- *Suggested time for this exam: 1 hour.*
- *Use the results from the course.*
- *You can use the course notations without recalling their definitions. If you have to use a new shortcut notation, give a definition.*
- **(SA)** *indicates that a short answer (typically one or two sentences, or one formula) is sufficient.*
- *Total number of marks = 20*

EXERCICE 1 (10 marks). We want to determine if the subscription to a public health insurance affects the number of visits to the doctor. To address this question, we use a sample of 3874 German individuals, aged between 25 and 64 and observed in 1984. We assume that the number of visits to the doctor in the last three months ($y$) conditional on the explanatory factors ($x$) has a Poisson distribution and that $\mathrm{E}(y|x) = \exp(x\beta)$. Therefore, the conditional density of $y$ given $x$ is given by

$$f(y|x,\beta) = \frac{\exp(-\exp(x'\beta))[\exp(x'\beta)]^y}{y!}.$$

The explanatory factors we consider are the following: the age of the individual, denoted **age**, the number of years of schooling (**educ**), the household income (in thousands of marks, the money of Germany in 1984), denoted **income**, the number of children younger than 16 in the household (**kids**) and an indicator variable equal to 1 if the individual has a public health insurance (**public**). The results are reported in Column (1) of Table 1.

1. [1 mark] **(SA)** Write the log-likelihood function of the sample.
2. [3 marks] Test if the impact of holding a public health insurance on the number of visits to the hospital is significant at the 5% level. You should write the null and alternative hypothesis, give the form of the test statistic, its behavior under the null hypothesis, the formal decision rule, and your conclusion based on the above results.
3. [3 marks] Calculate the estimated predicted probability of having 3 visits to the doctor for a 20 years old person, who has 10 years of education, with no children, holds a public health insurance and whose income is 5000 marks.
4. [3 marks] A new specification is estimated, where a quadratic term in age (denoted **age2**) and a quadratic term in education (denoted **educ2**) have been added to the list of covariates. Results are given in Column (2) of Table 1.

   Perform a LR test, at the 5% level, to test if these two variables are jointly significant. You should write the null and alternative hypothesis, give the form of the test statistic, its behavior under the null hypothesis, the formal decision rule, and your conclusion based on the above results.

EXERCICE 2 (10 marks). Suppose that the true wage equation is given by,

$$LW_i = \beta_1 + \beta_2 S_i + \beta_3 A_i + \varepsilon_i,$$

where $LW_i$, $S_i$, $A_i$ are log wage, completed years of schooling, and a measure of ability, respectively, for individual $i$. Let $S$ be an exogenous variable.

1. [1 mark] First suppose that we have no good measure of ability. Instead, we just estimate

   $$LW_i = \beta_1 + \beta_2 S_i + u_i. \tag{1}$$

   The OLS estimator of $\beta_2$ is given by

   $$\widehat{\beta}_2 = \frac{\widehat{\mathrm{cov}}(LW_i, S_i)}{\widehat{\mathrm{var}}(S_i)}.$$

   What is the sign of the bias of $\widehat{\beta}_2$? Give reasons in support of your claim.
2. [1 mark] Now suppose we find a measure of ability in some IQ test scores. Assuming that $IQ$ is a perfect measure of ability, $A$, we run the following OLS regression

   $$LW_i = \beta_1 + \beta_2 S_i + \beta_3 IQ_i + v_i, \tag{2}$$

   The results of OLS regressions for (1) and (2) is given in Table 2 below in lines 1 and 2, respectively. Standard deviation of schooling, $\sigma_S$ is 1.6 in the sample. Compute the estimated covariance between schooling $S$ and ability $A$.

3. [1 mark] In regression (2), we now suspect that $IQ$ is a noisy measure of $A$ making it potentially endogenous. We also use 2SLS to estimate the parameters in equation (2) and these are reported in line 3 of table 2. **The figures in parentheses are t-statistics rather than standard errors.** Compute the standard errors of $\widehat{\beta}_{3,OLS}$ and $\widehat{\beta}_{3,2SLS}$ (rounded to four decimal points).

4. [4 marks] How will you test the claim that $IQ$ is exogenous or endogenous in regression (2)? Write the null and alternative hypothesis, give the form of the test statistic, its behavior under the null hypothesis, and the formal decision rule. What is your conclusion in this particular application at a 1% and 5% significance level?

5. [3 marks] Is the effect of ability on log wage significantly positive? Construct an asymptotic test at 1% significance level using Regression 3 results.

Table 1: Estimation Results for Exercise 1

| Dependent variable: Doctor visits | (1) | (2) |
|---|---|---|
| _intercept | 0.4329 | 0.0122 |
| | (0.0921) | (0.2868) |
| age | 0.0228 | 0.0251 |
| | (0.0009) | (0.0073) |
| educ | -0.0238 | 0.0382 |
| | (0.0050) | (0.0372) |
| income | -0.0922 | -0.0931 |
| | (0.0075) | (0.0077) |
| kids | -0.1149 | -0.1165 |
| | (0.0213) | (0.0218) |
| public | 0.2672 | 0.2617 |
| | (0.0388) | (0.0389) |
| age2 | | $-0.2338 \cdot 10^{-4}$ |
| | | $(0.8046 \cdot 10^{-4})$ |
| educ2 | | $-24.23 \cdot 10^{-4}$ |
| | | $(14.44 \cdot 10^{-4})$ |
| $N$ | 3874 | 3874 |
| Sample log-likelihood | -15675.57 | -15674.115 |

Note: Standard errors in parentheses, Income in thousands of marks.

Table 2: Parameter Estimates for Exercise 2.
**Figures in parentheses are t-statistics rather than standard errors.**

| Line no. | Estimation Method | Coefficient of $S$ | Coefficient of $IQ$ |
|---|---|---|---|
| 1 | OLS | 0.08 (10.2) | – |
| 2 | OLS | 0.07 (9.4) | 0.002 (2.6) |
| 3 | 2SLS | 0.05 (6.0) | 0.004 (2.4) |

Quantiles $x$ of order $\beta$ from a $\mathcal{N}(0,1)$, defined as $\Phi(x) = \beta$.

| $\beta$ | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 | 0.99995 |
|---|---|---|---|---|---|---|---|---|
| $x$ | 1.2816 | 1.6449 | 1.96 | 2.3263 | 2.5758 | 3.0902 | 3.2905 | 3.8906 |

Quantiles of order $\beta$ of a $\chi^2$ with $p$ degrees of freedom such that $\Pr(X \leq x) = \beta$ with $X \sim \chi^2_p$.

| $p$ \ $\beta$ | 0.01 | 0.05 | 0.1 | 0.9 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.00 | 0.02 | 2.71 | 3.84 | 6.63 |
| 2 | 0.02 | 0.10 | 0.21 | 4.61 | 5.99 | 9.21 |
| 3 | 0.11 | 0.35 | 0.58 | 6.25 | 7.81 | 11.34 |
| 4 | 0.30 | 0.71 | 1.06 | 7.78 | 9.49 | 13.28 |
| 5 | 0.55 | 1.15 | 1.61 | 9.24 | 11.07 | 15.09 |

Intermediate Econometrics

— *Temps suggéré : 1 heure.*
— *Utilisez les résultats connus du cours.*
— *Vous pouvez utiliser les notations du cours sans rappeler leur signification. Si vous introduisez de nouvelles notations, donnez leur signification.*
— **(SA)** *indique qu'une réponse courte (une-deux phrases ou formules) est suffisante.*
— *Le total des points est de 20.*

EXERCICE 1 (10 points). Vous étudiez l'effet d'une assurance santé (publique) sur le nombre de visites chez le docteur. Vous avez un échantillon de 3874 résidents allemands âgés de 25 à 64 ans pour l'année 1984. Vous supposez que le nombre de visites chez le docteur dans le trimestre, $y$, est distribué suivant une loi de Poisson conditionnellement à des variables explicatives, $\boldsymbol{x}$, de sorte que $\mathrm{E}(y|\boldsymbol{x}) = \exp(\boldsymbol{x\beta})$. La distribution conditionnelle de $y$ sachant $\boldsymbol{x}$ est

$$f(y|\boldsymbol{x}, \boldsymbol{\beta}) = \frac{\exp(-\exp(\boldsymbol{x}'\boldsymbol{\beta}))[\exp(\boldsymbol{x}'\boldsymbol{\beta})]^y}{y!}.$$

Les variables explicatives sont : l'**age**, le niveau d'éducation en années (**educ**), le revenu annuel du ménage **income** (en milliers de marks, monnaie en vigueur en Allemagne en 1984), le nombre d'enfants de moins de 16 ans dans le foyer (**kids**), une variable dummy égale à 1 si l'individu a souscrit une assurance santé publique (**public**). Les résultats sont donnés dans la colonne (2) de la Table 1.

1. [1 mark] **(SA)** Ecrivez la vraisemblance de l'échantillon.
2. [3 marks] Testez si l'impact de l'assurance santé est siginificatif au niveau 5%. Ecrivez les hypothèses nulle et alternative, donnez la forme de la statistique de test, son comportement sous l'hypothèse nulle et la règle de décision formelle. Quelle est votre conclusion ?
3. [3 marks] Quel est la probabilité estimée de faire 3 visites chez le docteur pour un individu de 20 ans avec un niveau d'éducation de 10 ans, sans enfants, qui a une assurance santé, et dont le revenu est de 5000 marks.
4. [3 marks] Vous estimez un nouveau modèle en rajoutant un terme quadratique dans l'age (**age2**) et un terme quadratique dans l'éducation (**educ2**). Les résultats sont donnés dans la colonne (1) de la Table 1.
   Faites un test LR au niveau 5% pour tester la significativité jointe de ces deux nouvelles variables. Ecrivez les hypothèses nulle et alternative, donnez la forme de la statistique de test, son comportement sous l'hypothèse nulle et la règle de décision formelle. Quelle est votre conclusion ?

EXERCICE 2. [10 points] On suppose que la vraie équation de salaire est donnée par

$$LW_i = \beta_1 + \beta_2 S_i + \beta_3 A_i + \varepsilon_i,$$

où $LW_i, S_i, A_i$ sont, respectivement, le log du salaire, le nombre d'années d'éducation acquises, et une mesure d'aptitude (*ability*), pour l'individu $i$. On suppose que $S$ est une variable exogène.

1. [1 mark] D'abord, supposons que l'on ne dispose pas d'une bonne mesure de l'aptitude. On estime donc l'équation
   $$LW_i = \beta_1 + \beta_2 S_i + u_i. \tag{1}$$
   L'estimateur des MCO de $\beta_2$ est donné par
   $$\widehat{\beta}_2 = \frac{\widehat{\mathrm{cov}}(LW_i, S_i)}{\widehat{\mathrm{var}}(S_i)}.$$
   Quel est le signe du biais de $\widehat{\beta}_2$ ? Justifiez votre réponse.
2. [1 mark] A présent, supposons que l'on obtienne une mesure de l'aptitude à partir de résultats obtenus à un test de QI. En supposant que le QI est une mesure parfaite de l'aptitude, $A$, on estime l'équation suivante par MCO
   $$LW_i = \beta_1 + \beta_2 S_i + \beta_3 QI_i + v_i, \tag{2}$$
   Les résultats des régressions par MCO pour (1) et (2) sont donnés dans le Tableau 2 ci-dessous, aux lignes 1 et 2 respectivement. L'écart-type du nombre d'années d'éducation dans l'échantillon est de 1.6. Calculez la covariance estimée entre le nombre d'années d'éducation $S$ et l'aptitude $A$.
3. [1 mark] Dans la régression (2), on suspecte à présent que $QI$ est une mesure imparfaite de $A$, ce qui rend $QI$ potentiellement endogène. Nous utilisons les doubles moindres carrés (2MCO) pour estimer les paramètres de l'équation (2). Les paramètres estimés sont présentés à la ligne 3 du Tableau 2. **Les nombres entre parenthèse sont les t-statistiques, et non les erreurs standard.** Calculez les erreurs standard de $\widehat{\beta}_{3,MCO}$ et $\widehat{\beta}_{3,2MCO}$.

4. [4 marks] Comment testeriez-vous que $QI$ est exogène ou endogène dans la régression (2) ? Ecrivez les hypothèses nulle et alternative, donnez la forme de la statistique de test, son comportement sous l'hypothèse nulle et la règle de décision formelle. Quelle est votre conclusion dans cet exercice pour des niveaux de significativité de 1 % et 5 % ?
5. [3 marks] L'effet de l'aptitude sur le salaire est-il significativement positif ? Construisez un test asymptotique de niveau 1% en utilisant les résultats de la régression 3.

TABLE 1 – Résultats d'estimation pour l'exercice 1

| Dependent variable : Doctor visits | (1) | (2) |
|---|---|---|
| _intercept | 0.4329 | 0.0122 |
| | (0.0921) | (0.2868) |
| age | 0.0228 | 0.0251 |
| | (0.0009) | (0.0073) |
| educ | -0.0238 | 0.0382 |
| | (0.0050) | (0.0372) |
| income | -0.0922 | -0.0931 |
| | (0.0075) | (0.0077) |
| kids | -0.1149 | -0.1165 |
| | (0.0213) | (0.0218) |
| public | 0.2672 | 0.2617 |
| | (0.0388) | (0.0389) |
| age2 | | $-0.2338 \cdot 10^{-4}$ |
| | | $(0.8046 \cdot 10^{-4})$ |
| educ2 | | $-24.23 \cdot 10^{-4}$ |
| | | $(14.44 \cdot 10^{-4})$ |
| $N$ | 3874 | 3874 |
| Sample log-likelihood | -15675.57 | -15674.115 |

Note : Erreurs standards entre parenthèses, Revenu en milliers de marks.

TABLE 2 – Résultats d'estimation pour l'exercice 2

**Entre parenthèses les t-statistics, pas les erreurs standards.**

| Line no. | Estimation Method | Coefficient of $S$ | Coefficient of $IQ$ |
|---|---|---|---|
| 1 | OLS | 0.08<br>(10.2) | – |
| 2 | OLS | 0.07<br>(9.4) | 0.002<br>(2.6) |
| 3 | 2SLS | 0.05<br>(6.0) | 0.004<br>(2.4) |

Quantiles $x$ d'ordre $\beta$ pour une $\mathcal{N}(0,1)$, definis comme $\Phi(x) = \beta$.

| $\beta$ | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 | 0.99995 |
|---|---|---|---|---|---|---|---|---|
| $x$ | 1.2816 | 1.6449 | 1.96 | 2.3263 | 2.5758 | 3.0902 | 3.2905 | 3.8906 |

Quantiles d'ordre $\beta$ d'une $\chi^2$ à $p$ degrés de liberté tels que $\Pr(X \leq x) = \beta$ avec $X \sim \chi^2_p$.

| $p$ | $\beta$ | | | | | |
|---|---|---|---|---|---|---|
| | 0.01 | 0.05 | 0.1 | 0.9 | 0.95 | 0.99 |
| 1 | 0.00 | 0.00 | 0.02 | 2.71 | 3.84 | 6.63 |
| 2 | 0.02 | 0.10 | 0.21 | 4.61 | 5.99 | 9.21 |
| 3 | 0.11 | 0.35 | 0.58 | 6.25 | 7.81 | 11.34 |
| 4 | 0.30 | 0.71 | 1.06 | 7.78 | 9.49 | 13.28 |
| 5 | 0.55 | 1.15 | 1.61 | 9.24 | 11.07 | 15.09 |