**ECONOMETRICS II (ECO 2401S)**

**University of Toronto. Department of Economics. Spring 2011**
**Instructor: Victor Aguirregabiria**

**SOLUTION TO FINAL EXAM**
**Friday, April 15, 2011. From 9:00-12:00 (3 hours)**

---

**INSTRUCTIONS: The exam consists of three Problems. You have to answer all the questions. No study aids, including calculators, are allowed.**
**TOTAL MARKS = 100**

---

**PROBLEM 1 (35 points): Consider the Dynamic Panel Data model:**

$$Y_{it} = \beta\, Y_{it-1} + \alpha_i + u_{it}$$

**For simplicity, suppose that the number of periods in the dataset is $T = 3$, and that the variable $Y_{it}$ is in deviations with respect to its time–period-specifc mean such that, for every period $t$, $E(Y_{it}) = E(\alpha_i) = E(u_{it}) = 0$. We also assume that: $u_{it}$ is iid over time and individuals with variance $\sigma_u^2$; $\alpha_i$ is iid over individuals with variance $\sigma_\alpha^2$; and $|\beta| < 1$.**

**Question 1.1. (5 points):**
**(a) Define Anderson-Hsiao estimator and write the moment condition(s) for this estimator.**
ANSWER: First, we transform the model in first differences. Given that the dataset contains only 3 periods, we can construct the equation in first differences only for the last period, $t = 3$:

$$\Delta Y_{i3} = \beta\, \Delta Y_{i2} + \Delta u_{i3}$$

Anderson-Hsiao estimator is an IV estimator of the equation in first-differences where we instrument the regressor $\Delta Y_{i2}$ using $Y_{i1}$ (in general, with more than 3 periods, we instrument $\Delta Y_{it-1}$ using $Y_{it-2}$). The moment condition that defines the Anderson-Hsiao estimator is $E(Y_{i1}\ \Delta u_{i3}) = 0$. Or, taking into account that the model implies that $\Delta u_{i3} = \Delta Y_{i3} - \beta\ \Delta Y_{i2}$:

$$E\left(\ Y_{i1}\ \left[\Delta Y_{i3} - \beta\ \Delta Y_{i2}\right]\right) = 0$$

**(b) Obtain the closed form expression of the estimator in terms of the data.**
ANSWER: Anderson-Hsiao estimator is the value of $\beta$ that solves the sample counterpart of the moment condition $E\left(\ Y_{i1}\ [\Delta Y_{i3} - \beta\ \Delta Y_{i2}]\right) = 0$. That is, Anderson-Hsiao estimator is the value $\hat{\beta}_{AH}$ that solves the equation:

$$\frac{1}{N}\sum_{i=1}^{N} y_{i1}\left[\Delta y_{i3} - \hat{\beta}_{AH}\ \Delta y_{i2}\right] = 0$$

And solving for $\hat{\beta}_{AH}$, we get:

$$\hat{\beta}_{AH} = \left(\sum_{i=1}^{N} y_{i1}\ \Delta y_{i2}\right)^{-1} \left(\sum_{i=1}^{N} y_{i1}\ \Delta y_{i3}\right)$$

**Question 1.2. (10 points):**
**(a) Derive the expression of the Asymptotic Variance of the Anderson-Hsiao estimator in terms of the parameters $\beta$, $\sigma_u^2$, and $\sigma_\alpha^2$. Explain all your derivations.**
ANSWER: Solving the expression $\Delta y_{i3} = \beta\ \Delta y_{i2} + \Delta u_{i3}$ into the formula of the estimator $\hat{\beta}_{AH}$, we have that $\hat{\beta}_{AH} = \beta + \left(\sum_{i=1}^{N} y_{i1}\ \Delta y_{i2}\right)^{-1} \left(\sum_{i=1}^{N} y_{i1}\ \Delta u_{i3}\right)$, and

$$\sqrt{N}\left(\hat{\beta}_{AH} - \beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N} y_{i1}\ \Delta y_{i2}\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i1}\ \Delta u_{i3}\right)$$

By the LLN, $\frac{1}{N}\sum_{i=1}^{N} y_{i1}\ \Delta y_{i2} \rightarrow_p E\left(y_{i1}\ \Delta y_{i2}\right)$. By the CLT, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i1}\ \Delta u_{i3} \rightarrow_d N(0, Var(y_{i1}\ \Delta u_{i3}))$. And by combining these two results with Mann-Wald Theorem, we have that:

$$\sqrt{N}\left(\hat{\beta}_{AH} - \beta\right) \rightarrow_d N\left(0\ ,\ \frac{Var(y_{i1}\Delta u_{i3})}{E\left(y_{i1}\ \Delta y_{i2}\right)^2}\right)$$

Therefore, the asymptotic variance of $\hat{\beta}_{AH}$ is $\dfrac{1}{N}\frac{Var(y_{i1}\Delta u_{i3})}{E(y_{i1}\ \Delta y_{i2})^2}$.

Now, I obtain this variance in terms of the parameters $\beta$, $\sigma_u^2$, and $\sigma_\alpha^2$. First, note that $Var(y_{i1}\Delta u_{i3}) = E(y_{i1}^2 \Delta u_{i3}^2)$, and by the law of iterative expectations, $E(y_{i1}^2 \Delta u_{i3}^2) = E(y_{i1}^2)\ E(\Delta u_{i3}^2)$. Given that $E(\Delta u_{i3}^2) = 2\sigma_u^2$, and $E\left(y_{i1}^2\right) = \dfrac{\sigma_\alpha^2}{(1-\beta)^2} + \dfrac{\sigma_u^2}{1-\beta^2}$, we have that $Var(y_{i1}\Delta u_{i3}) = 2\sigma_u^2 [\dfrac{\sigma_\alpha^2}{(1-\beta)^2} + \dfrac{\sigma_u^2}{1-\beta^2}]$. Next, we have that:

$$\begin{aligned} E\left(y_{i1}\ \Delta y_{i2}\right) &= E\left([u_{i1} + \beta u_{i0} + ...]\ \ [u_{i2} + (\beta-1)u_{i1} + \beta(\beta-1)u_{i0} + ...]\right) \\ &= \frac{(\beta-1)\ \sigma_u^2}{1-\beta^2} = \frac{-\sigma_u^2}{1+\beta} \end{aligned}$$

Then,

$$\begin{aligned} AVar(\hat{\beta}_{AH}) &= \frac{1}{N}\ \frac{Var(y_{i1}\Delta u_{i3})}{E\left(y_{i1}\ \Delta y_{i2}\right)^2} = \frac{1}{N}\ \frac{2\sigma_u^2\ \left[\dfrac{\sigma_\alpha^2}{(1-\beta)^2} + \dfrac{\sigma_u^2}{1-\beta^2}\right]}{\left(\dfrac{-\sigma_u^2}{1+\beta}\right)^2} \\ &= \frac{2}{N}\ \left(\frac{1+\beta}{1-\beta}\right)\left[1 + \left(\frac{1+\beta}{1-\beta}\right)\frac{\sigma_\alpha^2}{\sigma_u^2}\right] \end{aligned}$$

**(b) Based on this expression, discuss the potential weak instrument(s) problem for this estimator.**
ANSWER: We can see that as $\beta$ approaches to 1 the variance of the AH estimator goes to infinity. This is due to a weak instruments problem. As $\beta$ approaches to 1, the proportion of the variance

of $\Delta y_{i2}$ explained by the instrument $y_{i1}$ (i.e., the R-square in the auxiliary regression of $\Delta y_{i2}$ on $y_{i1}$) goes to zero.

**Question 1.3. (10 points): Suppose now that the equation $Y_{it} = \beta\ Y_{it-1} + \alpha_i + u_{it}$ holds not only during the sample periods but also for many periods before. Therefore, for every individual $i$, the random variable $Y_{it}$ follows a stationary AR(1) process before and during the sample period.**

**(a) Show that, under this stationarity assumption, there is an additional moment restriction that can be use to estimate the parameter $\beta$. Write that moment condition and explain why it holds.**

ANSWER: Under this assumption, we have that for any period $t$, $Y_{it} = \frac{\alpha_i}{1-\beta} + u_{it} + \beta u_{it-1} + \beta^2 u_{it-2} + ...$, and $\Delta Y_{it}$ does not depend on $\alpha_i$. Consider the equation in levels at period $t = 3$, $Y_{i3} = \beta\ Y_{i2} + \alpha_i + u_{i3}$. In this equation, under the stationarity condition, $\Delta Y_{i2}$ is a valid instrument for $Y_{i2}$ because: (1) $Y_{i2}$ and $\Delta Y_{i2}$ are correlated; and (2) $E(\Delta Y_{i2}\ [\alpha_i + u_{i3}]) = 0$. The moment condition is:

$$E\left(\Delta Y_{i2}\ \ [Y_{i3} - \beta\ Y_{i2}]\right) = 0$$

Interpretation: Since $Y_{i2}$ is predetermined with respect to the iid transitory shock $u_{i3}$, the only reason why $Y_{i2}$ is an endogenous regressor is that it is correlated with the individual effect $\alpha_i$. But under the stationarity assumption, the first difference $\Delta Y_{i2}$ is 'purged' of the part of $Y_{i2}$ that depends on $\alpha_i$. This is why $\Delta Y_{i2}$ is a valid instrument for $Y_{i2}$.

**(b) Obtain the closed form expression of the IV estimator based on that moment condition.**

ANSWER: $\hat{\beta}_{BB}$ is the value of $\beta$ that solves for the sample counterpart of the moment condition $E\left(\Delta Y_{i2}\ \ [Y_{i3} - \beta\ Y_{i2}]\right) = 0$. That is,

$$\frac{1}{N}\sum_{i=1}^{N} \Delta y_{i2} \left[y_{i3} - \hat{\beta}_{BB}\ y_{i2}\right] = 0$$

And solving for $\hat{\beta}_{BB}$, we get:

$$\hat{\beta}_{BB} = \left(\sum_{i=1}^{N} \Delta y_{i2}\ y_{i2}\right)^{-1} \left(\sum_{i=1}^{N} \Delta y_{i2}\ y_{i3}\right)$$

**Question 1.4. (10 points): Let $\hat{\beta}_{BB}$ be the IV estimator in Question 1.3.**

**(a) Derive the expression of the Asymptotic Variance of this BB estimator in terms of the parameters $\beta$, $\sigma_u^2$, and $\sigma_\alpha^2$. Explain all your derivations.**

ANSWER: Solving the expression $y_{i3} = \beta\ y_{i2} + u_{i3}$ into the formula of the estimator $\hat{\beta}_{BB}$, we have that $\hat{\beta}_{BB} = \beta + \left(\sum_{i=1}^{N} \Delta y_{i2}\ y_{i2}\right)^{-1} \left(\sum_{i=1}^{N} \Delta y_{i2}\ u_{i3}\right)$.

$$\sqrt{N}\left(\hat{\beta}_{BB} - \beta\right) = \left(\frac{1}{N}\sum_{i=1}^{N} \Delta y_{i2}\ y_{i2}\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta y_{i2}\ u_{i3}\right)$$

By the LLN, $\frac{1}{N}\sum_{i=1}^{N} \Delta y_{i2}\ y_{i2} \to_p E\left(\Delta y_{i2}\ y_{i2}\right)$. By the CLT, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Delta y_{i2}\ u_{i3} \to_d N(0, Var(\Delta y_{i2}\ u_{i3}))$. And by combining these two results with Mann-Wald Theorem, we have that:

$$\sqrt{N}\left(\hat{\beta}_{BB} - \beta\right) \to_d N\left(0\ ,\ \frac{Var(\Delta y_{i2}\ u_{i3})}{E\left(\Delta y_{i2}\ y_{i2}\right)^2}\right)$$

Therefore, the asymptotic variance of $\hat{\beta}_{BB}$ is $\frac{1}{N}\frac{Var(\Delta y_{i2}\ u_{i3})}{E(\Delta y_{i2}\ y_{i2})^2}$.

Now, I obtain this variance in terms of the parameters $\beta$, $\sigma_u^2$, and $\sigma_\alpha^2$. First, note that $Var(\Delta y_{i2}\ u_{i3}) = E(\Delta y_{i2}^2\ u_{i3}^2)$, and by the law of iterative expectations, $E(\Delta y_{i2}^2\ u_{i3}^2) = E(\Delta y_{i2}^2)\ E(u_{i3}^2)$. Given that $E(u_{i3}^2) = \sigma_u^2$, and $E(\Delta y_{i2}^2) = \dfrac{2\sigma_u^2}{1+\beta}$, we have that $Var(\Delta y_{i2}\ u_{i3}) = \dfrac{2\sigma_u^4}{1+\beta}$. Next, we have that:

$$
\begin{aligned}
E\left(\Delta y_{i2}\ y_{i2}\right) &= E\left([u_{i2} + (\beta-1)u_{i1} + (\beta-1)\beta u_{i0} + ...]\ \ [u_{i2} + \beta u_{i1} + \beta^2 u_{i0} + ...]\right) \\
&= \sigma_u^2\ \left[1 + (\beta-1)\ \frac{1}{1-\beta^2}\right] = \frac{\sigma_u^2}{1+\beta}
\end{aligned}
$$

Then,

$$
\begin{aligned}
AVar(\hat{\beta}_{BB}) &= \frac{1}{N}\ \frac{Var(\Delta y_{i2}\ u_{i3})}{E\left(\Delta y_{i2}\ y_{i2}\right)^2} = \frac{1}{N}\ \frac{\dfrac{2\sigma_u^4}{1+\beta}}{\left(\dfrac{\sigma_u^2}{1+\beta}\right)^2} \\
&= \frac{2}{N}\ (1+\beta)
\end{aligned}
$$

**(b) Based on this expression, discuss the potential weak instrument(s) problem for this estimator.**

ANSWER: The variance of the BB estimator also increases with $\beta$. However, in contrast to the AH estimator, this variance does not goes to infinite as $\beta$ approaches 1. Even if $\beta = 1$, the variance of the BB estimator is finite, i.e., $4/N$.

**PROBLEM 2 (40 points): Suppose that the admissions office of a university has hired you to estimate the effect of financial help (student scholarships) on students' decisions to attend that university. Suppose that you have a dataset with information from the $N$ students who were accepted in the university last year. For every student $i$ in the sample, you observe the following variables: parental income, $M_i$; an index of high school grades and academic performance, $G_i$; the dollar amount of financial support offered by the university to the student, $S_i$; and a binary variable $Y_i \in \{0, 1\}$ such that $Y_i = 1$ iff student $i$ chose to attend this university.**

**Question 2.1. (10 points): Consider the Binary Choice Model:**

$$Y_i = 1\{\ X_i\beta - u_i \geq 0\}$$

**where $X_i = (1,\ \ln M_i,\ G_i,\ S_i)$, $\beta = (\beta_0,\ \beta_1,\ \beta_2,\ \beta_3)$, and the error term $u_i$ is assumed iid $N(0, 1)$.**

**(a) Write the log-likelihood function of this model and data.**

ANSWER: Given a random sample $\{y_i, x_i : i = 1, 2, ...n\}$, the log likelihood function is:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{N} y_i \ \ln \Pr(Y_i = 1 \mid X_i = x_i) + (1 - y_i)\ \ln \Pr(Y_i = 1 \mid X_i = x_i) \\ &= \sum_{i=1}^{N} y_i \ \ln \Pr(u_i \leq x_i\beta | x_i) + (1 - y_i)\ \ln \Pr(u_i > x_i\beta | x_i) \end{aligned}$$

Given that $u_i$ is iid $N(0, 1)$ and independent of $X_i$, we have that $\Pr(u_i \leq x_i\beta|x_i) = \Phi(x_i\beta)$ and $\Pr(u_i > x_i\beta|x_i) = 1 - \Phi(x_i\beta)$, where $\Phi$ is the CDF of the standard normal. Therefore,

$$l(\beta) = \sum_{i=1}^{N} y_i \ \ln \Phi(x_i\beta) + (1 - y_i)\ \ln(1 - \Phi(x_i\beta))$$

**(b) Describe an algorithm to compute the Maximum Likelihood Estimator of $\beta$.**

ANSWER: The MLE is the value of $\beta$ that maximizes the function $l(\beta)$. This log-likelihood function is globally concave in $\beta$. Therefore, a gradient method such as Newton-Raphson or BHHH always converges to the MLE regardless the initial value that we use to initialize the algorithm. Here I describe the Newton-Raphson algorithm. Newton-Raphson is an iterative algorithm that starts with an arbitrary initial guess, $\hat{\beta}^{(0)}$, and generates a sequence of vectors $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$, ... $\hat{\beta}^{(K)}$, ... using the following iterative equation: at iteration $K \geq 1$,

$$\hat{\beta}^{(K)} = \hat{\beta}^{(K-1)} - \left( \frac{\partial^2 l(\hat{\beta}^{(K-1)})}{\partial\beta\partial\beta'} \right)^{-1} \frac{\partial l(\hat{\beta}^{(K-1)})}{\partial\beta}$$

The method iterates until the sequence of $\hat{\beta}'s$ satisfies a convergence criterion such as, say, $||\hat{\beta}^{(K)} - \hat{\beta}^{(K-1)}|| < 10^{-6}$.

**(c) Define the Conditional Choice Probability (CCP) function $P(x) \equiv \Pr(Y_i = 1 \mid X_i = x)$. Based on this model, write the expression of the CCP function and of the effect on $P(x)$ of a marginal change in the amount of subsidy $S_i$.**

ANSWER: As explained in the answer to Question 2.1(a), $\Pr(Y_i = 1 \mid X_i = x) = \Pr(u_i \leq x\beta | x) = \Phi(x\beta)$. Therefore, $P(x) = \Phi(x\beta)$. The effect on $P(x)$ of a marginal change in the amount of subsidy $S_i$ is:

$$\frac{\partial P(x)}{\partial S} = \frac{\partial \Phi(x\beta)}{\partial S} = \beta_3 \ \phi(x\beta)$$

where $\phi(.)$ is the PDF of the standard normal.

**Question 2.2. (10 points): Suppose that the admission office is interested in maximizing an objective function that depends on the expected value of: (1) the size of the incoming class (i.e., $\sum_{i=1}^{N} Y_i$); (2) the total grades of the incoming class (i.e., $\sum_{i=1}^{N} Y_i G_i$); and (3) the total cost of the scholarships (i.e., $\sum_{i=1}^{N} Y_i S_i$). For instance, suppose that the admission office has the following objective function:**

$$W = \ln\left( \ \sum_{i=1}^{N} E\left(Y_i | X_i\right) \ \right) + \ln\left( \ \sum_{i=1}^{N} G_i \, E\left(Y_i | X_i\right) \ \right) - \ln\left( \ \sum_{i=1}^{N} S_i \, E\left(Y_i | X_i\right) \ \right)$$

**(a) Based on our model, derive the expression for the marginal change in $W$ associated with a marginal increase in the scholarship amount $S_i$ for every accepted student.**
ANSWER: Given that $E(Y_i | X_i = x_i) = P(x) = \Phi(x_i\beta)$, we have that:

$$W = \ln\left(\sum_{i=1}^{N} \Phi(x_i\beta)\right) + \ln\left(\sum_{i=1}^{N} G_i \, \Phi(x_i\beta)\right) - \ln\left(\sum_{i=1}^{N} S_i \, \Phi(x_i\beta)\right)$$

Suppose that we increase the scholarship amount of every accepted student by $\delta$. Let $W(\delta)$ be the value of the objective function. We want to obtain $\frac{\partial W(\delta=0)}{\partial \delta}$. Using the previous expression for $W$, we have that:

$$\begin{aligned} \frac{\partial W(0)}{\partial \delta} &= \frac{\partial \ln\left(\sum_{i=1}^{N} \Phi(x_i\beta)\right)}{\partial \delta} + \frac{\partial \ln\left(\sum_{i=1}^{N} G_i \, \Phi(x_i\beta)\right)}{\partial \delta} - \frac{\partial \ln\left(\sum_{i=1}^{N} S_i \, \Phi(x_i\beta)\right)}{\partial \delta} \\ &= \frac{\sum_{i=1}^{N} \beta_3 \ \phi(x_i\beta)}{\sum_{i=1}^{N} \Phi(x_i\beta)} + \frac{\sum_{i=1}^{N} \beta_3 \ G_i \ \phi(x_i\beta)}{\sum_{i=1}^{N} G_i \ \Phi(x_i\beta)} - \frac{\sum_{i=1}^{N} \Phi(x_i\beta)}{\sum_{i=1}^{N} S_i \ \Phi(x_i\beta)} - \frac{\sum_{i=1}^{N} \beta_3 \ S_i \ \phi(x_i\beta)}{\sum_{i=1}^{N} S_i \ \Phi(x_i\beta)} \end{aligned}$$

**(b) Based on your MLE in Question 2.1., propose an estimator for this marginal effect.**
ANSWER: The marginal effect $\frac{\partial W(0)}{\partial \delta}$ depends on the vector of parameters $\beta$. An estimator of this marginal effect $\frac{\partial W(0)}{\partial \delta}$ can be obtained by simply replacing $\beta$ by our MLE $\hat{\beta}$.

**Question 2.3. (20 points): The consistency of the estimates in Questions 2.1 and 2.2 depend on our correct specification of the distribution of the error term $u_i$. In contrast, the Least Absolute Deviations (LAD) estimator of $\beta$ in a Binary Choice Model is a consistent estimator as long as the error term $u_i$ is median independent of the regressors in $X_i$.**
**(a) Define the LAD estimator of our Binary Choice Model and show that this estimator is consistent if and only if $median(u_i | X_i) = 0$.**
ANSWER: Define the following sample criterion function for our BCM:

$$Q(\beta) = \sum_{i=1}^{N} \mid y_i - 1\{x_i\beta \geq 0\} \mid$$

where $|y_i - 1\{x_i\beta \geq 0\}|$ is the absolute deviation of $y_i$ with respect to the "predictor" $1\{x_i\beta \geq 0\}$. The LAD estimator of $\beta$ is defined as the value of $\beta$ that minimizes $Q(\beta)$. This LAD estimator is a consistent estimator of $\beta$ iff the "predictor" that we use to construct the absolute deviations is the conditional median of $Y_i$ given $X_i$. That is, for consistency of the LAD estimator, the LAD criterion function should be $\sum_{i=1}^{N} |y_i - \ median(Y_i|X_i = x_i)|$. In our BCM, we have that:

$$median(Y_i|X_i = x_i) = median(1\{x_i\beta - u_i \geq 0\} \mid x_i)$$

For a monotonic function, the median of the function is equal to the function of the median. Then,

$$median(Y_i|X_i = x_i) = 1\{x_i\beta - median(u_i|x_i) \geq 0\} = 1\{x_i\beta \geq 0\}$$

Therefore, the LAD estimator of $\beta$ is consistent iff $median(u_i|x_i) = 0$.

**(b) Obtain the relationship between the LAD criterion function and the Score function:**

$$S(\beta) = \textstyle\sum_{i=1}^{N} y_i \ 1\{x_i\beta \geq 0\} + (1 - y_i) \ 1\{x_i\beta < 0\}$$

ANSWER: Note that the absolute deviation $\mid y_i - 1\{x_i\beta \geq 0\} \mid$ can be written as:

$$\begin{aligned}
\mid y_i - 1\{x_i\beta \geq 0\} \mid &= y_i \ 1\{x_i\beta < 0\}) + (1 - y_i) \ 1\{x_i\beta \geq 0\} \\
&= y_i \ (1 - 1\{x_i\beta \geq 0\}) + (1 - y_i) \ (1 - 1\{x_i\beta < 0\}) \\
&= 1 - y_i \ 1\{x_i\beta \geq 0\} - (1 - y_i) \ 1\{x_i\beta < 0\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
Q(\beta) &= \sum_{i=1}^{N} \mid y_i - 1\{x_i\beta \geq 0\} \mid \\
&= N - \left[\textstyle\sum_{i=1}^{N} y_i \ 1\{x_i\beta \geq 0\} + (1 - y_i) \ 1\{x_i\beta < 0\}\right] \\
&= N - S(\beta)
\end{aligned}$$

**(c) Define the Maximum Score Estimator (MSE). Discuss its limitations.**
ANSWER: The MSE is defined as the value of $\beta$ that maximizes the Score function $S(\beta)$. Given that $Q(\beta) = N - S(\beta)$, it is clear that maximizing the Score function is equivalent to minimizing the sum of the absolute deviations $Q(\beta)$. Therefore, for Binary Choice Models, the LAD and MSE estimators are the same estimator. For Binary Choice Models, this estimator has 4 main limitations: (a) the rate of convergence is lower than root-N; (b) the asymptotic distribution is not normal, and it is not a standard distribution; (c) the sample criterion function is a step-wise function, discontinuous at many points, and with zero gradient at the rest of the points: we cannot use gradient optimization methods to optimize this type of criterion function; and (d) for a finite sample, there is not a unique value of $\beta$ that minimizes the criterion function.

**(d) Define the Smoothed Maximum Score Estimator (SMSE).**

ANSWER: To define and to understand the SMSE, it is convenient to re-write the score function $S(\beta)$.

$$
\begin{aligned}
S(\beta) &= \sum_{i=1}^{N} y_i \; 1\{x_i\beta \geq 0\} + (1-y_i) \; 1\{x_i\beta < 0\} \\
&= \sum_{i=1}^{N} y_i \; 1\{x_i\beta \geq 0\} + (1-y_i) \; (1 - 1\{x_i\beta \geq 0\}) \\
&= \sum_{i=1}^{N} (1-y_i) + \sum_{i=1}^{N} (2y_i - 1) \; 1\{x_i\beta \geq 0\}
\end{aligned}
$$

Therefore, maximizing in $\beta$ the score $S(\beta)$ is equivalent to maximizing in $\beta$ the function $\sum_{i=1}^{N} (2y_i - 1)$ $1\{x_i\beta \geq 0\}$. The SMSE is defined as the value of $\beta$ that maximizes the Smoothed Score function:

$$
SS(\beta) = \sum_{i=1}^{N} (2y_i - 1) \; \Phi\left(\frac{x_i\beta}{b_N}\right)
$$

where $\Phi(.)$ is the CDF of the standard normal (we could use any other smoothed CDF), and $b_N$ is a bandwidth parameter. Note that $SS(\beta)$ is just the result of replacing in $S(\beta)$ the step function $1\{x_i\beta \geq 0\}$ by the smooth function $\Phi\left(\frac{x_i\beta}{b_N}\right)$. For root-N consistency of the SSME, we need that as $N$ goes to infinite, $b_N$ goes to zero but $N \; b_N$ goes to infinity. The SMSE deals with the four problems of the MSE that have been mentioned in the answer to Question 2.3(c).

**(e) Suppose that you have a consistent estimator of $\beta$. Explain how to estimate consistently the distribution of the error term $u_i$.**

ANSWER: Given a consistent estimator $\hat{\beta}$, we can construct the values $z_i \equiv x_i\hat{\beta}$ for every observation $i$ in the sample. The model implies that:

$$
E(Y_i \mid z_i = z) = \Pr(u_i \leq z) = F_u(z)
$$

where $F_u(.)$ is the CDF of $u$. Therefore, we can estimate nonparametrically the CDF $F_u(.)$ by using a nonparametric estimator of the regression function (or conditional mean function) $E(Y_i \mid z_i = z)$. For instance, we can use the following Nadaraya-Watson (Kernel) estimator:

$$
\hat{F}_u(z) = \hat{E}(Y_i \mid z) = \frac{\sum_{i=1}^{N} y_i \; \phi\left(\dfrac{z_i - z}{b_N}\right)}{\sum_{i=1}^{N} \phi\left(\dfrac{z_i - z}{b_N}\right)}
$$

where $\phi(.)$ is the PDF of the standard normal (we could use other kernel function too) and $b_N$ is a bandwidth parameter such that as $N$ goes to infinite, $b_N$ goes to zero but $N \; b_N$ goes to infinity.

**(f) Given consistent estimates of $\beta$ and of the distribution of $u$, explain how to construct an estimator of the marginal effect defined in Question 2.2 above.**

ANSWER: For a general distribution function of $u$, we have that:

$$
\frac{\partial W(0)}{\partial \delta} = \frac{\sum_{i=1}^{N} \beta_3 \; f_u(x_i\beta)}{\sum_{i=1}^{N} F_u(x_i\beta)} + \frac{\sum_{i=1}^{N} \beta_3 \; G_i \; f_u(x_i\beta)}{\sum_{i=1}^{N} G_i \; F_u(x_i\beta)} - \frac{\sum_{i=1}^{N} F_u(x_i\beta)}{\sum_{i=1}^{N} S_i \; F_u(x_i\beta)} - \frac{\sum_{i=1}^{N} \beta_3 \; S_i \; f_u(x_i\beta)}{\sum_{i=1}^{N} S_i \; F_u(x_i\beta)}
$$

where $f_u(.)$ is the PDF associated to the CDF $F_u$. We have shown above how to obtain consistent estimates of $\beta$ and $F_u$. Note that the Nadaraya-Watson estimator $\hat{F}_u(.)$ is a continuous and differentiable function. Therefore, we can get a consistent estimator of the density function of $u$ as $\hat{f}_u(z) = d\hat{F}_u(z)/dz$. Finally, we obtain a consistent estimator of $\frac{\partial W(0)}{\partial \delta}$ by replacing $\beta$, $F_u$, and $f_u$ with our consistent estimates of these objects.

**PROBLEM 3 (25 points): Consider the dynamic panel data logit model,**

$$y_{it} = I\{\ \beta\ y_{i,t-1} + \alpha_i - u_{it} \geq 0\ \}$$

**for** $i = 1, 2, ..., N$ **and** $t = 1, 2, ..., T$**, where** $u_{it}$ **is iid with a logistic distribution. Suppose that the number of periods in the sample is** $T = 4$**. Let** $Y_i = \{y_{i1},\ y_{i2},\ y_{i3},\ y_{i4}\}$ **be the choice history for individual** $i$**. We distinguish four sets of choice histories:**

$$\begin{array}{rcl} A & = & \{y_1, 1, 0, y_4\} \\ B & = & \{y_1, 0, 1, y_4\} \\ C & = & \{y_1, 1, 1, y_4\} \\ D & = & \{y_1, 0, 0, y_4\} \end{array}$$

**Question 3.1 (10 POINTS):**

**(a) Obtain the expression for the probability** $\Pr(Y_i \mid Y_i \in A \cup B,\ \alpha_i,\ y_{i1})$ **for this model.**

ANSWER: First, by definition of conditional probability and taking into account that $A$ and $B$ are disjoint events, we have that:

$$\Pr(Y_i | Y_i \in A \cup B, \alpha_i, y_{i1}) = \frac{\Pr(Y_i \mid \alpha_i,\ y_{i1})}{\Pr(A \cup B \mid \alpha_i,\ y_{i1})} = \frac{\Pr(Y_i | \alpha_i, y_{i1})}{\Pr(A | \alpha_i, y_{i1}) + \Pr(B | \alpha_i, y_{i1})}$$

The next step is to obtain the expressions for the probabilities $\Pr(A|\alpha_i, y_{i1})$ and $\Pr(B|\alpha_i, y_{i1})$. By the chain rule of conditional probabilities, and by the Markov structure of the model:

$$\begin{aligned} \Pr(A|\alpha_i, y_{i1}) &= \Pr(y_{i2} = 1 | \alpha_i, y_{i1})\ \ \Pr(y_{i3} = 0 | \alpha_i, y_{i2} = 1)\ \ \Pr(y_{i4} = y_4 | \alpha_i, y_{i3} = 0) \\ &= \frac{\exp(\beta y_1 + \alpha_i)}{1 + \exp(\beta y_1 + \alpha_i)} \frac{1}{1 + \exp(\beta + \alpha_i)} \frac{\exp(y_4 \alpha_i)}{1 + \exp(\alpha_i)} \\ &= \frac{\exp(\beta y_1 + \alpha_i)\ \ \exp(y_4 \alpha_i)}{\delta_i} \end{aligned}$$

where $\delta_i \equiv [1 + \exp(\beta y_1 + \alpha_i)]\ [1 + \exp(\beta + \alpha_i)]\ [1 + \exp(\alpha_i)]$. And

$$\begin{aligned} \Pr(B|\alpha_i, y_{i1}) &= \Pr(y_{i2} = 0 | \alpha_i, y_{i1})\ \ \Pr(y_{i3} = 1 | \alpha_i, y_{i2} = 0)\ \ \Pr(y_{i4} = y_4 | \alpha_i, y_{i3} = 1) \\ &= \frac{1}{1 + \exp(\beta y_1 + \alpha_i)} \frac{\exp(\alpha_i)}{1 + \exp(\alpha_i)} \frac{\exp(y_4[\beta + \alpha_i])}{1 + \exp(\beta + \alpha_i)} \\ &= \frac{\exp(\alpha_i)\ \ \exp(y_4[\beta + \alpha_i])}{\delta_i} \end{aligned}$$

Then, taking into account that $\Pr(Y_i | Y_i \in A \cup B, \alpha_i, y_{i1}) = \frac{\Pr(Y_i|\alpha_i, y_{i1})}{\Pr(A|\alpha_i, y_{i1}) + \Pr(B|\alpha_i, y_{i1})}$ we have that:

$$\begin{aligned} \Pr(A | A \cup B, \alpha_i, y_{i1}) &= \frac{\exp(\beta y_1 + \alpha_i)\ \ \exp(y_4 \alpha_i)}{\exp(\beta y_1 + \alpha_i)\ \ \exp(y_4 \alpha_i) + \exp(\alpha_i)\ \ \exp(y_4[\beta + \alpha_i])} \\ &= \frac{\exp(\beta y_1)\ \ \exp(\alpha_i)\ \ \exp(y_4 \alpha_i)}{[\exp(\beta y_1) + \exp(\beta y_4)]\ \ \exp(\alpha_i)\ \ \exp(y_4 \alpha_i)} \\ &= \frac{\exp(\beta y_1)}{\exp(\beta y_1) + \exp(\beta y_4)} \\ &= \frac{1}{1 + \exp(\beta [y_4 - y_1])} \end{aligned}$$

By definition, $\Pr(B|A\cup B,\alpha_i,y_{i1}) = 1-\Pr(A|A\cup B,\alpha_i,y_{i1})$, and therefore:

$$\Pr(B|A\cup B,\alpha_i,y_{i1}) = \frac{\exp(\beta[y_4-y_1])}{1+\exp(\beta[y_4-y_1])}$$

**(b) Show that this probability does not depend on the individual effect $\alpha_i$.**
ANSWER: It is obvious that $\frac{1}{1+\exp(\beta[y_4-y_1])}$ and $\frac{\exp(\beta[y_4-y_1])}{1+\exp(\beta[y_4-y_1])}$ do not depend on $\alpha_i$.

**Question 3.2. (15 POINTS):**
**(a) Obtain the expression of the conditional likelihood function $\sum_{i=1}^{N}\log\Pr(Y_i|Y_i\in A\cup B,\ y_{i1})$.**
ANSWER: Given a random sample of individuals, this Conditional log-likelihood function is:

$$\begin{aligned}
l(\beta) &= \sum_{i=1}^{N} 1\{Y_i\in A\}\ \ln\Pr(A|A\cup B,y_{i1}) + 1\{Y_i\in B\}\ \ln\Pr(B|A\cup B,y_{i1}) \\
&= \sum_{i=1}^{N} 1\{Y_i\in A\}\ \ln\left(\frac{1}{1+\exp(\beta[y_{i4}-y_{i1}])}\right) + 1\{Y_i\in B\}\ \ln\left(\frac{\exp(\beta[y_{i4}-y_{i1}])}{1+\exp(\beta[y_{i4}-y_{i1}])}\right) \\
&= \sum_{i=1}^{N} 1\{Y_i\in B\}\ \beta\ [y_{i4}-y_{i1}] - \sum_{i=1}^{N} 1\{Y_i\in A\cup B\}\ \ln(1+\exp(\beta\ [y_{i4}-y_{i1}]))
\end{aligned}$$

**(b) Derive the likelihood equation for the MLE of the parameter $\beta$.**
ANSWER: The likelihood equation is simply the first order condition of optimality $\frac{\partial l(\hat{\beta})}{\partial\beta} = 0$. Then, taking the derivative of the log-likeklihood equation with respect to $\beta$, we get:

$$\sum_{i=1}^{N} 1\{Y_i\in B\}\ [y_{i4}-y_{i1}] - \sum_{i=1}^{N} 1\{Y_i\in A\cup B\}\ \frac{[y_{i4}-y_{i1}]\ \exp\left(\hat{\beta}[y_{i4}-y_{i1}]\right)}{1+\exp\left(\hat{\beta}[y_{i4}-y_{i1}]\right)} = 0$$

**(c) Based on the likelihood equation, show that the Conditional MLE of $\beta$ can be written as:**

$$\hat{\beta} = \ln\left(\frac{\#\{1,1,0,0\}+\#\{0,0,1,1\}}{\#\{0,1,0,1\}+\#\{1,0,1,0\}}\right)$$

**where the notation $\#\{y_1,y_2,y_3,y_4\}$ represents the number of individuals in the sample with a choice history $\{y_1,y_2,y_3,y_4\}$.**
ANSWER: First, notice that the histories with $[y_{i4}-y_{i1}] = 0$ do not have any contribution to the likelihood equation. Therefore, the only histories that contribute to the likelihood equation are histories of type $A$ or $B$ with $[y_{i4}-y_{i1}] \neq 0$. These histories are: $\{1,1,0,0\}$ and $\{0,1,0,1\}$ of type $A$, and $\{0,0,1,1\}$ and $\{1,0,1,0\}$ of type $B$. For each of these histories, we have the following contributions to the left-hand-side of the likelihood equations. For $\{1,1,0,0\}$:

$$\frac{\exp\left(-\hat{\beta}\right)}{1+\exp\left(-\hat{\beta}\right)} = \frac{1}{1+\exp\left(\hat{\beta}\right)}$$

For $\{0,1,0,1\}$:

$$-\frac{\exp\left(\hat{\beta}\right)}{1+\exp\left(\hat{\beta}\right)}$$

For $\{0,0,1,1\}$:

$$1-\frac{\exp\left(\hat{\beta}\right)}{1+\exp\left(\hat{\beta}\right)}=\frac{1}{1+\exp\left(\hat{\beta}\right)}$$

For $\{1,0,1,0\}$:

$$-1+\frac{\exp\left(-\hat{\beta}\right)}{1+\exp\left(-\hat{\beta}\right)}=-1+\frac{1}{1+\exp\left(\hat{\beta}\right)}=\frac{-\exp\left(\hat{\beta}\right)}{1+\exp\left(\hat{\beta}\right)}$$

Therefore, we can write the likelihood equation as:

$$\begin{aligned}&\#\{1,1,0,0\}\left[\frac{1}{1+\exp\left(\hat{\beta}\right)}\right]+\#\{0,1,0,1\}\left[\frac{-\exp\left(\hat{\beta}\right)}{1+\exp\left(\hat{\beta}\right)}\right]\\&+\#\{0,0,1,1\}\left[\frac{1}{1+\exp\left(\hat{\beta}\right)}\right]+\#\{1,0,1,0\}\left[\frac{-\exp\left(\hat{\beta}\right)}{1+\exp\left(\hat{\beta}\right)}\right]\\=\;&0\end{aligned}$$

Multiplying this equation by $1+\exp\left(\hat{\beta}\right)$, we have:

$$[\#\{1,1,0,0\}+\#\{0,0,1,1\}]=[\#\{0,1,0,1\}+\#\{1,0,1,0\}]\;\;\exp\left(\hat{\beta}\right)$$

And solving for $\hat{\beta}$, we get:

$$\hat{\beta}=\ln\left(\frac{\#\{1,1,0,0\}+\#\{0,0,1,1\}}{\#\{0,1,0,1\}+\#\{1,0,1,0\}}\right)$$