Université Toulouse 1 Capitole
Ecole d'économie de Toulouse

Année universitaire 2015-2016

Session 1

Semestre 2

Master 1 Economics, Economics & Statistics, Economics & Law

Epreuve : Applied Econometrics

Date de l'épreuve : 27 janvier 2016

Durée de l'épreuve : 1h

Liste des documents autorisés : none.

Liste des matériels autorisés : none.

Nombre de pages (y compris page de garde) : 15

Instructions

- *Duration: 1 hour.*
- *No document allowed*
- *You have to answer to questions related to **3 topics** of your choice.*
- *Answers have to be written on the last page of the exam, that will have to be included in the exam double sheet.*
- *There is a unique correct answer for each question.*
- *Grading for each question:*
  - *No answer: 0 point*
  - *Correct answer: 1 point*
  - ***!!!!! Wrong answer: -0.5 point !!!!!***
- *Total number of points = 18*

Topic 1. DEV: Natural Experiments in Development Economics (Matteo Bobba)

1. Suppose you observe outcomes $Y_{i,t}$ for individuals i over different time periods t. At time $D_t = 1$ some groups of individuals ($T_g = 1$) are exposed to an exogenous shock, which likely affects individual outcomes. Consider the following specification:

$$Y_{i,t} = \alpha + \gamma T_g + \lambda D_t + \beta \left[T_g \times D_t\right] + \epsilon_{i,g,t}.$$

The parameter β captures the effect of the shock on outcomes and allows you to get rid of time-invariant unobserved factors at the individual-level.

A. True

B. False

2. Jayachandran's paper "Selling Labor Low" employs an instrumental variable approach mainly because the explanatory variable of interest, crop yields, is likely measured with error.

A. True

B. False

3. Regression discontinuity designs require a lot of observations in order to fit well the data around the threshold of the assignment variable.

A. True

B. False

4. Consider the following simple linear regression model

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

where $Cov(X_i, \epsilon_i) \neq 0$. A good instrumental variable Z_i is one that:

A. It is difficult to exclude from the equation for Y_i

B. It is only weakly correlated with the variable X_i

C. It affects Y_i only through its effect on X_i

D. It is randomly assigned across i

5. Which of the following piece of evidence in the paper by Banerjee and Duflo "Do Firms Want to Borrow More?" is NOT a test for the parallel trend assumption?
    A. The fact that the firms in the sample solely borrow from the bank for which the authors collect the data from
    B. The fact that they found symmetric effects of credit expansion and contraction on firms revenues and costs
    C. The fact that the reform only affects loan amounts and not the selection of firms getting a positive debt increment
    D. The fact that there are no effects on sales, costs and profits for the subsample without a change in credit limit

6. Regression discontinuity designs are invalid if individuals can precisely manipulate the assignment variable. How can you test this assumption?
    A. Check that pre-determined covariates have the same distribution just below and above the threshold
    B. Plot the first-stage relationship and inspect the jump in treatment assignment
    C. Look for some "placebo" test in order to assess the validity of the exclusion restriction
    D. Check for the presence of any "bumps" in the density of the assignment variable at the threshold

Topic 2. IO: Industrial Organization (Yinghua He)

1. In a binary discrete choice model, we can estimate the unknown parameters $(\alpha, \beta, \gamma, \sigma)$ in the following model when observing the data $(x_1, d_1), ..., (x_i, d_i), ..., (x_n, d_n)$ and $d_i \in \{0, 1\}$:

$$\begin{aligned} u_i &= \alpha + x_i\beta + \sigma\epsilon_i, \\ d_i &= 1_{[u_i > \gamma]}, \end{aligned}$$

where $\epsilon \sim \mathcal{N}(0, 1)$.

A. True

B. False

2. Suppose that we use a nested logit model to study the choice of transportation mode among $(Driving, Bus, ExpressBus, Train)$. We put Bus and $ExpressBus$ together as a group. In this way, we allow the utility shocks (ϵ) of choosing Bus or $ExpressBus$ to be correlated.

A. True

B. False

3. Consider a model of product choice where consumer i in market m assigns indirect utility u_{imj} to product j:

$$\begin{aligned} u_{imj} &= \alpha\left(y_{im} - p_{mj}\right) + x_j\beta + \xi_{mj} + \epsilon_{imj}, \\ u_{im0} &= \alpha y_{im} + \epsilon_{im0}, \end{aligned}$$

where $j = 0$ indicates the outside option; ϵ_{imj} has the type I extreme value distribution. If we have market share data of each product from $m = 1, ..., M$ markets, we can estimate the unknown coefficients (α, β) only if we have the distribution of income (y_{im}) from each market.

A. True

B. False

4. In the same logit demand model as in the previous question, let us denote the market share of j in market m by S_{mj}. The price elasticity of S_{mj} with respect to p_j is:

A. $-\alpha p_{mj}\left(1 - S_{mj}\right)$

B. $-\alpha p_{mj}\left(1 - S_{mk}\right)$

C. $-\alpha p_{mj} S_{mj}$

D. $-\alpha p_{mj} S_{mk}$

5. Suppose that we have the following model for market share of $j(= 1, ..., J)$ in market $m(= 1, ..., M)$:

$$S_{mj} = \frac{\exp(\delta_{mj})}{1 + \sum_{k=1}^{J} \exp(\delta_{mk})},$$

where $\delta_{mj} = x_j\beta - \alpha p_{mj} + \xi_{mj}$. Moreover, we know the market share of the outside option is:

$$S_{m0} = \frac{1}{1 + \sum_{k=1}^{J} \exp(\delta_{mk})}.$$

We have data on S_{m0}, S_{mj}, x_j, and p_{mj} from each product in each market. Which one of the following statements is incorrect?

A. We can use the above equations to derive an equation for $j = 1, ..., J$ and $m = 1, ..., M$:

$$\ln \frac{S_{mj}}{S_{m0}} = x_j\beta - \alpha p_{mj} + \xi_{mj}.$$

B. To use the above equations to estimate the unknown coefficients (β, α), the number of observations is $J \times M$.

C. The estimates will be biased if we estimate the model in (A) by OLS, because price p_{mj} is endogenous.

D. We can use our data and estimates to estimate δ_{mj}, but we cannot know the exact value of δ_{mj}.

6. Suppose that in a market all firms are identical and earn the following profits:

$$\pi(N) = V(N, \mathbf{Z}, \mathbf{W}, \alpha, \beta)S(\mathbf{Y}, \lambda) - F(\mathbf{W}, \gamma) + \epsilon \equiv V(N)S - F + \epsilon,$$

where

- N is the number of incumbent firms in a given market.
- S is the potential market size (e.g. population).
- $V(N)$ are the variable profits per firm and per capita.
- F are fixed costs.
- $\mathbf{Z}, \mathbf{W}, \mathbf{Y}$ are observed market characteristics.
- ϵ unobserved market-level heterogeneity.

We assume that firms enter as long as it is profitable, which implies that the probability of observing n firms in a given market is:

A. $\Pr[N = n] = \Pr[F - V(n+1)S \leq \epsilon < F - V(n)S]$

B. $\Pr[N = n] = \Pr[F - V(n)S \leq \epsilon < F - V(n+1)S]$

C. $\Pr[N = n] = \Pr[V(n)S - F + \epsilon \geq 0]$

D. $\Pr[N = n] = \Pr[V(n+1)S - F + \epsilon < 0]$

Topic 3. FE: Financial Econometrics (Jihyun Kim)

1. Consider a return process with heteroskedasticity

$$r_t = \bar{r} + u_t, \quad \mathbb{E}[u_t] = 0, \ \mathbb{E}[u_t^2] = \sigma_t^2.$$

The null hypothesis of no return predictability can be tested by the variance ratio test suggested by Lo and MacKinlay (1988).

A. True

B. False

2. Consider predictive regressions with various investment horizon K

$$\sum_{k=1}^{K} r_{t+k} = \beta_K x_t + u_{t,K},$$
$$x_t = \rho x_{t-1} + \epsilon_t \quad \text{with } |\rho| < 1, \ \rho \neq 0.$$

We have $\beta_K \leq \beta_{K'}$ for any $K < K'$.

A. True

B. False

3. Let (R_t) be a time series of a risky return which covaries positively with tomorrow's consumption. If the consumption based asset pricing model is true, then we should not reject the null hypothesis of positive risk premium $\mathbb{E}_t[R_{t+1} - R^f] > 0$ where R^f is the risk free rate. Hint: The model yields

$$\frac{\mathbb{E}_t[R_{t+1} - R^f]}{R^f} = -Cov_t\left(R_{t+1}, \beta\frac{U'(C_{t+1})}{U'(C_t)}\right).$$

A. True

B. False

4. Consider the following predictive regression model

$$y_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t,$$
$$x_t = \rho x_{t-1} + u_t,$$

where $|\rho| < 1$. Assume that ϵ and u are jointly normal satisfying $\mathbb{E}[\epsilon_t|x_{t-1}] = 0$, $\mathbb{E}[\epsilon_t^2|x_{t-1}] = \sigma_\epsilon^2$, $\mathbb{E}[u_t^2] = \sigma_u^2$ and $\mathbb{E}[\epsilon_t u_t] = 0$. Which statement is false?

A. The OLS estimate $\hat{\beta}_1$ is consistent.

B. The OLS estimate $\hat{\beta}_1$ is biased.

C. The t-statistic for $\hat{\beta}_1$ converges to standard normal distribution under $\beta_1 = 0$.

D. $\mathbb{E}[x_t^2] = \sigma_u^2/(1-\rho^2)$.

5. Suppose that a set of time series $(y_t, x_t, z_t)_{t=1}^T$ is given by

$$\begin{aligned} y_t &= \alpha + \beta_t x_t + \gamma z_t + \eta_t \\ x_t &= \rho x_{t-1} + u_t, \\ \beta_t &= \beta + v_t, \end{aligned} \tag{1}$$

where x_t, z_t, η_t and v_t are independent each other. The model (1) has a time varying coefficient β_t. But I ignore the time varying β_t, and consider the OLS estimate of the following regression

$$y_t = b_0 + b_1 x_t + b_2 z_t + \epsilon_t.$$

Which statement is false?

A. $\hat{b}_0$ converges to α.

B. $\hat{b}_1$ converges to β.

C. $\hat{b}_2$ converges to γ.

D. The White test statistic LM converges to χ^2 distribution for some degrees of freedom.

6. The CAPM is estimated in equation for monthly returns to three United States stocks and gold for the period April 1990 to July 2004.

Stock	β_0	β_1	$\bar{R}^2$	LM(1)	LM(2)	White
Gold	-0.003 (0.238)	-0.098 (0.066)	0.014	1.452 (0.228)	7.530 (0.023)	2.579 (0.275)
Exxon	0.012 (0.000)	0.502 (0.000)	0.235	0.567 (0.452)	1.115 (0.573)	1.022 (0.600)
GE	0.016 (0.000)	1.144 (0.000)	0.440	5.458 (0.019)	7.014 (0.030)	5.336 (0.069)
MS	0.012 (0.069)	1.447 (0.000)	0.333	3.250 (0.071)	6.134 (0.047)	0.197 (0.906)

The test statistics are the LM test for jth order autocorrelation, and White test of heteroskedasticity. Note that p-values are in parenthesis. Which statement is false?

A. Only GE is rejected in the White test at 10% level.

B. Gold has the highest proportion of risk that is diversifiable.

C. For GE, the ratio of the idiosyncratic risk to the total risk is 0.440.

D. The gold and MS CAPMs exhibit second order autocorrelation, but not first autocorrelation at 5% level.

Topic 4. PROD: Production Econometrics and Cost Functions Estimation (Norbert Ladoux)

1. It is generally better to estimate a flexible form of the cost function instead of a flexible form of the production function because it is impossible to derive the input demand functions corresponding to a flexible production function.
   A. True
   B. False

Table 1: OLS and IZEF estimates of a TL form

Parameter	OLS	IZEF	Parameter	OLS	IZEF	Parameter	OLS	IZEF
a_K	0.279	0.057	g_{LL}	0.101	0.075	g_{EM}	−0.013	−0.004
	(0.035)	(0.001)		(0.035)	(0.007)		(0.008)	(0.009)
g_{KK}	0.045	0.030	g_{LE}	0.041	−0.004	g_{EY}	−0.031	0.000
	(0.004)	(0.006)		(0.034)	(0.002)		(0.003)	
g_{KL}	0.031	−0.000	g_{LM}	−0.123	−0.071	a_M	0.119	0.645
	(0.011)	(0.004)		(0.048)	(0.011)		(0.127)	(0.003)
g_{KE}	0.000	−0.010	g_{LY}	−0.028	0.000	g_{MK}	−0.062	−0.019
	(0.010)	(0.003)		(0.021)			(0.013)	(0.010)
g_{KM}	−0.015	−0.019	a_E	0.205	0.044	g_{ML}	−0.161	−0.071
	(0.015)	(0.010)		(0.018)	(0.001)		(0.039)	(0.011)
g_{KY}	−0.043	0.000	g_{EK}	−0.004	−0.010	g_{ME}	−0.053	−0.004
	(0.007)			(0.002)	(0.003)		(0.038)	(0.009)
a_L	0.398	0.253	g_{EL}	0.029	−0.004	g_{MM}	0.150	0.094
	(0.114)	(0.002)		(0.005)	(0.002)		(0.054)	(0.023)
g_{LK}	0.021	−0.000	g_{EE}	0.011	0.019	g_{MY}	0.102	0.000
	(0.012)	(0.004)		(0.005)	(0.005)		(0.024)	

2. In table 1 here above, we can deduce with certainty that IZEF estimates of the corresponding Translog cost function is such that one of the Allen elasticity is significantly equal to 1.
   A. True
   B. False

3. In table 1 here above, we can deduce with certainty that IZEF estimates of the corresponding Translog cost function is such that the translog cost function is not concave.
   A. True
   B. False

4. Consider that the Zellner method applied to a Translog form has given the following estimates of the own-price and Allen elasticities of substitution

$$\begin{aligned}\varepsilon_{KK} &= -0.388,\ \varepsilon_{LL} = -0.450,\ \varepsilon_{EE} = -0.536,\ \varepsilon_{MM} = -0.223\\ \sigma_{KL} &= 0.975, \sigma_{KE} = 3.285,\ \sigma_{KM} = 0.428,\ \sigma_{LE} = 0.640,\ \sigma_{LM} = 0.589,\ \sigma_{EM} = 0.853\end{aligned}$$

   A. Those results are compatible with the underlying theoretical model.
   B. Those results are not compatible with the underlying theoretical model.
   C. One of the Allen elasticities should be negative.
   D. The sum of own-price elasticities should be zero.

5. y is production level, C is total cost, p_i $(i = K, L, E, M)$ are the input prices of respectively capital, labor energy and materials, β_{ij} the parameters of the model. Consider the following unit cost function:

$$\begin{aligned}\frac{C}{y} &= \beta_{KK}p_K + \beta_{KL}(p_K p_L)^{1/2} + \beta_{KE}(p_K p_E)^{1/2} + \beta_{KM}(p_K p_M)^{1/2}\\ &\quad +\beta_{LK}(p_L p_K)^{1/2} + \beta_{LL}p_L + \beta_{LE}(p_L p_E)^{1/2} + \beta_{LM}(p_L p_M)^{1/2}\\ &\quad +\beta_{EK}(p_E p_K)^{1/2} + \beta_{EL}(p_E p_L)^{1/2} + \beta_{EE}p_E + \beta_{EM}(p_E p_M)^{1/2}\\ &\quad +\beta_{MK}(p_M p_K)^{1/2} + \beta_{ML}(p_M p_L)^{1/2} + \beta_{ME}(p_M p_E)^{1/2} + \beta_{MM}p_M\end{aligned}$$

   where $\beta_{ij} = \beta_{ji}\ \forall i, j = K, L, E, M$.
   A. A direct estimation of this cost function generally gives good results.
   B. There is no reason to estimate simultaneously the input demand functions even if contemporaneous residuals are correlated.
   C. It is absolutely necessary to consider a simultaneous estimation of the cost function and the input demand functions.
   D. The own-price elasticities resulting from the estimation of this cost function will not be necessarily negative.

6. Consider the following system of equations :

$$\begin{aligned}\frac{K}{y} &= \beta_{KK} + \beta_{KL}\left(\frac{p_L}{p_K}\right)^{1/2} + \beta_{KE}\left(\frac{p_E}{p_K}\right)^{1/2} + \beta_{KM}\left(\frac{p_M}{p_K}\right)^{1/2} + u_K\\ \frac{L}{y} &= \beta_{LL} + \beta_{KL}\left(\frac{p_K}{p_L}\right)^{1/2} + \beta_{LE}\left(\frac{p_E}{p_L}\right)^{1/2} + \beta_{LM}\left(\frac{p_M}{p_L}\right)^{1/2} + u_L\\ \frac{E}{y} &= \beta_{EE} + \beta_{KE}\left(\frac{p_K}{p_E}\right)^{1/2} + \beta_{LE}\left(\frac{p_L}{p_E}\right)^{1/2} + \beta_{EM}\left(\frac{p_M}{p_E}\right)^{1/2} + u_E\\ \frac{M}{y} &= \beta_{MM} + \beta_{KM}\left(\frac{p_K}{p_M}\right)^{1/2} + \beta_{LM}\left(\frac{p_L}{p_M}\right)^{1/2} + \beta_{EM}\left(\frac{p_E}{p_M}\right)^{1/2} + u_M\end{aligned}$$

   You are using a large sample to estimate this system of equations and you impose the symmetry constraints. If contemporaneous residuals only are correlated, you choose to estimate the model by
   A. using the Zellner method applied to 3 of the 4 equations of the system.
   B. using the Zellner method applied to the 4 equations of the system.
   C. using the OLS method.
   D. using limited information maximum likelihood method.

Topic 5. FAM: Economics of the Family (Nicolas Pistolesi)

1. In their paper "Changes in the labor supply behavior of married women", Blau and Kahn show that female labor supply has become less elastic relative to own and spouse wages over the period 1980-2000.
    A. True
    B. False

2. In their paper "The effect of child support payments on the labor supply and female family heads", Graham and Beller estimate that AFDC has the same effect as other non-wage income on female labor supply.
    A. True
    B. False

3. In their paper "Children and their parent's labor supply: Evidence from exogenous variation in family size", Angrist and Evans show that families with two children of the same sex are more likely to have a third one.
    A. True
    B. False

4. In their article, Blau and Kahn correct for non-random sample selection since
    A. The sample of women is not representative of the sample of men.
    B. The sample of working women is not representative of the population of women.
    C. Men and women are very different.
    D. The sample of men is not representative of the population of men.

5. In their paper, Graham and Beller
    A. estimate the effect of the different non-wage income on female labor supply.
    B. test if female labor supply depends on family size.
    C. test if child support is an efficient policy.
    D. test if women with dependent children are better off single than married.

6. In their article, Angrist and Evans show that
    A. Parents' labor supply depends on the sex of their children.
    B. Male and female labor supply is affected by the sex composition of their children.
    C. Women labor supply is affected by the sex composition of their children.
    D. In families with many children women tend to work more.

Topic 6. EDUC: Education Economics (François Poinas)

For the first three questions, we consider an ordered probit model. y denotes the observed dependent variable, taking values $\{1, 2, 3\}$. y^* is a latent variable. We assume $y_i^* = \beta_1 x_i + \beta_2 z_i + \varepsilon_i$, where x and z are observed variables and ε is an error term such that $\varepsilon \sim \mathcal{N}(0, 1)$. The relation between y and y^* is given by: $y = 1$ if $y^* < \gamma_1$; $y = 2$ if $\gamma_1 < y^* < \gamma_2$ and $y = 3$ if $\gamma_2 < y^*$, where γ_1 and γ_2 are cutoff values.

1. β_1 is the marginal effect of the variable x on the probability that y is equal to 1.
   A. True
   B. False

2. The probability that individual i chooses $y = 2$ is given by $\Phi(\gamma_2 - \beta_1 x_i - \beta_2 z_i) - \Phi(\gamma_1 - \beta_1 x_i - \beta_2 z_i)$ where $\Phi(\cdot)$ denotes the standard Normal cumulative distribution function.
   A. True
   B. False

3. An econometrician wants to explain the impact of the distance to work and the family income on the transportation mode chosen by workers. The econometrician has access to a cross-section sample of workers for whom he observes the following variables:
   - *transport*: transport mode used to go to work. This variable takes the value 1 if the individual uses a car, 2 if the individual uses the train and 3 otherwise,
   - *distance*: distance to work (in kilometers),
   - *income*: family income (in thousands of euros).

   The ordered probit model described above, where *transport* is the dependent variable (y) and *distance* and *income* are the explanatory variables (x and z), is appropriate for the analysis
   A. True
   B. False

4. An econometrician wants to estimate the causal impact of class size on schooling attainments. He uses a sample of pupils registered in a school. He estimates the following equation by OLS:
$$grade_i = \beta_1 + \beta_2\, class\, size_i + \boldsymbol{x}_i'\boldsymbol{\gamma} + \varepsilon_i$$
   where *grade* is the yearly average grade of pupil i, *class size* is the number of pupils registered in the class of pupil i and $\boldsymbol{x}$ is a vector of control variables including gender, age and parental education.

   Which of the following statements is a reason to suspect that the OLS estimator of β_2 will underestimate the causal impact of class size on the grade?
   A. More able pupils are selected in larger classes.
   B. More able pupils are selected in smaller classes.
   C. Girls are selected in larger classes.
   D. Girls are selected in smaller classes.

5. In the article of Bandiera, Larcinese, Rasul (2010)[1], what is the empirical strategy adopted to estimate the causal impact of class size on the grade?

   A. The authors use the fact that some students are assigned to classes of different sizes based on their birth date. They use the birth date as an instrumental variable.

   B. The authors run a randomized field experiment by assigning randomly students to small and large classes. The random assignment permits to construct groups of students who share the same characteristics, except the size of the class they are assigned to.

   C. The authors exploit an administrative rule of the university which stipulates that, for a given course, a class should not contain more than 50 students. They estimate a causal impact by comparing the grades obtained by students registered in courses with slightly less than 50 students and the grades obtained in courses with slightly more than 50 students, as students are split into 2 classes of around 25 students in this later case.

   D. The authors exploit the fact that the same student is registered in different courses that differ with respect to the number of students registered. They use a fixed effect panel data model to control for time-unvarying characteristics that stay stable across courses.

6. Bandiera, Larcinese, Rasul (2010) estimate the following quantile regression model:

$$Quant_\theta(y_{ic}|\cdot) = \gamma_\theta N_c + \delta_\theta X_c$$

   where y_{ic} is the exam grade of student i in course c, $Quant_\theta$ denotes quantile θ, N_c is the class size, X_c is a set of course characteristics. The results of the quantile regression performed at each percentile are represented in Figure 1 below. What is the empirical evidence suggested by the figure?

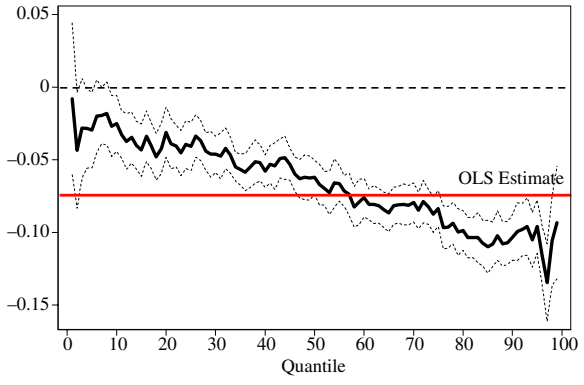
   A. The impact of class size is negative and larger (in absolute value) for students having higher grades.

   B. The impact of class size is negative and lower (in absolute value) for students having higher grades.

   C. The impact of class size is positive and larger (in absolute value) for students having higher grades.

   D. The impact of class size is positive and lower (in absolute value) for students having higher grades.

Figure 1: Quantile Regression Estimates of the Effect Size



Note: The figure shows the estimated effect size on the test score at each percentile of the distribution of student test scores an the associated 95% confidence interval.

[1] Bandiera, O., Larcinese, V., Rasul, I., 2010, "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students", *The Economic Journal*, vol. 120 (December), pp. 1365-1398

ANSWERS SHEET

Stick your 3rd barcode sticker here

Please, indicate for each column the label among the 3 topics chosen (ex: "IO", "FAM", etc.) and report your answers in the corresponding column of the following table. Then, include this sheet (and this sheet only) in the exam double sheet.

WARNING: IF, FOR A GIVEN COLUMN, THE NAME OF THE TOPIC IS NOT INDICATED, ANSWERS WILL NOT BE TAKEN INTO ACCOUNT AND NO MARK WILL BE GIVEN FOR THE TOPIC.

	Topic 1: _____	**Topic 2:** _____	**Topic 3:** _____
Question 1			
Question 2			
Question 3			
Question 4			
Question 5			
Question 6			