

Econométrie 2, 2018-2019

Consignes et sujets

Jérémy L'Hour
assistant-econometrie@ensae.fr

March 15, 2019

1. Consignes

Mode de fonctionnement, attribution, notation. Le projet est à effectuer et rendre par binôme d'étudiants d'un même groupe de TD. Lorsque les groupes sont de taille impaire, et une fois le plus possible de binômes formés, il est possible de former un groupe de trois étudiants.

Il vous est demandé d'envoyer le numéro de vos trois sujets préférés à votre chargé de TD pour le lundi 18 mars à 18h. L'attribution des sujets répond aux contraintes suivantes: (i) chaque sujet doit être pris une fois dans un groupe de TD, (ii) un sujet peut être pris deux fois au plus dans un groupe de TD. C'est à chaque chargé de TD de déterminer la répartition des sujets, en essayant de répondre aux préférences des élèves. C'est le concepteur du sujet qui va corriger tous les projets réalisés à partir de son sujet. La triche et le plagiat seront sévèrement sanctionnés.

Date de rendu. Vendredi 17 mai, 18h: **réponses rédigées ET code.**

Notation. Le sujet est noté sur 20 points. Le langage de programmation est au choix parmi R, STATA, SAS ou Python.

Liste des sujets.

Num.	Intitulé	Rendre à
1	Position sur le marché du travail et inégalités liées à l'origine	lucas.girard @ensae.fr
2	Influence d'une chaîne de TV conservatrice sur le vote	jeremy.l.hour @ensae.fr
3	Impact de l'arrestation sur la probabilité de récidive	jeremy.l.hour @ensae.fr
4	La beauté et la couleur de peau en politique	o.couperier @gmail.com
5	Impact du salaire minimum dans la restauration rapide	yannick.guyonvarch @ensae.fr
6	Estimation des rendements privés et sociaux de l'enseignement supérieur	elio.nimier-david @ensae.fr
7	Création d'un indice de prix pour l'immobilier commercial	roxane.morel @developpement-durable.gouv.fr
8	Fécondité et participation des femmes au marché du travail en Afrique sub-saharienne	raphael.sh.lee @gmail.com

2. Pièges à éviter

2.1. Utilisation des méthodes

- Tout résultat d'estimation doit faire figurer un écart-type dont on précisera le modèle de calcul (niveau de cluster, par exemple).
- La matrice de corrélation des régresseurs ne permet PAS d'étudier l'hypothèse de non-corrélation entre régresseurs et résidus. La conclusion "il y a peu de corrélation entre régresseurs donc on peut penser qu'il n'y a pas de gros problème d'endogénéité" est fausse. L'hypothèse d'exogénéité est difficilement testable et repose sur une justification soigneuse avant toute estimation.
- En revanche, si les variables sont très corrélées, les estimateurs peuvent être très peu précis.
- Le fait que les X soient corrélés avec le Y ne permet pas de conclure à l'existence d'endogénéité (corrélation entre X et ε).
- Quand vous faites un Probit ou un Logit, les coefficients ne s'interprètent pas directement. Il faut passer par les effets marginaux.
- Attention à bien calculer les effets marginaux quand une variable et sa transformation sont dans le modèle. Attention par exemple pour les calculs d'effets marginaux avec age et age au carré où age au carré apparaît car il a été calculé à la main puis intégré dans le modèle.
- Les coefficients dans les modèles de type Logit ou Probit ordonnés avec seuils connus ont la même interprétation que dans une régression linéaire classique (c'est simplement une régression par intervalles due au fait que l'information n'est que partielle).
- Faites toujours figurer les écarts-types dans vos tableaux de résultats, les étoiles ne sont pas suffisantes.

- L’utilisation d’un modèle binomial de type Logit ou Probit ne se justifie PAS pas le fait que l’un des X soit binaire mais par le fait que Y soit binaire.
- Y compris entre 0 et 1 ne justifie par l’utilisation du modèle binomial : il faut que Y vaille SOIT 0, SOIT 1.
- Dans les modèles polytomiques non ordonnés, il y a toujours une référence. Par exemple dans un modèle de choix de transport, si c’est la voiture personnelle qui aurait été normalisée, alors les coefficients associés aux autres modalités s’interpréteront par rapport à l’alternative “voiture personnelle”.
- L’hypothèse IANP (Indépendance aux Alternatives Non Pertinentes) est rarement mentionnée et encore moins discutée dans les modèles de type “multinomial Logit”. Il est important de le faire.
- Adopter le réflexe d’écrire le modèle et ses hypothèses. La vraisemblance aussi, si c’est la méthode choisie.
- Attention au traitement des valeurs manquantes dans vos données. Elles sont parfois codées à 0 par inattention ce qui peut poser problème dans la suite de vos estimations. Une justification de la méthode de traitement est nécessaire (e.g. suppression, remplacement par la valeur la plus fréquentes, remplacement en utilisant un algorithme de type plus-proche-voisin). Eventuellement: est-ce que la prise en compte de la sélection est nécessaire dans le problème ?
- N’incluez pas des variables endogènes quand elles ne vous intéressent pas.

2.2. Présentation, problèmes de forme

- Utilisez le nom français de la quantité mesurée dans vos phrases: votre variable sera le nombre d’année d’études et non pas *nb_an_educ*.
- Faire un copier-coller de sortie STATA n’est pas acceptable : il a des packages qui permettent de sortir un beau tableau LaTeX sans aucun effort.

- Les graphiques doivent porter un message : réfléchissez à la meilleure façon de représenter ce message et rendez-le facilement compréhensible pour le lecteur. Les statistiques descriptives doivent déjà vous donner une idée des variables pertinentes à inclure et sont l’occasion de démontrer vos compétences en visualisation des données.
- Pensez au sens des phrases que vous écrivez. Personne ne passe de la catégorie “homme” à “femme” pour le sexe; on n’est pas “célibataire” puis tout à coup “divorcé”.
- Au même titre que vous devez faire attention à l’échelle de mesure de vos variables, vous devez aussi préciser les unités de mesure. On n’écrit pas “ un salaire de 3000 ” sans préciser que ce sont des euros.