ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Winter 2017
Instructor: Victor Aguirregabiria

SOLUTION TO FINAL EXAM
Tuesday, April 18, 2017. From 2:00pm-5:00pm (3 hours)

INSTRUCTIONS:
- **This is a closed-book exam.**
- **No study aids, including calculators, are allowed.**
- **Please, answer all the questions.**

TOTAL MARKS = 100

PROBLEM 1 (50 points). Let y_{it} be the logarithm of output produced by firm i at period t. The researcher has a panel dataset $\{y_{it} : i = 1, 2, ...N;\ t = 1, 2, ..., T\}$ where the number of firms N is large and the number of periods T is small. The researcher is interested in the relationship between firm's growth, $\Delta y_{it} \equiv y_{it} - y_{i,t-1}$, and firm size, $y_{i,t-1}$. She proposes the following model:

$$\Delta y_{it} = \delta\ y_{i,t-1} + \alpha_i + \gamma_t + u_{it}$$

where $E(\alpha_i) = E(u_{it}) = 0$; $E(\alpha_i\ u_{it}) = 0$; $var(\alpha_i) = \sigma_\alpha^2$; and u_{it} is not serially correlated. Note that this model is equivalent to $y_{it} = \delta\ y_{i,t-1} + \alpha_i + \gamma_t + u_{it}$ with $\beta \equiv 1 + \delta$.

Question 1.1. [5 points] According to Gibrat's hypothesis (often referred as Gibrat's Law), there is not a causal relationship between firm size and firm growth, i.e., $\delta = 0$ or $\beta = 1$. However, empirical applications presenting a Within-Group estimator of the regression of Δy_{it} on $y_{i,t-1}$ typically provide negative and a significant estimate of the parameter δ. Explain why this negative relationship can be spurious and not causal. For the case with $T = 3$, show that the asymptotic bias of the Within-Groups estimator is negative (Hint: when T=3 the WG estimator is the same as OLS in first differences).

ANSWER: The model is a dynamic panel data model. In this class of model, the Fixed-Effects (FE) or Within-Groups (WG) estimator is inconsistent as N goes to infinity and T is fixed. More specifically, this is an AR(1) panel data model and the WG estimator of β (or of δ) is asymptotically downward biased: $p\lim_{N\to\infty} \widehat{\beta}_{WG} < \beta$. Nickell (Econometrica, 1981) obtained the closed form expression of the asymptotic biased of the WG estimator for this AR(1) panel data model and showed that this biased is negative. For general T, this expression is convoluted. Here we consider the simpler case with $T = 3$.

We have the model in levels at $t = 2$ and $t = 3$:

$$y_{i2} = \beta \, y_{i1} + \alpha_i + \gamma_2 + u_{i2}$$

$$y_{i3} = \beta \, y_{i2} + \alpha_i + \gamma_3 + u_{i3}$$

The WG transformation is:

$$y_{i3} - \left(\frac{y_{i2} + y_{i3}}{2}\right) = \beta \left[y_{i2} - \left(\frac{y_{i1} + y_{i2}}{2}\right)\right] + \left[\gamma_3 - \left(\frac{\gamma_2 + \gamma_3}{2}\right)\right] + \left[u_{i3} - \left(\frac{u_{i2} + u_{i3}}{2}\right)\right]$$

Multiplying RHS and LHS by 2, we have that (for $T = 3$) this transformation is equivalent to the transformation in first differences:

$$y_{i3} - y_{i2} = \beta \, [y_{i2} - y_{i1}] + [\gamma_3 - \gamma_2] + [u_{i3} - u_{i2}]$$

Therefore, for $T = 3$, the WG estimator is equivalent to OLS in the regression of $[y_{i3} - y_{i2}]$ on $[y_{i2} - y_{i1}]$. Note that $[\gamma_3 - \gamma_2]$ is simply a constant term. Therefore,

$$\widehat{\beta}_{WG} = \frac{\sum_{i=1}^{N} [y_{i3} - y_{i2}] \; [y_{i2} - y_{i1}]}{\sum_{i=1}^{N} [y_{i2} - y_{i1}]^2} \quad \rightarrow_p \quad \frac{\mathbb{E}([y_{i3} - y_{i2}] \; [y_{i2} - y_{i1}])}{\mathbb{E}([y_{i2} - y_{i1}]^2)}$$

and this is equal to $\beta + \dfrac{\mathbb{E}([u_{i3} - u_{i2}] \; [y_{i2} - y_{i1}])}{\mathbb{E}([y_{i2} - y_{i1}]^2)}$. Therefore, the sign of the asymptotic bias is equal to the sign of $\mathbb{E}([u_{i3} - u_{i2}] \; [y_{i2} - y_{i1}])$.

In the AR(1) model, we have that,

$$y_{it} = u_{it} + \beta \; u_{i,t-1} + \beta^2 \; u_{i,t-2} + \beta^3 \; u_{i,t-3} + ...$$

This implies that,

$$\begin{aligned} \mathbb{E}([u_{i3} - u_{i2}] \; [y_{i2} - y_{i1}]) &= \mathbb{E}(u_{i3} \; y_{i2}) - \mathbb{E}(u_{i3} \; y_{i1}) - \mathbb{E}(u_{i2} \; y_{i2}) + \mathbb{E}(u_{i2} \; y_{i1}) \\ &= 0 - 0 - \sigma_u^2 + 0 = -\sigma_u^2 < 0 \end{aligned}$$

Question 1.2.[10 points] Describe in detail the Arellano-Bond moment conditions and the corresponding GMM estimator of β in this model. Present the closed-form expression for this estimator in this model.

ANSWER: To describe the AB estimator, it is convenient to see the model as a system of $T - 2$ equations in first differences:

$$\begin{cases} \text{Equation at } t = 3: & \Delta y_{i3} = \beta \; \Delta y_{i2} + \Delta\gamma_3 + \Delta u_{i3} \\ \text{Equation at } t = 4: & \Delta y_{i4} = \beta \; \Delta y_{i3} + \Delta\gamma_4 + \Delta u_{i4} \\ \vdots & \vdots \\ \text{Equation at } t = T: & \Delta y_{iT} = \beta \; \Delta y_{iT-1} + \Delta\gamma_T + \Delta u_{iT} \end{cases}$$

For each equation we have different valid instruments or moment conditions:

$$\begin{array}{ll} \text{For } t=3: & E[\Delta u_{i3}]=0;\ E[y_{i1}\ \Delta u_{i3}]=0 \\ \text{For } t=4: & E[\Delta u_{i4}]=0;\ E[y_{i1}\ \Delta u_{i4}]=0;\ E[y_{i2}\ \Delta u_{i4}]=0 \\ \vdots & \qquad\qquad\qquad \vdots \\ \text{For } t=T: & E[\Delta u_{iT}]=0;\ E[y_{i1}\ \Delta u_{iT}]=0;\ ...\ ;\ E[y_{i,T-2}\ \Delta u_{iT}]=0 \end{array}$$

The total number of moment conditions is $q=(T-2)+\dfrac{(T-2)(T-1)}{2}$. We can represent these q moment conditions in vector form as:

$$\mathbb{E}\left[\mathbf{Z}_i^{AB}\left(\Delta\mathbf{Y}_i-\beta\ \Delta\mathbf{Y}_{i(-1)}\right)\right]$$

where $\Delta\mathbf{Y}_i$ and $\Delta\mathbf{Y}_{i(-1)}$ are $(T-2)\times 1$ vectors, and $\mathbf{Z}_i^{AB}$ is the $q\times(T-2)$ matrix of instruments. Let $m_N(\beta)$ be the $q\times 1$ vector of sample moment conditions, i.e., the sample counterpart of $\mathbb{E}\left[\mathbf{Z}_i^{AB}\left(\Delta\mathbf{Y}_i-\beta\ \Delta\mathbf{Y}_{i(-1)}\right)\right]$

$$m_N(\beta)\ =\ N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{AB}\left(\Delta\mathbf{Y}_i-\beta\ \Delta\mathbf{Y}_{i(-1)}\right)$$

The Arellano-Bond GMM estimator is:

$$\hat{\beta}_{AB}=\underset{\{\beta\}}{\arg\min}\ m_N(\beta)'\ \hat{\Omega}^{-1}\ m_N(\beta)$$

where $\hat{\Omega}$ is a consistent estimator of the optimal weighting matrix, $\Omega=E\left(\mathbf{Z}_i\ \Delta\mathbf{U}_i\ \Delta\mathbf{U}_i'\ \mathbf{Z}_i\right)$, with $\Delta\mathbf{U}_i=(\Delta u_{i3},\Delta u_{i4},...,\Delta u_{iT})'$. Deriving the first order conditions and solving for $\hat{\beta}_{AB}$ we get:

$$\hat{\beta}_{AB}=\frac{\left(\displaystyle\sum_{i=1}^{N}\ \mathbf{Z}_i^{AB}\ \Delta\mathbf{Y}_{i(-1)}\right)'\hat{\Omega}^{-1}\left(\displaystyle\sum_{i=1}^{N}\mathbf{Z}_i^{AB}\ \Delta\mathbf{Y}_i\right)}{\left(\displaystyle\sum_{i=1}^{N}\ \mathbf{Z}_i^{AB}\ \Delta\mathbf{Y}_{i(-1)}\right)'\hat{\Omega}^{-1}\left(\displaystyle\sum_{i=1}^{N}\ \mathbf{Z}_i^{AB}\ \Delta\mathbf{Y}_{i(-1)}\right)}$$

Question 1.3. [10 points] Obtain the expression of the asymptotic variance of the Arellano-Bond estimator [Note: For simplicity, consider that $T=3$.] Show that this variance is infinite if Gibrat's hypothesis holds.

ANSWER: When $T=3$ we have only one equation in first differences, $\Delta y_{i3}=\beta\ \Delta y_{i2}+\Delta\gamma_3+\Delta u_{i3}$, and only two moment conditions: $\mathbb{E}[\Delta u_{i3}]=0$ and $\mathbb{E}[y_{i1}\ \Delta u_{i3}]=0$. In this case, AB estimator is equivalent to the Anderson-Hsiao estimator where the endogenous regressor Δy_{i2} is instrumented with y_{i1}. The expression of this estimator is:

$$\hat{\beta}_{AB}=\frac{\displaystyle\sum_{i=1}^{N}y_{i1}\ \Delta y_{i3}}{\displaystyle\sum_{i=1}^{N}y_{i1}\ \Delta y_{i2}}$$

Its asymptotic distribution is:

$$\sqrt{N}\left(\hat{\beta}_{AB}-\beta\right) = \frac{N^{-1/2}\sum_{i=1}^{N} y_{i1}\ \Delta u_{i3}}{N^{-1}\sum_{i=1}^{N} y_{i1}\ \Delta y_{i2}}$$

$$\underset{(as\ N\rightarrow\infty)}{\rightarrow_d} \quad N\left(0, \frac{Var\left(y_{i1}\ \Delta u_{i3}\right)}{E\left(y_{i1}\ \Delta y_{i2}\right)^2}\right)$$

Therefore,

$$Var\left(\hat{\beta}_{AB}\right) = \frac{1}{N}\frac{Var\left(y_{i1}\ \Delta u_{i3}\right)}{\mathbb{E}\left(y_{i1}\ \Delta y_{i2}\right)^2}$$

Now, we show that $\mathbb{E}\left(y_{i1}\ \Delta y_{i2}\right) = \dfrac{-\sigma_u^2}{1+\beta}$ and $Var\left(y_{i1}\ \Delta u_{i3}\right) = 2\sigma_u^2\ Var\left(y_{i1}\right)$. First, note that $Var(y_{i1}\Delta u_{i3}) = \mathbb{E}(y_{i1}^2\Delta u_{i3}^2)$, and by the law of iterative expectations, $\mathbb{E}(y_{i1}^2\Delta u_{i3}^2) = \mathbb{E}(y_{i1}^2)\ \mathbb{E}(\Delta u_{i3}^2)$. Given that $\mathbb{E}(\Delta u_{i3}^2) = 2\sigma_u^2$, and $\mathbb{E}\left(y_{i1}^2\right) = \dfrac{\sigma_\alpha^2}{(1-\beta)^2} + \dfrac{\sigma_u^2}{1-\beta^2}$, 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:

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

Then,

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

Under Gibrat's hypothesis, we have that $\beta = 1$, and it is clear that this variance is equal to infinite.

Question 1.4. [10 points] Describe in detail the Blundell-Bond moment conditions and the corresponding System-GMM estimator of β in this model.

ANSWER: To describe BB moment conditions, it is convenient to see the model as a system of $T-2$ equations in levels:

$$\begin{cases}\text{Equation at } t=3: & y_{i3} = \beta\ y_{i2} + \gamma_3 + \alpha_i + u_{i3}\\ \text{Equation at } t=4: & y_{i4} = \beta\ y_{i3} + \gamma_4 + \alpha_i + u_{i4}\\ \vdots & \vdots\\ \text{Equation at } t=T: & y_{iT} = \beta\ y_{iT-1} + \gamma_T + \alpha_i + u_{iT}\end{cases}$$

For each equation we have different valid instruments or moment conditions:

$$\begin{array}{ll} \text{For } t=3: & \mathbb{E}[\alpha_i + u_{i3}] = 0;\ \mathbb{E}[\Delta y_{i2}\ (\alpha_i + u_{i3})] = 0 \\ \text{For } t=4: & \mathbb{E}[\alpha_i + u_{i4}] = 0;\ \mathbb{E}[\Delta y_{i2}\ (\alpha_i + u_{i2})] = 0;\ \mathbb{E}[\Delta y_{i3}\ (\alpha_i + u_{i4})] = 0 \\ \quad\vdots & \qquad\qquad\qquad\vdots \\ \text{For } t=T: & \mathbb{E}[\alpha_i + u_{iT}] = 0;\ \mathbb{E}[\Delta y_{i2}\ (\alpha_i + u_{iT})] = 0;\ \ldots\ ;\ \mathbb{E}[\Delta y_{iT-1}\ (\alpha_i + u_{iT})] = 0 \end{array}$$

The total number of moment conditions is $q = (T-2) + \dfrac{(T-2)(T-1)}{2}$. We can represent these q moment conditions in vector form as:

$$\mathbb{E}\left[\mathbf{Z}_i^{BB}\left(\mathbf{Y}_i - \beta\ \mathbf{Y}_{i(-1)}\right)\right]$$

where $\mathbf{Y}_i$ and $\mathbf{Y}_{i(-1)}$ are $(T-2)\times 1$ vectors, and $\mathbf{Z}_i^{BB}$ is the $q \times (T-2)$ matrix of instruments.

The system-GMM estimator combines AB and BB moment conditions. We can represent these moment conditions as:

$$\mathbb{E}\left[\begin{pmatrix} \mathbf{Z}_i^{AB} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_i^{BB} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{Y}_i - \beta\ \Delta\mathbf{Y}_{i(-1)} \\ \mathbf{Y}_i - \beta\ \mathbf{Y}_{i(-1)} \end{pmatrix}\right] = \mathbb{E}\left[\begin{array}{c} \mathbf{Z}_i^{AB}\left[\Delta\mathbf{Y}_i - \beta\ \Delta\mathbf{Y}_{i(-1)}\right] \\ \mathbf{Z}_i^{BB}\left[\mathbf{Y}_i - \beta\ \mathbf{Y}_{i(-1)}\right] \end{array}\right] = 0$$

Let $m_N(\beta)$ be the $2q \times 1$ vector of sample moment conditions. The sample counterpart is:

$$m_N(\beta) \quad = \quad N^{-1}\sum_{i=1}^{N}\left[\begin{array}{c} \mathbf{Z}_i^{AB}\left[\Delta\mathbf{Y}_i - \beta\ \Delta\mathbf{Y}_{i(-1)}\right] \\ \mathbf{Z}_i^{BB}\left[\mathbf{Y}_i - \beta\ \mathbf{Y}_{i(-1)}\right] \end{array}\right]$$

The System-GMM estimator is:

$$\hat{\beta}_{Sys} = \arg\min_{\{\beta\}}\ m_N(\beta)'\ \hat{\Omega}^{-1}\ m_N(\beta)$$

where $\hat{\Omega}$ is a consistent estimator of the optimal weighting matrix. Deriving the first order conditions and solving for $\hat{\beta}_{Sys}$ we get:

$$\hat{\beta}_{Sys} = \frac{\left(\displaystyle\sum_{i=1}^{N} \begin{array}{c} \mathbf{Z}_i^{AB}\Delta\mathbf{Y}_{i(-1)} \\ \mathbf{Z}_i^{BB}\mathbf{Y}_{i(-1)} \end{array}\right)' \hat{\Omega}^{-1} \left(\displaystyle\sum_{i=1}^{N} \left[\begin{array}{c} \mathbf{Z}_i^{AB}\Delta\mathbf{Y}_i \\ \mathbf{Z}_i^{BB}\mathbf{Y}_i \end{array}\right]\right)}{\left(\displaystyle\sum_{i=1}^{N} \begin{array}{c} \mathbf{Z}_i^{AB}\Delta\mathbf{Y}_{i(-1)} \\ \mathbf{Z}_i^{BB}\mathbf{Y}_{i(-1)} \end{array}\right)' \hat{\Omega}^{-1} \left(\displaystyle\sum_{i=1}^{N} \begin{array}{c} \mathbf{Z}_i^{AB}\Delta\mathbf{Y}_{i(-1)} \\ \mathbf{Z}_i^{BB}\mathbf{Y}_{i(-1)} \end{array}\right)}$$

Question 1.5. [10 points] Obtain the expression of the asymptotic variance of Blundell-Bond estimator. [Note: For simplicity, consider that $T = 3$ and use only Blundell-Bond moment conditions.] Show that this variance is finite for any value of $\beta \equiv 1 + \delta$ within $[0, 1]$, including $\beta = 1$.

ANSWER: When $T = 3$ we have only one equation in levels with valid BB instruments: $y_{i3} = \beta\ y_{i2} + \gamma_3 + \alpha_i + u_{i3}$, and two moment conditions: $\mathbb{E}[\alpha_i + u_{i3}] = 0$ and $\mathbb{E}[\Delta y_{i2}\ (\alpha_i + u_{i3})] = 0$. In this case, the BB estimator is equivalent to the following IV estimator:

$$\hat{\beta}_{BB} = \frac{\sum_{i=1}^{N} \Delta y_{i2}\ y_{i3}}{\sum_{i=1}^{N} \Delta y_{i2}\ y_{i2}}$$

Its asymptotic distribution is:

$$\begin{aligned} \sqrt{N}\left(\hat{\beta}_{BB} - \beta\right) &= \frac{N^{-1/2}\sum_{i=1}^{N} \Delta y_{i2}\ \ (\alpha_i + u_{i3})}{N^{-1}\sum_{i=1}^{N} \Delta y_{i2}\ y_{i2}} \\ &\to_d \frac{N\left(0, Var[\Delta y_{i2}\ \ (\alpha_i + u_{i3})]\right)}{\mathbb{E}\left(\Delta y_{i2}\ y_{i2}\right)} \end{aligned}$$

Therefore,

$$Var\left(\hat{\beta}_{BB}\right) = \frac{1}{N}\frac{Var[\Delta y_{i2}\ \ (\alpha_i + u_{i3})]}{\mathbb{E}\left(\Delta y_{i2}\ y_{i2}\right)^2}$$

Given that: $Var[\Delta y_{i2}\ \ (\alpha_i + u_{i3})] = \mathbb{E}\left(\Delta y_{i2}^2\right)\ \mathbb{E}\left((\alpha_i + u_{i3})^2\right) = 2\sigma_u^2\ \left(\sigma_u^2 + \sigma_\alpha^2\right)/\left(1+\beta\right)$; and $\mathbb{E}\left(\Delta y_{i2}\ y_{i2}\right) = \sigma_u^2/\left(1+\beta\right)$; we have that:

$$\begin{aligned} Var\left(\hat{\beta}_{BB}\right) &= \frac{1}{N}\frac{2\sigma_u^2\left(\sigma_u^2 + \sigma_\alpha^2\right)/\left(1+\beta\right)}{\sigma_u^4/\left(1+\beta\right)^2} \\ &= \frac{1}{N}\ 2\left(1+\beta\right)\left(1 + \frac{\sigma_\alpha^2}{\sigma_u^2}\right) \end{aligned}$$

In contrast to the AB estimator, the variance of this estimator is finite when $\beta = 1$.

Question 1.6. [5 points] Given the System-GMM estimator, consider testing the null hypothesis $\beta = 1$. Comment potential problem(s) with the standard t-test for this null hypothesis.

ANSWER: The derivation of Blundell and Bond moment conditions require the assumption that the model is stationary such that the condition $|\beta| < 1$ should hold. Therefore, when β is exactly equal to 1, the AB moment conditions do not have any identification power and the BB moment conditions do not hold. This implies that when in the DGP with true $\beta = 1$ the parameter β is not identified, at least from the AB and BB moment conditions. The standard regularity conditions for the t-test (and for Wald test, LM test, and LR test) do not hold. Andrews (Econometrica, 2001) shows that under these conditions the t-test statistic does not have the well-known standard normal distribution and it should be corrected to construct a valid test.

PROBLEM 2 (50 points). Consider the random coefficients multinomial choice model,

$$Y_n = \arg \max_{j \in \{0,1,...,J\}} [\beta_n \ X_j + \varepsilon_{nj}]$$

where n is the index for individuals/observations, and j is the index for choice alternatives. X_j is a (continuous) characteristic of choice alternative j, e.g., the price of product j. β_n is a random coefficient with the following properties: $\beta_n = \beta + \sigma \ v_n$, where v_n is *i.i.d.* standard normal. Variables ε_{nj} are *i.i.d.* over (n, j) with a Type I extreme value distribution and independent of v_n. Define the conditional choice probabilities (CCPs) $P_j(X) \equiv \Pr(Y_n = j \mid X)$ with $X = (X_j : j = 0, 1, ..., J)$. The researcher is interested in the estimation of this conditional choice probabilities, and especially in the estimation of the partial effects $\dfrac{\partial P_j(\mathbf{X})}{\partial X_k}$. The dataset consists of a cross-section of individuals with information about their choices: $\{y_n : n = 1, 2, ..., N\}$. The value of the vector of product characteristics, X, is the same for every individual/observation in the sample, i.e., all the individuals have the same set of feasible choice alternatives. The sample size N is large, i.e., asymptotics in N.

Question 2.1. [5 points] **Describe a nonparametric estimator of the CCPs $P_j(X) = \Pr(Y_n = j|\mathbf{X})$. Explain why this nonparametric approach cannot identify the partial effects $\dfrac{\partial P_j(\mathbf{X})}{\partial X_k}$.**

ANSWER: Note that the vector of product characteristics $\mathbf{X}$ does not have any sample variation. Therefore, we can obtain a nonparametric estimator of $P_j(\mathbf{X})$ only for the value of $\mathbf{X}$ that we observe. This nonparametric estimator can be simply the frequency estimator of $\Pr(Y_n = j)$:

$$\widehat{P_j(\mathbf{X})} = N^{-1} \sum\nolimits_{n=1}^{N} 1\{Y_n = j\}$$

It is clear that this nonparametric estimator does not provide any information about how $P_j(\mathbf{X})$ varies when we change a component of $\mathbf{X}$. More specifically, it does not provide any information about $\dfrac{\partial P_j(\mathbf{X})}{\partial X_k}$.

Question 2.2. [10 points] **Consider the Random utility model with parameters $\theta = (\beta, \sigma)$. Describe the log-likelihood function of this model and data. Obtain the first order conditions (likelihood equations) that define the MLE of θ. Show that, for any Random Utility Model, these likelihood equations have the following form:**

$$\sum_{j=0}^{J} \frac{\partial P_j(\theta)}{\partial \theta} \ \frac{1}{P_j(\theta)} \left[\frac{N_j}{N} - P_j(\theta) \right] = 0$$

where N_j is the number of observations in the sample with $Y_n = j$.

ANSWER: The log-likelihood function is:

$$\ell(\theta) = \sum_{n=1}^{N} \sum_{j=0}^{J} 1\{Y_n = j\} \ \ln P_j(\theta)$$

The first order conditions of optimality are:

$$\frac{\partial \ell(\theta)}{\partial \theta} = \sum_{n=1}^{N} \sum_{j=0}^{J} 1\{Y_n = j\} \ \frac{\partial P_j(\theta)}{\partial \theta} \ \frac{1}{P_j(\theta)} = 0$$

Interchanging the sums, and taking into account that $N_j = \sum_{n=1}^{N} 1\{Y_n = j\}$, we have:

$$\sum_{j=0}^{J} \frac{N_j}{N} \ \frac{\partial P_j(\theta)}{\partial \theta} \ \frac{1}{P_j(\theta)} = 0$$

We also have that $\sum_{j=0}^{J} P_j(\theta) = 1$, and therefore $\sum_{j=0}^{J} \dfrac{\partial P_j(\theta)}{\partial \theta} = 0$, such that:

$$\begin{aligned} \sum_{j=0}^{J} \frac{N_j}{N} \ \frac{\partial P_j(\theta)}{\partial \theta} \ \frac{1}{P_j(\theta)} - \sum_{j=0}^{J} \frac{\partial P_j(\theta)}{\partial \theta} &= 0 \\ \Rightarrow \qquad \sum_{j=0}^{J} \frac{\partial P_j(\theta)}{\partial \theta} \ \frac{1}{P_j(\theta)} \left[\frac{N_j}{N} - P_j(\theta) \right] &= 0 \end{aligned}$$

Question 2.3. [10 points] Suppose that the model is restricted to a standard logit model, i.e., no random coefficients, $\sigma = 0$. Obtain the expressions for $P_j(\beta)$ and $\dfrac{\partial P_j(\beta)}{\partial \beta}$ in this model. Show that the first order condition that defines the MLE of β is:

$$\sum_{j=0}^{J} X_j \left[\frac{N_j}{N} - P_j(\beta) \right] = 0$$

[Hint: $\dfrac{\partial P_j(\beta)}{\partial \beta} = \dfrac{\partial P_j(\beta)}{\partial \delta_j} \dfrac{\partial \delta_j}{\partial \beta} + \sum_{k \neq j} \dfrac{\partial P_j(\beta)}{\partial \delta_k} \dfrac{\partial \delta_k}{\partial \beta}$, with $\delta_j \equiv \beta \ X_j$]. Propose an algorithm to compute the MLE of β. Comment on the computational properties of this algorithm.

ANSWER: When $\sigma = 0$ we have the standard logit model with:

$$P_j(\beta) = \Lambda_j(\beta) = \frac{\exp\{\beta \ X_j\}}{\sum_{k=0}^{J} \exp\{\beta \ X_k\}} = \frac{\exp\{\delta_j\}}{\sum_{k=0}^{J} \exp\{\delta_k\}}$$

with $\delta_j \equiv \beta \ X_j$. For the logit model, we have that $\dfrac{\partial \Lambda_j}{\partial \delta_j} = \Lambda_j(1 - \Lambda_j)$; and for $k \neq j$, $\dfrac{\partial \Lambda_j}{\partial \delta_k} = -\Lambda_j \Lambda_k$. Therefore,

$$\begin{aligned} \frac{\partial P_j(\beta)}{\partial \beta} &= \frac{\partial \Lambda_j(\beta)}{\partial \delta_j} \frac{\partial \delta_j}{\partial \beta} + \sum_{k \neq j} \frac{\partial \Lambda_j(\beta)}{\partial \delta_k} \frac{\partial \delta_k}{\partial \beta} \\ &= \Lambda_j(1 - \Lambda_j) \ X_j - \sum_{k \neq j} \Lambda_j \Lambda_k \ X_k \\ &= \Lambda_j \left[X_j - \overline{X}(\beta) \right] \end{aligned}$$

with $\overline{X}(\beta) \equiv \sum_{k=0}^{J} \Lambda_k(\beta)\ X_k$. This implies that $\dfrac{\partial P_j(\beta)}{\partial \beta}\ \dfrac{1}{P_j(\beta)} = \dfrac{P_j(\beta)\left[X_j - \overline{X}(\beta)\right]}{P_j(\beta)} = X_j - \overline{X}(\beta)$. The first order conditions becomes:

$$\sum_{j=0}^{J} \left[X_j - \overline{X}(\beta)\right] \left[\frac{N_j}{N} - P_j(\beta)\right] = 0$$

Note that $\sum_{j=0}^{J} \overline{X}(\beta) \left[\dfrac{N_j}{N} - P_j(\beta)\right] = \overline{X}(\beta) \sum_{j=0}^{J} \left[\dfrac{N_j}{N} - P_j(\beta)\right] = \overline{X}(\beta)\ [1-1] = 0$. And we can write the first order conditions as:

$$\sum_{j=0}^{J} X_j \left[\frac{N_j}{N} - P_j(\beta)\right] = 0$$

We can use a Newton's method to compute the MLE of β. The expression of the Newton's method is:

$$\widehat{\beta}_{K+1} = \widehat{\beta}_K - \left[\frac{\partial^2 \ell(\widehat{\beta}_K)}{\partial \beta^2}\right]^{-1} \left[\frac{\partial \ell(\widehat{\beta}_K)}{\partial \beta}\right]$$

where $\dfrac{\partial \ell(\beta)}{\partial \beta} = \sum_{j=0}^{J} X_j \left[\dfrac{N_j}{N} - P_j(\beta)\right]$, and $\dfrac{\partial^2 \ell(\widehat{\beta}_K)}{\partial \beta^2} = -\sum_{j=0}^{J} X_j\ P_j(\beta)\ \left[X_j - \overline{X}(\beta)\right]$. The log-likelihood function of the logit model is globally concave in β, such that the Newton's method always converges to the MLE.

Question 2.4. [10 points] In the context of this standard logit model, describe the property of Independence of Irrelevant Alternatives. Explain why this is not an attractive property when the researcher is interested in the estimation of the Average Partial Effects $\dfrac{\partial P_j(\mathbf{X})}{\partial X_k}$.

ANSWER: The logit model imposes the restriction that the ratio between the probabilities of two alternatives, say j and i, depends ONLY on the utilities / characteristics of these alternatives:

$$\frac{P_j(\mathbf{X})}{P_k(\mathbf{X})} = \exp\{\beta[X_j - X_k]\}$$

Therefore, if we change the choice set by adding or/and removing alternatives, the ratios between probabilities should not change. This property can generate unrealistic predictions, e.g., blue bus / red bus example.

The IIA property restricts also the partial effects $\dfrac{\partial P_j(\mathbf{X})}{\partial X_k}$. For any alternative $j \neq k$:

$$\frac{\partial P_j(\mathbf{X})}{\partial X_k} = -\beta\ P_j(\mathbf{X})\ P_k(\mathbf{X})$$

Suppose that X_k is the price of product k. This expression implies that a change in the price of product k affects the "market shares" (CCPs) of two products, j and j', proportionally to the their market shares. If $P_j(\mathbf{X}) = P_{j'}(\mathbf{X})$, then the effect is the same, regardless of whether product k is very similar to product j but very different to product j' in terms of product characteristics X.

Question 2.5. [15 points] Consider the Random Coefficients Logit model (i.e., $\sigma > 0$). The log-likelihood function of this model is not globally concave in β and σ. This complicates the computation of the MLE. Instead, consider the following simulation-based estimator. For every observation n in the sample, let v_n be a random draw from the standard normal distribution. Then, using $v_n's$ as if they were observable explanatory variables, estimate β and σ in the standard logit model $Y_n = \arg\max_{j\in\{0,1,...,J\}}[\beta\ X_j + \sigma\ X_j\ v_n + \varepsilon_{nj}]$.

(a) Write the expression of the (pseudo) log-likelihood function that treats simulated $v_n's$ 'as if' they were observable.

(b) Show that the first order conditions that define the estimator of β and σ are:

$$\sum_{n=1}^{N}\sum_{j=0}^{J} X_j\ \left[1\{Y_n = j\} - \Lambda_j(v_n,\theta)\right] \quad = \quad 0$$

$$\sum_{n=1}^{N}\sum_{j=0}^{J} X_j\ v_n\ \left[1\{Y_n = j\} - \Lambda_j(v_n,\theta)\right] \quad = \quad 0$$

with $\Lambda_j(v_n,\theta) \equiv \exp\{\beta X_j + \sigma X_j v_n\} / \sum_{k=0}^{J} \exp\{\beta X_k + \sigma X_k v_n\}$.

(c) Show that this estimator is consistent as N goes to infinity.

ANSWER:

(a) The log-likelihood function is:

$$\ell(\theta; v) = \sum\nolimits_{n=1}^{N} \sum\nolimits_{j=0}^{J} 1\{Y_n = j\}\ \ln \Lambda_j(v_n,\theta)$$

where $\Lambda_j(v_n,\theta) \equiv \exp\{\beta X_j + \sigma X_j v_n\} / \sum_{k=0}^{J} \exp\{\beta X_k + \sigma X_k v_n\}$.

(b) The first order conditions of optimality are:

$$\frac{\partial \ell(\theta; v)}{\partial \beta} = 0 \Rightarrow \quad \sum\nolimits_{n=1}^{N} \sum\nolimits_{j=0}^{J} 1\{Y_n = j\}\ \frac{\partial \Lambda_j(v_n,\theta)}{\partial \beta} \frac{1}{\Lambda_j(v_n,\theta)} \quad = 0$$

$$\frac{\partial \ell(\theta; v)}{\partial \sigma} = 0 \Rightarrow \quad \sum\nolimits_{n=1}^{N} \sum\nolimits_{j=0}^{J} 1\{Y_n = j\}\ \frac{\partial \Lambda_j(v_n,\theta)}{\partial \sigma} \frac{1}{\Lambda_j(v_n,\theta)} \quad = 0$$

Define $\delta_{jn} = \beta X_j + \sigma X_j v_n$. Then,

$$\begin{aligned}
\frac{\partial \Lambda_j(v_n,\theta)}{\partial \beta} \quad &= \quad \frac{\partial \Lambda_j(v_n,\theta)}{\partial \delta_{jn}} \frac{\partial \delta_{jn}}{\partial \beta} + \sum\nolimits_{k\neq j} \frac{\partial \Lambda_j(v_n,\theta)}{\partial \delta_{kn}} \frac{\partial \delta_{kn}}{\partial \beta} \\
&= \quad \Lambda_{jn}(1-\Lambda_{jn})\ X_j - \sum\nolimits_{k\neq j} \Lambda_{jn}\Lambda_{kn}\ X_k \\
&= \quad \Lambda_{jn}\left[X_j - \overline{X}_n(v_n,\theta)\right]
\end{aligned}$$

with $\overline{X}(v_n,\theta) \equiv \sum_{k=0}^{J} \Lambda_k(v_n,\theta)\ X_k$. And,

$$\begin{aligned}
\frac{\partial \Lambda_j(v_n,\theta)}{\partial \sigma} &= \frac{\partial \Lambda_j(v_n,\theta)}{\partial \delta_{jn}} \frac{\partial \delta_{jn}}{\partial \sigma} + \sum_{k\neq j} \frac{\partial \Lambda_j(v_n,\theta)}{\partial \delta_{kn}} \frac{\partial \delta_{kn}}{\partial \sigma} \\
&= \Lambda_{jn}(1-\Lambda_{jn})\ X_j v_n - \sum_{k\neq j} \Lambda_{jn}\Lambda_{kn}\ X_k v_n \\
&= \Lambda_{jn}\ v_n\ \left[X_k - \overline{X}_n(v_n,\theta)\right]
\end{aligned}$$

Note that $\sum_{j=0}^{J} \overline{X}(v_n,\theta)\,[1\{Y_n=j\} - \Lambda_j(v_n,\theta)] = \overline{X}(v_n,\theta) \sum_{j=0}^{J} [1\{Y_n=j\} - \Lambda_j(v_n,\theta)] = \overline{X}(v_n,\theta)$ $[1-1]=0$. Then, we can write the first order conditions as:

$$\begin{aligned}
\sum_{n=1}^{N}\sum_{j=0}^{J} X_j\ \left[1\{Y_n=j\} - \Lambda_j(v_n,\theta)\right] &= 0 \\
\sum_{n=1}^{N}\sum_{j=0}^{J} X_j\ v_n\ \left[1\{Y_n=j\} - \Lambda_j(v_n,\theta)\right] &= 0
\end{aligned}$$

(c) Consistency of the estimator as N goes to infinity. Note that as N goes to infinity,

$$\begin{aligned}
N^{-1}\sum_{n=1}^{N}\sum_{j=0}^{J} X_j\,[1\{Y_n=j\} - \Lambda_j(v_n,\theta)] &\to_p \sum_{j=0}^{J} X_j\ \mathbb{E}\left(1\{Y_n=j\} - \Lambda_j(v_n,\theta)\right) \\
N^{-1}\sum_{n=1}^{N}\sum_{j=0}^{J} X_j v_n\,[1\{Y_n=j\} - \Lambda_j(v_n,\theta)] &\to_p \sum_{j=0}^{J} X_j\ \mathbb{E}\left(v_n\,[1\{Y_n=j\} - \Lambda_j(v_n,\theta)]\right)
\end{aligned}$$

Now, we show that, at the true parameters θ^0, the expectations $\mathbb{E}\left(1\{Y_n=j\} - \Lambda_j(v_n,\theta^0)\right)$ and $\mathbb{E}\left(v_n\left[1\{Y_n=j\} - \Lambda_j(v_n,\theta^0)\right]\right)$ are equal to zero.

$$\begin{aligned}
\mathbb{E}\left(1\{Y_n=j\} - \Lambda_j(v_n,\theta^0)\right) &= \mathbb{E}\left(1\{Y_n=j\}\right) - \mathbb{E}_{v_n}\left(\Lambda_j(v_n,\theta^0)\right) \\
&= \mathbb{E}\left(1\{Y_n=j\}\right) - P_j(\theta^0) = 0
\end{aligned}$$

And using iterative expectations,

$$\begin{aligned}
\mathbb{E}\left(v_n\left[1\{Y_n=j\} - \Lambda_j(v_n,\theta^0)\right]\right) &= \mathbb{E}_{v_n}\left(v_n\ \mathbb{E}_{\varepsilon_n|v_n}\left[1\{Y_n=j\} - \Lambda_j(v_n,\theta^0)\right]\right) \\
&= \mathbb{E}_{v_n}\left(v_n\ \left[\Lambda_j(v_n,\theta^0) - \Lambda_j(v_n,\theta^0)\right]\right) = 0
\end{aligned}$$

This result implies that this pseudo maximum likelihood estimator can be interpreted as a Method of Moments estimator where the moments are valid and identify the parameters (β,σ). The model satisfies standard regularity conditions of the Method of Moments estimator. Therefore, this estimator of θ^0 is root-N consistent and asymptotically normal estimator. The main computational advantage of the estimator is that the pseudo log-likelihood function is globally concave in (β,σ) such that Newton's method always converges to the consistent pseudo-MLE.