can apply the Gauss Markov Theorem; The OLS estimator has minimum variance.

7. (NLS, IV, GMM)

(a) Noting that the homogeneous ODE, $y' - \beta y = 0$ has the solution $y_h = Ce^{bx}$, we can find a particular solution $y_p = \gamma + \alpha x$. Hence the solution is $y = \gamma + \alpha x + Ce^{bx}$. From the initial condition we have $C = 1$. The model would be $y_i = \gamma + \alpha x_i + e^{\beta x_i} + \varepsilon_i$ with $\varepsilon_i \sim iid(0, \sigma^2)$.

(b) The normal equation for a nonlinear model $y = F(x; \gamma, \alpha, \beta)$ is $F'(\gamma, \alpha, \beta)(y - F(\gamma, \alpha, \beta)) = 0$. Since $F' = (1, x, xe^{\beta x})^T$, the normal equation is

$$\begin{aligned}
\sum \left[y_i - (\gamma + \alpha x_i + e^{\beta x_i})\right] &= 0, \\
\sum x_i \left[y_i - (\gamma + \alpha x_i + e^{\beta x_i})\right] &= 0, \\
\sum x_i e^{\beta x_i} \left[y_i - (\gamma + \alpha x_i + e^{\beta x_i})\right] &= 0.
\end{aligned}$$

(c) Given a value of β, it becomes a standard LS problem. We can solve for γ and α in terms of β then the problem is one dimensioanl optimization with respect to β only.

(d) "normal equation" become $z'(y - Ey) = 0$. In terms of GMM, you have k moment conditions. In this overidentified case, these equation cannot be solved directly and GLS must be used on them. Thus minimize $(y - Ey)'z(z'z)^{-1}z'(y - Ey)$ with FOC (normal equations) $F'z(z'z)^{-1}z'(y - Ey) = 0$.

(e) We can compute the estimator from (b) and (d). If x is endogenous, they should be different significantly.