Problem Set #2

ECO 8000
Automne 2025

Date de remise : 8-déc-2025. Veuillez soumettre votre travail par courriel à `sam.gyetvay@gmail.com` avec l'objet : ECO8000 + TP2 + votre nom de famille + numéro d'étudiant.

Rappel : La collaboration est permise (et encouragée), mais chaque étudiant doit rédiger et soumettre sa propre solution. Rappelez-vous que chaque travail pratique compte seulement pour 15% de votre note finale, tandis que la présentation + la participation + le projet final comptent pour 70%. Cela est intentionnel : l'objectif est que vous appreniez, pas simplement que vous obteniez la bonne réponse. Si vous obtenez la bonne réponse en copiant un autre étudiant ou ChatGPT, vous n'apprendrez rien.

Instructions : Je souhaite que vous rédigiez votre document en LaTeX. Un compilateur LaTeX en ligne populaire parmi les étudiants et les chercheurs est Overleaf (`www.overleaf.com`). C'est un bon exercice pour lorsque vous devrez écrire un article. Pour les questions qui nécessitent du code, veuillez inclure votre script (fichier STATA `.do`, script `.R`, ou autre). Ajoutez des commentaires dans votre code et indiquez quelle section correspond à chaque question afin que je puisse le suivre facilement.

Vous pouvez trouver ce document en ligne à `www.samgyetvay.com/teaching/eco8000/tp2.pdf`

Question 1 : Emplois différenciés (Card, Cardoso, Heining et Kline, 2018) (10 points) Dans cette question, nous explorerons quelques extensions simples du modèle étudié en classe. Une infinité continue de masse L de travailleurs i choisissent dans quelle entreprise $j \in \{1, ..., J\}$ travailler.

Question 1 (a) : (1 point) Supposons que l'utilité qu'un travailleur i retire en travaillant pour l'entreprise j est donnée par :

$$U_{ij} = \beta \ln w_j + a_j + \epsilon_{ij}$$

et que chaque travailleur choisit l'entreprise qui maximise son utilité. Interprétez β, a_j et ϵ_{ij}. Quels facteurs réels a_j et ϵ_{ij} pourraient-ils représenter ? Donnez des exemples précis. Dans ce modèle, les travailleurs choisissent-ils toujours l'entreprise qui paie le salaire le plus élevé ? Pourquoi ou pourquoi pas ?

Question 1 (b) : (1 point) Supposons que les ϵ_{ij} sont des valeurs extrêmes de Type I indépendantes et identiquement distribuées. C'est-à-dire, supposons que la fonction de répartition $F(\epsilon) = \exp(-\exp(-\epsilon))$.

On peut montrer[1] que si $U_{ij} = V_{ij} + \epsilon_{ij}$

$$P(U_{ij} > U_{ik} \ \forall \ j \neq k) = \frac{\exp(V_{ij})}{\sum_k \exp(V_{ik})}$$

En utilisant ce résultat, montrez que

$$\ln L_j = \ln \lambda + \beta \ln w_j + a_j$$

Qu'est-ce que λ ? Quelle est cette équation ? Que représente β ? Que se passe-t-il lorsque w_{ij} ou a_j augmentent ? Que se passe-t-il lorsque L est plus grand ? Que se passe-t-il lorsque les salaires des autres entreprises $(w_{ik})_{k \neq j}$ augmentent ?

Question 1 (c) : (1 point) Supposons que l'entreprise j produise une quantité Y_j en utilisant la fonction de production $Y_j = A_j L_j$, où A_j est la productivité spécifique de l'entreprise et constante. Les entreprises vendent leur production au prix p_j. L'entreprise j choisit w_j et L_j pour maximiser les profits $\pi_j = p_j Y_j - w_j L_j$ sous la contrainte d'offre de travail obtenue à la partie (b), en traitant λ et p_j comme constants. Écrivez le problème de maximisation du profit sous contrainte de l'entreprise à l'aide d'un lagrangien. Notez par μ le multiplicateur de Lagrange associé à la contrainte d'offre de travail. (*Indice* : il peut être plus simple d'écrire la contrainte d'offre de travail en termes de L_j.)

Question 1 (d) : (1 point) Calculez les conditions du premier ordre par rapport à w_j, L_j et μ, puis réarrangez pour montrer que le salaire optimal satisfait

$$w_j^* = p_j A_j \frac{\beta}{1 + \beta}$$

Interprétez cette équation. Quelle est la relation entre le salaire et le produit marginal du travail ? Les entreprises plus productives paient-elles des salaires plus élevés ? Que se passe-t-il lorsque β augmente ?

Question 2 : Simulations de données de panel. Dans cette question, vous allez simuler des données de panel, estimer des régressions en double différence, et tracer des graphiques d'étude d'événement (event-study). J'ai fourni du code utile pour cette question dans le cours sur la recherche :

`https://www.samgyetvay.com/teaching/eco8000/search.pdf`

Question 2 (a) : (2 points) Simulez un jeu de données de panel synthétique avec 100 individus

[1] Voir chapitre 3 de Train (2009) https://eml.berkeley.edu/books/choice2.html pour une démonstration. Au fait, le livre de Train est excellent et les premiers chapitres valent vraiment la peine d'être lus, particulièrement pour les doctorants.

(indexés par i) observés sur 21 périodes (t). Générez un indicateur binaire de traitement D_i égal à 1 pour les travailleurs traités et 0 pour les non-traités. Tirez D_i indépendamment selon $P(D_i = 1) = 0.5$. Générez un indicateur de traitement variant dans le temps D_{it} : pour les travailleurs traités, D_{it} passe de 0 à 1 à la période 10. Pour les non-traités, $D_{it} = 0$ à toutes les périodes. Générez des effets fixes individuels $\alpha_i \sim N(\mu_1, \sigma_\alpha)$ si $D_i = 1$ et $\alpha_i \sim N(\mu_0, \sigma_\alpha)$ si $D_i = 0$. Générez des effets fixes temporels $\tau_t = \sqrt{t} + \eta_t$ où $\eta_t \sim N(0, \sigma_\eta)$. Simulez les revenus Y_{it} selon l'équation suivante :

$$Y_{it} = \alpha_i + \tau_t + \beta D_{it} + \epsilon_{it}$$

Commencez avec les valeurs de paramètres $\beta = 5$, $\sigma_\alpha = 2$, $\sigma_\eta = 0.025$, $\mu_{alpha0} = 1$ et $\mu_{alpha1} = 3$. Tracez Y_{it} en fonction de t pour quelques travailleurs. Commentez la dynamique des revenus. Modifiez certaines valeurs de paramètres et tracez de nouveau quelques travailleurs. Discutez intuitivement le rôle de chaque paramètre. (*Indications* : fixez une graine aléatoire pour assurer la reproductibilité : utilisez `set seed 123` au début du script. Utilisez des globales pour les paramètres (p. ex. `global beta = 2`, `global sd_alpha = 2`) pour faciliter leur manipulation. Pour générer les effets individuels et temporels, utilisez l'astuce suivante : générez d'abord des tirages indépendants pour chaque ligne, puis utilisez `bysort` afin de sélectionner la première observation de chaque individu ou période :

```
gen alpha_i=rnormal($mu_alpha1, $sd_alpha) if D_i == 1
replace alpha_i=rnormal($mu_alpha0, $sd_alpha) if D_i == 0
bys id: replace alpha_i=alpha_i[1]
gen eta_t=rnormal(0,$sd_eta)
bys t: gen tau_t=sqrt(t)+eta_t[1])
```

Question 2 (b) : (2 points) Créez un graphique montrant le revenu moyen des travailleurs traités et non-traités au fil du temps. C'est-à-dire, tracez les estimés de $\bar{Y}_{1t} = E[Y_{it}|D_i = 1]$ et $\bar{Y}_{0t} = E[Y_{it}|D_i = 0]$ pour chaque t. (*Indication* : utilisez `collapse (mean) Ybar_t = Y_it, by(D_i t)` pour créer un jeu de données où chaque ligne correspond à une combinaison traitement-année. Pour conserver vos données simulées en mémoire, entourez le `collapse` avec `preserve` et `restore`.)

Question 2 (c) : (2 points) Générez une variable $Post_t$ égale à 1 si $t \geq 10$ et estimez la régression en double différence :

$$Y_{it} = \gamma_1 D_i + \gamma_2 Post_t + \delta D_i \times Post_t + \epsilon_{it}$$

Que représentent γ_1, γ_2 et δ ? Calculez la double différence manuellement en calculant

$$E[Y_{it}|D_i = 1, Post_t = 1] - E[Y_{it}|D_i = 1, Post_t = 0] - (E[Y_{it}|D_i = 0, Post_t = 1] - E[Y_{it}|D_i = 0, Post_t = 0])$$

(*Indication* : réutilisez `collapse`, mais cette fois `collapse (mean) Y_it, by(D_i Post_t)`.) Comparez cette valeur à l'estimation de δ provenant de la régression.

Question 2 (d) : (4 points) Nous allons maintenant simuler des données pour un choc de revenu avec des effets de traitement dynamiques. Simulez les données de panel comme auparavant, mais avant de générer Y_{it}, définissez le temps relatif $s = t - 10$, qui varie donc de -10 à $+10$. Que l'issue soit générée par

$$Y_{it} = \alpha_i + \tau_t + \sum_{s=-10}^{10} \beta_s \cdot D_i \cdot \mathbf{1}\{t - t_0 = s\} + \varepsilon_{it},$$

où les β_s sont calculés selon la fonction suivante :

$$\beta_s = \begin{cases} \kappa_{Pre} \cdot s, & s < 0, \\ \kappa_{Post} \cdot s, & s \geq 0, \end{cases}$$

avec $\kappa_{Pre} = 0$ et $\kappa_{Post} = 0.5$. Tracez le revenu moyen des travailleurs traités et non-traités comme à la partie (b). Commentez la dynamique des revenus. (*Indication* : pour générer les coefficients β_s, vous pouvez utiliser `gen beta_s = ($kappa_pre*s)*(s<0) + ($kappa_post*s)*(s>=0)`.)

Question 2 (e) : (4 points) Estimez la régression d'étude d'événement suivante avec effets fixes individuels et temporels :

$$Y_{it} = \alpha_i + \tau_t + \sum_{s=-10}^{10} \widehat{\beta}_s \cdot \mathbf{1}\{t - t_0 = s\} + u_{it},$$

et tracez les estimateurs ponctuels et intervalles de confiance à 95% pour $\widehat{\beta}_s$ pour $s \in [-10, 10]$ (en mettant le $s = -1$ omis à zéro sur l'axe horizontal). Y a-t-il des pré-tendances ? (*Indication* : vous pouvez estimer la régression en utilisant `reghdfe Y_it i.s##c.D_i, absorb(id tau_t))` suivie de `estimates store` et `coefplot es_reg, drop(_cons) xline(0) vertical ci`. Une autre option est d'utiliser le package `xtevent`, qui contient une commande permettant d'estimer des event-studies et de tracer les résultats directement. Enfin, vous pouvez construire le graphique manuellement avec

`twoway (rcap ci_upper ci_lower s, sort) (scatter beta_s s, sort)`.

Pour utiliser cette troisième option, vous devrez extraire les coefficients dans votre jeu de données.

Question 2 (f) : (4 points) Répétez les parties (d) et (e), mais cette fois fixez $\kappa_{Pre} = 0.2$ et $\kappa_{Post} = 0.8$. Commentez brièvement les résultats. Pouvez-vous détecter la pré-tendance ?

Question 3 : Card et Krueger (1994). Dans cette question, vous allez reproduire certaines figures et tableaux du célèbre article sur l'effet du salaire minimum sur l'emploi et les revenus. Pour commencer, téléchargez le jeu de données sur le site web de David Card à :

`https://davidcard.berkeley.edu/data_sets`

Avant de commencer la question, lisez attentivement l'article en entier si vous ne l'avez pas déjà fait. Vous pouvez trouver l'article ici :

`https://davidcard.berkeley.edu/papers/njmin-aer.pdf`

Question 3 (a) : Reproduisez la Figure 1 de l'article.

Question 3 (b) : Reproduisez les trois premières colonnes du Tableau 3 (Magasins par État, (i) PA, (ii) NJ et (iii) NJ-PA).

Question 3 (c) : Reproduisez le Tableau 4. Vous devez uniquement reproduire les quatre premières lignes (vous pouvez ignorer les lignes 5 et 6).