Correction – Quiz 2 sur le Chapitre 2
(données de panel)

(L.G.) – Cette version : 25 février 2022

ENSAE 2A – Économétrie 2 – Printemps 2022

Sauf mention contraire, les notations utilisées cherchent à reprendre celles du cours. On suppose également l'hypothèse permanente faite dans le cours d'échantillonnage i.i.d. inter-unités/clusters (voir le dernier point du slide 3 du Chapitre 2).

Structure et résumé du Chapitre 2 sur les panels

Comme dans le cours, les hypothèses, résultats, équations, etc. sont toujours écrits implicitement pour tout i dans $\{1, \dots, n\}$ (on omet d'ailleurs parfois l'indice i comme cela est fait à certains moments du cours) car on suppose toujours les variables i.i.d. inter unités i.

De plus, s'il n'y a pas de précision contraire, elles sont également écrites implicitement pour tout $t \in \{1, \dots, T\}$.

Le modèle de base considéré est, en décomposant comme dans le cours l'erreur ν_{it} en un effet individuel α_i et un choc idiosyncratique ε_{it},

$$Y_{it} = X'_{it}\beta_0 + \nu_{it} = X'_{it}\beta_0 + \alpha_i + \varepsilon_{it}, \tag{1}$$

où l'on cherche à estimer le paramètre β_0 considéré comme le paramètre causal ou structurel d'intérêt.

Les trois sections principales du Chapitre 2 (en oubliant l'introduction et l'exemple récapitulatif) correspondent à différentes hypothèses obligeant à différentes méthodes pour estimer de façon consistante le paramètre β_0 et également, en second ordre d'importance après la consistance, pour *l'estimer de façon efficace ou optimale*, c'est-à-dire avec la meilleure précision asymptotique au sens de plus faible variance asymptotique pour l'estimateur.

1. **"Exogénéité des résidus mais autocorrélation"** (première section). **On suppose l'exogénéité des régresseurs dans** (1), c'est-à-dire, $\mathbb{E}[\boldsymbol{X_{it}\nu_{it}}] = \mathbf{0}$.
   Une condition suffisante pour cela est de supposer d'une part [1] $\mathbb{E}[X_{it}\alpha_i] = 0$ et d'autre part *l'exogénéité faible* [2] des régresseurs X_{it} dans (1) : $\mathbb{E}[X_{it}\varepsilon_{it'}] = 0$, $\forall (t, t')$ *tels que* $t' \geq t$.
   En effet, l'exogénéité faible implique en particulier que les régresseurs sont non corrélés aux chocs idiosyncratiques contemporains : $\mathbb{E}[X_{it}\varepsilon_{it}] = 0$, $\forall t$.

1. L'hypothèse $\mathbb{E}[X_{it}\alpha_i] = 0$ est parfois appelée hypothèse de "*random effects*" au sens où les effets individuels α_i propres à chaque unité/cluster i sont supposés "random", aléatoires au sens du langage ordinaire : comme s'ils étaient tirés au hasard, indépendamment de tout et en particulier donc indépendamment des régresseurs X_{it} (et donc non corrélés avec ces derniers).

2. Attention à la terminologie : "*exogénéité faible*" et "*exogénéité stricte*" ont des sens bien précis dans le cadre de données de panel et ont trait à l'absence de corrélation (ou à l'orthogonalité de façon générale, équivalente dès lors qu'on peut supposer sans perte de généralité les ε_{it} centrés ce qui est le cas dans le cours car β_0 dans (1) contient une constante) entre les erreurs *idiosyncratiques* ε_{it} et les régresseurs X_{it}. En comparaison, le terme "*exogénéité*" sans autre précision fait référence au sens habituel introduit avec des données en coupe au Chapitre 1 : l'absence de corrélation (ou l'orthogonalité) entre *l'ensemble du terme d'erreur* ν_{it} et les régresseurs X_{it}.

On a ainsi, par la décomposition du terme d'erreur et la linéarité de l'espérance, pour tout t, $\mathbb{E}[X_{it}\nu_{it}] = \mathbb{E}[X_{it}\alpha_i]+\mathbb{E}[X_{it}\varepsilon_{it}] = 0+0 = 0$. De ce fait, l'écriture (1), qui est l'écriture linéaire "causale" au sens où elle fait intervenir le paramètre d'intérêt β_0, coïncide sous ces hypothèses avec l'écriture linéaire "projection".

Par conséquent, **l'estimateur des MCO empilés $\widehat{\beta}_{\text{MCO}}$ de Y_{it} sur X_{it}** est un **estimateur consistant et asymptotiquement normal** de β_0 mais *attention à l'estimation de l'écart-type asymptotique de $\widehat{\beta}_{\text{MCO}}$* $\longrightarrow$ erreur-type prenant en compte l'autocorrélation des observations au sein d'une même unité (via l'option `cluster` dans Stata) pour une inférence correcte.

2. **On autorise**[3] **$\mathbb{E}[X_{it}\alpha_i] \neq 0$** mais on suppose toujours au moins *l'exogénéité faible* (sous-entendu des régresseurs X_{it} dans (1)).

   Dans ce cas, l'estimateur des MCO empilés $\widehat{\beta}_{\text{MCO}}$ n'est plus un estimateur consistant de β_0 car on n'a plus l'exogénéité des régresseurs dans (1) puisqu'alors $\mathbb{E}[X_{it}\nu_{it}] = \mathbb{E}[X_{it}\alpha_i] \neq 0$ $\longrightarrow$ il faut faire autre chose.

   (a) **"$\mathbb{E}[X_{it}\alpha_i] \neq 0$ mais *exogénéité stricte* : $\mathbb{E}[X_{it}\varepsilon_{it'}] = 0, \forall(t,t')$"** (deuxième section).

   Si l'on suppose *l'exogénéité stricte* des régresseurs X_{it} dans (1) : $\mathbb{E}[X_{it}\varepsilon_{it'}] = 0$, $\forall(t,t')$, c'est-à-dire que les régresseurs sont supposés être non corrélés aux chocs idiosyncratiques non seulement contemporains et futurs (ce qui est *l'exogénéité faible* : $t' \geq t$) mais également aux chocs idiosyncratiques passés ($t' < t$), alors on peut construire (par des transformations éliminant α_i) **deux estimateurs consistants et asymptotiquement normaux de β_0** :

   — **l'estimateur des différences premières $\widehat{\beta}_{\text{FD}}$** = l'estimateur des MCO empilés *sur les données différenciées : ΔY_{it} et ΔX_{it}, $i = 1,\ldots,n$, $t = 2,\ldots,T$*, c'est-à-dire en évolutions temporelles entre deux dates consécutives (attention, on perd une date du fait de cette transformation), et

   — **l'estimateur within $\widehat{\beta}_{\text{W}}$** = l'estimateur des MCO empilés *sur les données en écart ou variation autour de la moyenne intra (= within) unité : $\widetilde{Y}_{it}$ et $\widetilde{X}_{it}$, $i = 1,\ldots,n$, $t = 1,\ldots,T$.*

   On peut alors se demander lequel utiliser et, pour cela, se poser la question de leur optimalité ou efficacité asymptotique : lequel est le "plus précis" par rapport à l'autre, c'est-à-dire, a une variance asymptotique plus faible (au sens des matrices semi-définies positives) ? La réponse dépend de la dynamique des chocs idiosyncratiques :

   — $\widehat{\beta}_{\text{W}}$ est plus précis lorsque les $(\varepsilon_t)_t$ sont peu corrélés : peu de dépendance temporelle ou peu d'autocorrélation intra-unité plus largement dans les chocs idiosyncratiques, alors que

   — $\widehat{\beta}_{\text{FD}}$ est plus précis lorsque les $(\varepsilon_t)_t$ sont très corrélés : beaucoup de dépendance temporelle ou beaucoup d'autocorrélation intra-unité dans les chocs idiosyncratiques (donc $(\Delta\varepsilon_t)_t$ peu corrélés).

   En théorie, on pourrait également utiliser l'estimateur GMM en deux étapes pour construire l'estimateur optimal de β_0 au sens de celui ayant la plus faible variance asymptotique. En pratique néanmoins, cet estimateur est souvent assez instable numériquement et peu précis à distance finie, c'est-à-dire pour une taille d'échantillon donnée, non-asymptotiquement.

3. Par opposition au cadre "*random effects*" où l'on suppose $\mathbb{E}[X_{it}\alpha_i] = 0$, on parle parfois de "*fixed effects*" pour l'hypothèse $\mathbb{E}[X_{it}\alpha_i] \neq 0$. C'est aussi pour cela qu'on appelle parfois l'estimateur within $\widehat{\beta}_{\text{W}}$ l'estimateur "*fixed effects*".

(b) **"$\mathbb{E}[X_{it}\alpha_i] \neq 0$ mais *exogénéité faible* : "$\mathbb{E}[X_{it}\varepsilon_{it'}] = 0, \forall t' \geq t$"** (troisième section).

Si l'on ne suppose plus l'exogénéité stricte mais seulement l'exogénéité faible. Alors

i. Les estimateurs des différences premières $\widehat{\beta}_{\mathrm{FD}}$ et within $\widehat{\beta}_{\mathrm{W}}$ ne sont plus des estimateurs consistants [4] de $\beta_0 \longrightarrow$ il faut trouver autre chose.

ii. L'idée : l'exogénéité faible donne des conditions de moments permettant d'identifier β_0 dans le modèle en différences premières [5] : $\mathbb{E}[X_{is}\Delta\varepsilon_{it}] = 0, \forall s < t$.

iii. On pourrait calculer l'estimateur GMM en deux étapes optimal à partir de ces conditions de moments mais il s'avère instable numériquement en pratique.

iv. C'est pourquoi on utilise souvent seulement certaines de ces conditions de moment pour instrumenter ΔX_t dans le modèle en différences premières

$$\Delta Y_t = \Delta X_t'\beta_0 + \Delta\varepsilon_t, \quad \forall t \in \{2, \ldots, T\}$$

par une combinaison linéaire des régresseurs passés $(X_s)_{s<t} \longrightarrow$ **estimateur 2MC dans le modèle en différences premières**.

Il y a notamment deux choix simples typiques : instrumenter ΔX_t par

— la variable (en niveau) retardée ("*lagged*") : X_{t-1}, ou bien par

— l'évolution temporelle, la variation retardée : $\Delta X_{t-1} = X_{t-1} - X_{t-2}$

En pratique, le choix dépend essentiellement de la qualité de la régression de première étape (condition de rang/pertinence de l'instrument) : quel instrument est le plus fort, le plus corrélé avec ΔX_t ?

Pour les deux choix, la seconde condition de validité de l'instrument (condition d'exogénéité) est satisfaite du fait des conditions de moments provenant de l'exogénéité faible : $\mathbb{E}[X_s\Delta\varepsilon_t] = 0, \forall s < t$.

Question 1 (exogénéités et expression des MCO)

On considère le modèle principal du Chapitre 2, où l'on décompose le terme d'erreur ν_{it} en la somme $\alpha_i + \varepsilon_{it}$ (voir notamment les slides 4 et 5 du cours) :

$$Y_{it} = X_{it}'\beta_0 + \nu_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it}. \tag{2}$$

On suppose toujours un échantillonnage i.i.d. inter-unité statistique (indexée par i). De manière sous-entendue, cette équation (2) est ainsi valide pour tout individu/unité/cluster i. On pourrait l'écrire en omettant l'indice i, comme cela se fait souvent dans un cadre de données en coupes :

$$Y_t = X_t'\beta_0 + \alpha + \varepsilon_t.$$

Toutefois, on procède plutôt moins souvent à cette omission de l'indice i des unités dans un cadre de données de panel, par exemple pour rendre plus explicite que α est un α_i spécifique à chaque individu (α_i est l'hétérogénéité individuelle inobservée permanente, constante dans le temps, par opposition aux chocs temporaires, conjoncturelles ε_{it}). L'indice i est malgré cela fréquemment omis quand il n'y a pas d'ambiguïté, notamment lorsqu'on définit des moments théoriques en population (voir par exemple la matrice J à la slide 8 du Chapitre 2).

Implicitement, l'équation (2) tient également à toute date, pour tout $t \in \{1, \ldots, T\}$.

4. Le biais asymptotique de $\widehat{\beta}_{\mathrm{FD}}$ ne dépend pas de T alors que celui de $\widehat{\beta}_{\mathrm{W}}$ tend vers 0 quand $T \to +\infty$. S'il fallait en choisir un des deux, ce serait donc plutôt $\widehat{\beta}_{\mathrm{W}}$. Malgré cela, attention, ce biais peut rester grand à distance finie pour des T petits. Rappel : l'asymptotique considérée est pour T fixé alors que n tend vers l'infini.

5. On utilise le modèle en différences premières et non le modèle within car dans ce dernier le terme d'erreur $\widetilde{\varepsilon}_{it}$ fait intervenir tous les chocs idiosyncratiques $(\varepsilon_{it})_{t=1,\ldots,T}$, y compris donc les chocs passés sur lesquels on n'a plus d'hypothèse d'exogénéité, puisqu'on suppose désormais seulement l'exogénéité faible et non stricte.

(a) Écrivez formellement ce que signifie l'exogénéité de X_{it} dans l'équation (2) ("le régresseur X_{it} est exogène") dans le sens classique, vu avec des données en coupe au Chapitre 1.

De manière générale, on peut dire que *l'exogénéité d'un ou de régresseurs* X_{it} (sans la précision "faible" ou "stricte" propre aux données de panel) signifie que X_{it} n'est pas corrélé[6] avec le terme d'erreur intervenant dans le modèle linéaire "représentation causale", c'est-à-dire l'équation faisant intervenir le paramètre d'intérêt qu'on cherche à estimer.

Ici, l'équation (2) est justement cette représentation "causale"[7] puisqu'elle fait intervenir le paramètre d'intérêt β_0. Dans cette équation (2), le terme d'erreur (hétérogénéité individuelle propre à chaque observation inobservée) est ν_{it}. L'exogénéité de X_{it} signifie donc dans ce cadre

$$\forall t \in \{1, \ldots, T\},\ \mathbb{E}[X_{it}\nu_{it}] = 0, \tag{3}$$

où cette équation est implicitement écrite pour tout i puisque les unités statistiques indexées par i sont supposées i.i.d. On pourrait ainsi l'écrire en omettant l'indice i :

$$\forall t \in \{1, \ldots, T\},\ \mathbb{E}[X_t\nu_t] = 0.$$

Or, X_{it} est le vecteur colonne $K \times 1$ des régresseurs pour l'observation indexée par it, c'est-à-dire pour la t-ème observation (une date pour le cadre de panels temporels ou un élève pour le cadre de panel multi-niveaux par exemple) de la i-ème unité (un individu ou une classe par exemple). X_{it} contient la constante (sa première composante vaut 1) et l'équation précédente (3) est donc équivalente à

$$\forall t \in \{1, \ldots, T\},\ \mathbb{E}[\nu_{it}] = 0 \text{ et } \forall j \in \{2, \ldots, K\}, \mathbb{C}\mathrm{ov}(X_{itj}, \nu_{it}) = 0,$$

où X_{itj} désigne la j-ème composante (le j-ème régresseur) de X_{it} (la première composante $j = 1$ est la constante, d'où le fait que ν_{it} est centré et, par conséquent, l'égalité entre l'espérance du produit avec ν_{it} et la covariance avec ν_{it}).

(b) En comparaison, rappelez les définitions formelles de *l'exogénéité faible* et de *l'exogénéité stricte* telles que définies dans le cadre de données de panel, toujours en utilisant les notations de l'équation (2). *Indice* : attention ici à bien préciser les quantificateurs.

Dans le cadre de données de panel, les termes (qui s'entendent comme un tout en précisant bien l'adjectif associé) "exogénéité faible" et "exogénéité stricte" font référence à la non-corrélation / l'orthogonalité entre les régresseurs X_{it} et la partie idiosyncratique propre à chaque observation it du terme d'erreur, c'est-à-dire ε_{it} ; ε_{it} par opposition à

— l'ensemble du terme d'erreur $\nu_{it} \longrightarrow$ question de l'exogénéité des régresseurs X_{it} (sans adjectif, au sens classique de (a)) ; et à

6. Ou est orthogonal de manière plus générale ; les deux sont équivalents dès lorsqu'il y a une constante et qu'on peut donc supposer les termes d'erreur centrés sans perte de généralité (ce qui est toujours le cas dans le cours).

7. Vous pouvez dire "structurelle" si vous préférez au lieu de "causale" (les variables potentielles d'Économétrie 1 permettant de définir formellement la causalité n'étant ici pas explicitées) ; le point important en tout cas est de bien distinguer cette représentation linéaire où intervient le paramètre qu'on veut estimer (représentation "causale" ou "structurelle") par rapport à la représentation "projection linéaire" dans laquelle, par construction, le résidus sera toujours (sous réserve des faibles conditions de moment habituelles) orthogonal avec les régresseurs – c'est pourquoi on ne parle d'exogénéité que relativement au terme d'erreur d'une représentation causale car pour la représentation "projection linéaire" cela ne fait aucun sens, ne signifie rien puisqu'il y a, par construction de la projection, orthogonalité / non-corrélation des régresseurs avec le résidus de la projection linéaire. Remarque de terminologie : c'est aussi pourquoi j'essaye si possible de réserver le mot "terme d'erreur" pour le terme agrégeant tous les facteurs inconnus influençant Y par l'économètre intervenant dans l'équation causale et le mot "résidus" pour le résidus dans l'équation projection linéaire, qu'il faut effectivement vraiment voir uniquement comme le résidus, le reste d'une projection orthogonale.

— la partie constante α_i au sein d'une même unité du terme d'erreur $\longrightarrow$ question de la corrélation entre X_{it} et α_i ("*random effects*" ou "*fixed effects*").

L'exogénéité stricte signifie que les régresseurs sont orthogonaux / non-corrélés avec les chocs idiosyncratiques pour n'importe quelle paire de dates/d'observations, à la fois donc avec les chocs idiosyncratiques *futurs et passés* (dans la terminologie des panels temporels) :

$$\forall (t,s) \in \{1,\ldots,T\}^2,\ \mathbb{E}[X_{it}\varepsilon_{is}] = 0. \tag{4}$$

En comparaison, l'exogénéité faible signifie que les régresseurs sont orthogonaux / non-corrélés avec les chocs idiosyncratiques *contemporains et futurs uniquement* (mais non passés) :

$$\forall (t,s) \in \{1,\ldots,T\}^2 \text{ tel que } s \geq t,\ \mathbb{E}[X_{it}\varepsilon_{is}] = 0. \tag{5}$$

(c) Pour cette question (c) uniquement, on suppose qu'on a en fait des données de coupe et non de panels ($T = 1$ formellement : on observe les unités sur une seule période).
Écrivez dans ce cas l'expression de l'estimateur MCO de Y sur X et indiquez une condition nécessaire et suffisante pour qu'il soit un estimateur consistant du paramètre d'intérêt β_0.

Pour faciliter le lien avec la question (d) suivante, on laisse visible le double indice it même si ici t ne prend que la valeur 1 ($T = 1$, une seule observation par unité statistique)[8]. D'après le Chapitre 1, slide 5, l'estimateur MCO de Y sur X avec des données en coupe est égal à

$$\begin{aligned}\widehat{\beta}^{T=1}_{\text{MCO}} &:= \left(\frac{1}{n}\sum_{i=1}^{n} X_{i1}X'_{i1}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_{i1}Y_{i1}\right)\\ &= \left(\frac{1}{n\times 1}\sum_{i=1}^{n}\sum_{t=1}^{1} X_{it}X'_{it}\right)^{-1}\left(\frac{1}{n\times 1}\sum_{i=1}^{n}\sum_{t=1}^{1} X_{it}Y_{it}\right).\end{aligned}$$

Une condition nécessaire et suffisante pour que $\widehat{\beta}^{T=1}_{\text{MCO}}$ (l'exposant $T = 1$ explicite qu'il s'agit de l'expression dans le cas de données en coupe) soit un estimateur consistant de β_0 est que l'écriture linéaire "causale" (implicitement pour tout i puisque les unités sont supposées i.i.d.)

$$Y_{i1} = X'_{i1}\beta_0 + \nu_{i1}$$

coïncide avec l'écriture linéaire "projection linéaire". Pour cela, en supposant comme toujours (bien qu'implicitement souvent dans le cours et les TD) les hypothèses de moment habituelles vérifiées[9], il faut et il suffit que le terme d'erreur de l'écriture causale ν_{i1} soit orthogonal avec les régresseurs, c'est-à-dire que $\mathbb{E}[X_{i1}\nu_{i1}] = 0$; ce qui est justement aussi la définition de l'exogénéité de X_{i1} (voir (a) avec ici $T = 1$).

(d) On revient maintenant au cadre du cours avec des données de panel ($T > 1$).
Écrivez l'expression de l'estimateur des MCO dits "empilés" de Y sur X et expliquez ce nom. Vous pourrez pour cela faire le lien avec l'expression de l'estimateur des MCO habituels sur des données en coupe de la question (c).

L'estimateur des MCO dits empilés de Y sur X, qu'on note ici $\widehat{\beta}_{\text{MCO}}$ comme dans le cours, correspond exactement à l'estimateur des MCO habituel des données en coupe $\widehat{\beta}^{T=1}_{\text{MCO}}$ *mais en*

8. X_{i1} est donc le vecteur colonne $K \times 1$ des régresseurs pour la première et seule observation de l'unité i (et non la première composante qui serait X_{i11} avec les notations employées ici attention).

9. C'est-à-dire ici $\mathbb{E}[Y^2_{i1}] < +\infty$, $\mathbb{E}[\|X_{i1}\|^2] < +\infty$, et $\mathbb{E}[X_{i1}X'_{i1}]$ inversible.

utilisant toutes les observations : les T observations de chacune des unités $i \in \{1, \ldots, n\}$, d'où ce terme de MCO "empilés" ou "*pooled OLS*" en anglais qui indiquent qu'on se sert bien de toutes les observations.

Ainsi, alors qu'avec des données en coupe il suffit de sommer sur un seul indice $\sum_{i=1}^{n}$ pour faire la somme sur toutes les observations utilisées, avec des données en panel et le double indice it (la t-ème observation de l'unité i), pour faire la somme sur toutes les observations, il faut faire formellement la double somme $\sum_{i=1}^{n} \sum_{t=1}^{T}$. On utilise au total $n \times T$ observations. De là l'expression de l'estimateur MCO empilés de Y sur X en données de panel (Chapitre 2, slide 7)

$$\widehat{\beta}_{\text{MCO}} := \left(\frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} X_{it}' \right)^{-1} \left(\frac{1}{n \times T} \sum_{i=1}^{n} \sum_{t=1}^{T} X_{it} Y_{it} \right).$$

(e) Montrez que l'hypothèse d'exogénéité faible (au sens spécifique des données de panel) associée à l'hypothèse suivante [10]

$$\forall t \in \{1, \ldots, T\},\ \mathbb{E}[X_{it} \alpha_i] = 0 \tag{6}$$

impliquent une condition assurant que la représentation linéaire de l'équation (2) (qui fait intervenir le paramètre d'intérêt/causal β_0) coïncide avec la représentation "projection linéaire" [11] de la régression de Y sur X.

$\widehat{\beta}_{\text{MCO}}$ est l'estimateur MCO du modèle linéaire

$$\forall i \in \{1, \ldots, n\}, \forall t \in \{1, \ldots, T\},\ Y_{it} = X_{it}' \beta_0 + \nu_{it},$$

où l'on a précisé les quantificateurs pour i et t (souvent laissés implicites plutôt) pour souligner le fait qu'il s'agit de l'estimateur des MCO empilés.

Cela reste néanmoins un simple estimateur MCO. Dès lors, $\widehat{\beta}_{\text{MCO}}$ sera un estimateur consistant de β_0 dès lors que la représentation linéaire "causale" ou "structurelle" ci-dessus coïncide avec la représentation linéaire "projection".

Pour cela, il suffit à nouveau que le terme d'erreur ν_{it} soit orthogonal aux régresseurs X_{it} et cela sous-entendu pour toute observation, c'est-à-dire dans le cadre de panels avec le double indice des observations : pour tout $i \in \{1, \ldots, n\}$ (laissé implicite en général du fait de l'hypothèse d'échantillonnage i.i.d. des unités) et pour tout $t \in \{1, \ldots, T\}$.

Une condition assurant que la représentation linéaire de l'équation (2) coïncide avec la représentation "projection linéaire" est donc

$$\forall i \in \{1, \ldots, n\}, \forall t \in \{1, \ldots, T\},\ \mathbb{E}[X_{it} \nu_{it}] = 0. \tag{7}$$

Cette condition (7) est impliquée par l'hypothèse d'exogénéité faible (5) associée à l'hypothèse (6) de la question (e), hypothèse parfois qualifiée d'hypothèse de "*random effects*".

En effet, l'exogénéité faible (5) implique en particulier (en prenant $s = t$)

$$\forall t \in \{1, \ldots, T\},\ \mathbb{E}[X_{it} \nu_{it}] = 0.$$

Par conséquent, en utilisant la décomposition du terme d'erreur $\nu_{it} = \alpha_i + \varepsilon_{it}$ et la linéarité de l'espérance, on obtient, pour tout $i \in \{1, \ldots, n\}$ et pour tout $t \in \{1, \ldots, T\}$,

$$\mathbb{E}[X_{it} \nu_{it}] = \mathbb{E}[X_{it}(\alpha_i + \varepsilon_{it})] = \mathbb{E}[X_{it} \alpha_i] + \mathbb{E}[X_{it} \varepsilon_{it}] = 0 + 0 = 0,$$

ce qui est bien la condition (7).

10. Cette hypothèse est sous-entendu faite pour tout i, en raison de l'échantillonnage i.i.d. inter-unités.
11. Vous pouvez revoir Économétrie 1, Chapitre 3, slide 14 si besoin.

(f) En déduire que l'estimateur des MCO empilés de Y sur X est un estimateur consistant de β_0. Pourquoi préférer cet estimateur des MCO empilés à un estimateur MCO en coupe n'utilisant qu'une seule date (par exemple, $t = 1$ la première ou $t = T$ la dernière) ? Quelle hypothèse implicite sur l'effet causal de X sur Y est faite derrière ce choix ?

D'après la réponse précédente à la question (e), sous les hypothèses d'exogénéité faible (5) et de "*random effects*" (6), la représentation linéaire (2) faisant intervenir le paramètre β_0 coïncide avec la projection linéaire. Par construction des MCO, sous les conditions de moment habituelles supposées vérifiées ici, l'estimateur des MCO est alors [12] un estimateur consistant (et également asymptotiquement normal) du coefficient théorique de la projection, c'est-à-dire ici de β_0.

Les hypothèses d'exogénéité faible (5) et de "*random effects*" (6) impliquent donc que l'estimateur des MCO empilés $\widehat{\beta}_{\text{MCO}}$ est un estimateur consistent de β_0. C'est justement un des résultats principaux (avec aussi la normalité asymptotique) de la première section du Chapitre 2 "Exogénéité des résidus mais autocorrélation".

En fait, puisqu'on a $\mathbb{E}[X_{it}\nu_{it}] = 0$ pour tout $t \in \{1, \dots, T\}$ (et pour tout i de façon sous-entendue toujours), on pourrait aussi construire un estimateur MCO consistant de β_0 en utilisant seulement certaines dates (en utilisant la terminologie des panels temporels). En particulier, on pourrait utiliser une seule date $t_0 \in \{1, \dots, T\}$: l'estimateur MCO (en coupe car une seule date) de Y sur X en utilisant seulement la date t_0 s'écrit

$$\begin{aligned}
\widehat{\beta}^{t_0}_{\text{MCO}} &:= \left(\frac{1}{n}\sum_{i=1}^{n} X_{it_0}X'_{it_0}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_{it_0}Y_{it_0}\right) \\
&= \left(\frac{1}{n\times 1}\sum_{i=1}^{n}\sum_{t=t_0}^{t_0} X_{it}X'_{it}\right)^{-1}\left(\frac{1}{n\times 1}\sum_{i=1}^{n}\sum_{t=t_0}^{t_0} X_{it}Y_{it}\right).
\end{aligned}$$

Sous les mêmes hypothèses d'exogénéité faible (5) et de "*random effects*" (6), $\widehat{\beta}^{t_0}_{\text{MCO}}$ (où l'exposant indique qu'on utilise seulement la date t_0) est aussi un estimateur consistant de β_0.

Pourquoi alors utiliser $\widehat{\beta}_{\text{MCO}}$ et non $\widehat{\beta}^{t_0}_{\text{MCO}}$ pour un certain t_0 afin d'estimer β_0 ? $\widehat{\beta}_{\text{MCO}}$ utilise davantage d'information avec nT observations et, bien qu'il faille prendre en compte l'autocorrélation entre les observations d'une même unité statistique pour avoir une inférence correcte [13], $\widehat{\beta}_{\text{MCO}}$ sera donc plus précis pour estimer β_0 que $\widehat{\beta}^{t_0}_{\text{MCO}}$ qui n'utilise que n observations et laisse de côté de l'information.

L'hypothèse sous-jacente toutefois est que l'effet causal β_0 est constant au cours du temps (en utilisant toujours la terminologie des panels temporels, qui est une application majeure des

12. En toute rigueur, ce résultat présenté ainsi est une légère arnaque puisque ce qu'on sait des MCO vient du Chapitre 1 où les observations sont supposées être i.i.d. alors qu'ici justement, en données de panel, on autorise des corrélations et dépendances entre les observations d'une même unité. En fait, ici en panel, en supposant bien les unités indexées par i i.i.d. et des conditions habituelles d'existence de moments finis et d'inversibilité, il faut et il suffit que la variable aléatoire (en omettant l'indice i) $\sum_{t=1}^{T} X_t\nu_t$ soit centrée pour que $\widehat{\beta}_{\text{MCO}}$ converge en probabilité vers β_0. Par linéarité de l'espérance, une condition suffisante pour cela est bien d'avoir $\forall t \in \{1, \dots, T\}, \mathbb{E}[X_t\nu_t] = 0$. L'interprétation comme coïncidence entre les représentations "causale" et "projection linéaire" demeure toutefois intéressante pour la compréhension en comparaison avec le cadre de données en coupe et je préférais la présenter ainsi, même si c'est donc seulement une condition suffisante et non nécessaire sans autre hypothèse (ce serait le cas en ajoutant une hypothèse de stationnarité : $\mathbb{E}[X_t\nu_t]$ ne dépend pas de t : les observations indexées par t au sein d'une même unité peuvent être corrélées mais ont la même distribution marginale). Cela complique donc légèrement mais je préférais mettre cette note de bas de page pour ne pas vous arnaquer. Vous pouvez aussi revoir la preuve de la normalité asymptotique du cours pour plus de détails.

13. C'est le troisième message principal de cette première section "Exogénéité des résidus mais autocorrélation" après la consistance et la normalité asymptotique de $\widehat{\beta}_{\text{MCO}}$.

données de type panel / multi-niveaux). On peut regarder si les estimations obtenues par $\widehat{\beta}^{t_0}_{\text{MCO}}$ pour différents t_0 sont suffisamment proches (en tenant compte de l'incertitude statistique) pour tester cette hypothèse d'un effet β_0 constant au cours du temps (un unique β_0 au lieu de T β_{0t} différents indexées par $t \in \{1, \ldots, T\}$). Sinon, on capterait une sorte d'effet causal moyen au cours du temps mais qui n'est pas forcément très bien défini.

Question 2 (variance asymptotique des MCO empilés)

En Économétrie 1 (EM1), Chapitre 2, section "Clustering" (effets de grappes/clusters), à la slide 32, le Théorème 2 énonce la normalité asymptotique de l'estimateur des MCO empilés et donne l'expression de sa variance asymptotique :

$$V_{a,g} = \mathbb{E}\left(\sum_{i \in g} X_i X_i'\right)^{-1} \mathbb{E}\left[\left(\sum_{i \in g} X_i \varepsilon_i\right)\left(\sum_{i \in g} X_i \varepsilon_i\right)'\right] \mathbb{E}\left(\sum_{i \in g} X_i X_i'\right)^{-1} \tag{8}$$

Attention : les notations sont différentes de celles du Chapitre 2 d'Économétrie 2 (EM2).
EM1 : g indice les clusters / groupes / grappes / unités statistiques supposées i.i.d. et i indice les observations au sein de chaque unité, lesquelles sont possiblement corrélées intra-unité. [14]
EM2 : l'indice i est utilisé pour les clusters / groupes / grappes / unités statistiques (les individus dans des panels temporels) supposées i.i.d. alors que l'indice t est utilisé pour les observations au sein de chaque unité (t numérote ainsi les dates pour des panels temporels, les élèves au sein d'une classe, les salariés au sein d'une entreprise, etc.).

Dans le cours d'EM2, Chapitre 2, slide 9, on trouve également l'expression de la variance asymptotique de l'estimateur des MCO empilés :

$$\sqrt{n}\left(\widehat{\beta}_{\text{MCO}} - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, J^{-1} B J^{-1}) \tag{9}$$

(a) Vérifiez qu'il s'agit bien de la même expression (au changement de notations près) pour la variance asymptotique dans (8) et dans (9), c'est-à-dire que $V_{a,g}$ du cours d'EM1 est bien égal à $J^{-1} B J^{-1}$ du cours d'EM2.

En suivant les notation du cours d'EM2, on omet l'indice i puisque les variables sont supposées i.i.d. inter-unités $i \in \{1, \ldots, n\}$. En conséquence, la somme pour t allant de 1 à T (implicitement c'est la somme pour des observations au sein d'une même unité i bien sûr) a donc la même distribution quelle que soit l'unité i considérée. Ce sont ces sommes sur les observations d'une même unité qui sont les variables i.i.d. clés intervenant dans les preuves de convergence et de normalité asymptotique dans le cas des données de panel. En particulier, sont définies dans le cours (EM2, Chapitre 2, slides 8 et 9) les espérances suivantes :

$$B := \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} X_t \nu_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t \nu_t\right)'\right],$$
$$J := \mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right).$$

J est une matrice symétrique. En effet, par linéarité de l'espérance,

$$J = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[X_t X_t'].$$

14. Ce sont également les notations employées dans l'exercice 1 du TD3-4 d'EM2.

Or, par les propriétés de la transposée appliquée à un produit matriciel, $(X_t X_t')' = (X_t')' X_t' = X_t X_t'$ est symétrique. Donc $\mathbb{E}[X_t X_t']$ est symétrique également et par conséquent, par linéarité de la transposée, J est bien symétrique.

Dans le calcul de la variance asymptotique de l'estimateur MCO empilés $\widehat{\beta}_{\text{MCO}}$, les $1/T$, qui sont des scalaires et qui commutent donc avec toute matrice dans le produit $J^{-1}BJ^{-1}$, se simplifient et on obtient finalement

$$J^{-1}BJ^{-1} = \mathbb{E}\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \mathbb{E}\left[\left(\sum_{t=1}^{T} X_t \nu_t\right)\left(\sum_{t=1}^{T} X_t \nu_t\right)'\right] \mathbb{E}\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1}.$$

Aux trois changements de notations près entre les deux cours :

— g de EM1 $= i$ de EM2 (indice des unités / clusters / grappes / etc. = le niveau en amont, supérieur, celui où l'on suppose bien l'échantillonnage i.i.d.),

— i de EM1 $= t$ de EM2 (indice des observations au sein d'une unité = le niveau aval, inférieur, celui où on autorise des corrélations entre observations d'une même unité),

— ε de EM1 $= \nu$ de EM2 (le résidus "total" au sens de la somme de l'effet individuel α_i et des chocs idiosyncratiques ε_{it} avec les notations d'EM2),

il s'agit bien de l'expression du cours d'EM1 de l'équation (8).

(b) Ce résultat de normalité asymptotique de $\widehat{\beta}_{\text{MCO}}$ requiert-il une hypothèse d'homoscédasticité ? Le cas échéant, précisez l'expression formelle de cette condition.

Ces deux questions (b) et (c) font directement référence aux explications des slides 11 et 12 du Chapitre 2 d'EM2 qui discutent le résultat de normalité asymptotique de $\widehat{\beta}_{\text{MCO}}$ et l'expression de sa variance asymptotique.

Ce résultat est obtenu sans faire d'hypothèse d'homoscédasticité et il est donc robuste à l'hétéroscédasticité.

L'écriture formelle de l'hypothèse d'homoscédasticité [15] serait ici, en omettant l'indice i car données supposées i.i.d. inter-unité :

$$\forall t \in \{1, \dots, T\},\ \mathbb{E}[\nu_t^2 X_t X_t'] = \underbrace{\mathbb{E}[\nu_t^2]}_{\text{noté } \sigma^2} \mathbb{E}[X_t X_t'] = \sigma^2 \mathbb{E}[X_t X_t']. \tag{10}$$

En plus de l'homoscédasticité classique des données en coupe, cette hypothèse exprimée dans le cadre de données de panel y associe une hypothèse de stationnarité au sens où l'on suppose dans (10) que le moment d'ordre 2 (égale à la variance si ν_t est centré, ce qui est sans perte de généralité lorsqu'il y a une constante dans le modèle) du terme d'erreur ν_t ne dépend pas de t, d'où cette notation σ^2, non indicée par t.

La version du cours est légèrement différente, mais équivalente : elle est écrite (voir slide 11) pour deux composantes j et l quelconques de X_t.

L'expression est ici rappelée simplement pour insister sur le fait que la question hétéroscédasticité ou homoscédasticité des résidus est distincte de celle de l'autocorrélation ou non des résidus : dans (10), on peut remarquer qu'intervient seulement ν_t, un seul terme d'erreur et non un autre terme ν_s pour $s \neq t$. La question a à voir avec, pour un terme d'erreur individuel, isolé, si sa variance (son moment d'ordre 2) est ou non corrélée avec les régresseurs. Cela ne dit rien sur d'éventuelles corrélations avec d'autres termes d'erreur. C'est d'ailleurs pourquoi cette question d'hétéroscédasticité ou d'homoscédasticité apparaît également avec des données en coupe.

15. Écrit ici dans une forme faible par opposition à une forme forte qui ferait intervenir la variance ou le moment d'ordre 2 conditionnellement aux régresseurs.

(c) Ce résultat de normalité asymptotique de $\widehat{\beta}_{\mathrm{MCO}}$ impose-t-il des restrictions sur la corrélation des termes d'erreur au sein d'une même unité ? Le cas échéant, précisez l'expression formelle de cette condition.

Le résultat de normalité asymptotique de $\widehat{\beta}_{\mathrm{MCO}}$ n'impose aucune restriction sur la dépendance entre les termes d'erreur au sein d'une même unité, il est donc robuste à toute forme d'autocorrélation des $(\nu_t)_{t\in\{1,\ldots,T\}}$.

L'absence d'autocorrélation entre les termes d'erreur $(\nu_t)_{t\in\{1,\ldots,T\}}$ s'écrirait formellement

$$\forall (t,s) \in \{1,\ldots,T\}^2 \text{ tels que } s \neq t,\ \mathbb{E}[\nu_t \nu_s] = 0$$

dans une écriture inconditionnelle aux régresseurs. Dans cette équation, l'espérance du produit peut être remplacée par la covariance $\mathbb{C}\mathrm{ov}(\nu_t, \nu_s)$ dès lors qu'il y a une constante dans le modèle et qu'on peut donc centrer sans perte de généralité les termes d'erreur.

Une écriture qu'on pourrait qualifier d'absence d'autocorrélation conditionnellement aux régresseurs [16] serait

$$\forall (t,s) \in \{1,\ldots,T\}^2 \text{ tels que } s \neq t,\ \mathbb{E}[X_t \nu_t \nu_s X_s'] = 0.$$

Question 3 (types de données)

(a) Quelles différences, similitudes, relations y a-t-il entre ces trois types de données ?

1. données de panel
2. données de coupes répétées
3. données longitudinales

D'après la slide 3 du Chapitre 5 d'EM1, les données longitudinales sont de manière générale des données où l'on observe les mêmes variables à différentes dates dans le temps. Elles regroupent donc :

— les données de coupe répétées (“*repeated cross-section data*”) : on mesure les mêmes variables à plusieurs dates dans le temps mais pour des *unités différentes* ;

— les données de panel (temporel) : on est capable de suivre, d'identifier les unités dans le temps et on mesure donc les mêmes variables à plusieurs dates pour les *mêmes unités*.

Remarque : il y a parfois des conventions différentes où le terme de “données longitudinales” est utilisé comme synonyme de données de panel exclusivement (la question était surtout pour vous faire réfléchir à ces différences au-delà des terminologies concurrentes).

(b) Pour ces types de données, indiquez pour chacun de ces estimateurs vus en cours d'EM1 (Chapitre 5) et d'EM2 (Chapitre 2) s'ils peuvent ou non être appliqués

1. estimateur des MCO empilés (“pooled OLS”)
2. estimateur des différences premières (“first-difference”)
3. estimateur within
4. estimateur des différences de différences (“difference-in-differences”)

16. Il s'avère qu'on a besoin de cette version conditionnelle pour simplifier l'expression de la variance asymptotique (et la diminuer surtout) – voir par exemple la discussion des conditions assurant que $\widehat{\beta}_{\mathrm{FD}}$ est efficace / optimal asymptotiquement à la fin de la slide 18 du Chapitre 2 en remplaçant les $\Delta\varepsilon_t$ par ν_t et ΔX_t par X_t.

Le chapitre 5 du cours d'EM1 considère des méthodes qui peuvent s'appliquer à la fois aux données de panel et aux données de coupe répétés. On peut calculer l'estimateur des différences de différences (DID) avec les deux types de données longitudinales. Sur ce thème, voir par ailleurs l'exercice 2 du TD3-4 d'EM2 pour une équivalence entre l'estimateur DID et l'estimateur within dans le cadre de données de panel.

Pour l'estimateur des MCO empilés, la réponse est inutilement compliquée par des questions sémantiques. On peut faire des MCO avec toutes les observations dans des données de coupe répétées, et, en ce sens, on peut bien calculer l'estimateur des MCO empilés pour les deux types de données longitudinales. Néanmoins le terme “MCO empilés” est la plupart du temps employé dans un cadre de panel uniquement où on “empile” les variables observées à différentes dates pour un même individu.

L'estimateur des différences premières et l'estimateur within s'appliquent par contre exclusivement aux données de panel. Ils ne sont pas applicables sur les données de coupe répétées. En effet, on a besoin de suivre les *mêmes unités* pour considérer les variations temporelles (différences premières) ou les écarts/variations autour de la moyenne intra-unité (within) et ainsi se débarrasser du terme d'erreur individuel constant dans le temps α_i.

Question 4 (estimateur MCO et inférence)

On considère le modèle de panel standard étudié au Chapitre 2

$$Y_{it} = X_{it}'\beta_0 + \nu_{it},$$

où $i = 1, \ldots, n$ indice les individus et $t = 1, \ldots, T$ indice les dates. On suppose que pour tous individu i et date t, $\mathbb{E}[X_{it}\nu_{it}] = 0$, mais, intra-individu i, les variables $(\nu_{i1}, \ldots, \nu_{iT})$ peuvent être corrélées.

Alors, l'estimateur des MCO empilés de Y_{it} sur X_{it} (“pooled OLS” en anglais)

C'est le message principal de la première section du cours : “Exogénéité des résidus mais autocorrélation” (voir résumé du Chapitre 2 en début de document).

On suppose ici l'exogénéité des régresseurs :

$$\forall t \in \{1, \ldots, T\}, \ \mathbb{E}[X_{it}\nu_{it}] = 0,$$

mais on précise bien qu'on autorise de l'autocorrélation entre les résidus d'une même unité : pour tout $i \in \{1, \ldots, n\}$, les variables $(\nu_{i1}, \ldots, \nu_{iT})$ peuvent être corrélées.

L'estimateur des MCO empilés $\widehat{\beta}_{\text{MCO}}$ est alors un estimateur consistant et asymptotiquement normal de β_0, mais il faut utiliser un estimateur de son écart-type asymptotique prenant en compte les possibles corrélations entre les erreurs ν_{it}, ν_{is} de deux observations $t \neq s$ au sein d'une même unité i, c'est-à-dire, prenant en compte “l'autocorrélation des résidus ν”.

Cet estimateur de la variance asymptotique, $\widehat{V}$ à la slide 12 du cours, permet d'estimer de façon consistante la variance asymptotique de $\widehat{\beta}_{\text{MCO}}$ et d'avoir donc une inférence correcte en ne supposant rien sur la corrélation des observations intra-unité ni non plus sur l'homoscédasticité des résidus et donc de façon également robuste à l'hétéroscédasticité (voir aussi la Question 2 (b) pour la distinction entre ces deux notions).

1. est un estimateur consistant de β_0 et les erreurs-types [17] peuvent être obtenues comme dans des données de coupe habituelles sans tenir compte de l'hétéroscédasticité ni de l'autocorrélation (commande Stata `regress` sans option) – **Faux**, cela suppose l'homoscédasticité des résidus et l'absence d'autocorrélation.

17. = “standard-errors” en anglais, les estimateurs des écarts-types asymptotiques.

2. est un estimateur consistant de β_0 et les erreurs-types peuvent être obtenues comme dans des données de coupe habituelles en tenant compte de l'hétéroscédasticité (commande Stata `regress` avec l'option `robust`) – **Faux**, cela suppose l'absence d'autocorrélation.
3. est un estimateur consistant de β_0 et les erreurs-types peuvent être obtenues en tenant compte de l'hétéroscédasticité et de l'autocorrélation (commande Stata `regress` avec l'option `cluster(ident)` où la variable `ident` est la variable d'identifiant individuel) – **Vrai**.
4. n'est pas un estimateur consistant de β_0 – **Faux**.

Question 5 (MCO, first-difference, within, IV, GMM)

On considère le modèle de panel standard étudié au Chapitre 2, où l'on a décomposé le terme d'erreur comme dans le cours :

$$Y_{it} = X_{it}'\beta_0 + \nu_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it}.$$

On suppose $\forall t \in \{1, \dots, T\}, \mathbb{E}[X_{it}\alpha_i] \neq 0$ et $\forall (t, s) \in \{1, \dots, T\}^2, \mathbb{E}[X_{it}\varepsilon_{is}] = 0$.

Sous ces hypothèses, on peut utiliser comme estimateur consistant du paramètre d'intérêt β_0

On est exactement dans le cadre de la deuxième section du cours (voir résumé et structure du Chapitre 2 en début de document) :

— "*fixed effects*" : $\forall t \in \{1, \dots, T\}, \mathbb{E}[X_{it}\alpha_i] \neq 0$: la partie α_i constante intra-unité du terme d'erreur est autorisée à être corrélée avec les régresseurs X_{it} ;

— mais on suppose l'exogénéité stricte : $\forall (t, s) \in \{1, \dots, T\}^2, \mathbb{E}[X_{it}\varepsilon_{is}] = 0$: les régresseurs sont supposés être non corrélés avec les chocs idiosyncratiques futurs comme passés.

Dans ce cadre, on sait d'après le cours que l'estimateur des différences premières $\widehat{\beta}_{\text{FD}}$ et l'estimateur within $\widehat{\beta}_{\text{W}}$ sont des estimateurs consistants de β_0. Cela conduit directement à l'unique bonne réponse : la réponse 3.

La réponse 1 est fausse puisque l'estimateur des MCO empilés $\widehat{\beta}_{\text{MCO}}$ n'est quant à lui pas un estimateur consistant de β_0 dans ce cadre puisque les régresseurs X_{it} sont endogènes :

$$\begin{aligned}\forall t \in \{1, \dots, T\},\ \mathbb{E}[X_{it}\nu_{it}] &= \mathbb{E}[X_{it}(\alpha_i + \varepsilon_{it})] = \mathbb{E}[X_{it}\alpha_i] + \mathbb{E}[X_{it}\varepsilon_{it}] \\ &= \mathbb{E}[X_{it}\alpha_i] + 0 = \mathbb{E}[X_{it}\alpha_i] \neq 0.\end{aligned}$$

La réponse 2 propose un estimateur GMM utilisant les conditions de moment (écrites en omettant l'indice i des unités statistiques qui sont supposées i.i.d) :

$$\forall (s, t) \in \{1, \dots, T\}^2 \text{ tels que } s < t,\ \mathbb{E}[X_s(Y_t - X_t'\beta_0)] = 0. \tag{11}$$

Or, ces conditions de moment (11) ne permettent pas d'identifier β_0 puisqu'elles ne tiennent pas, sont fausses sous nos hypothèses : ici, dans ce cadre, β_0 n'est justement pas tel qu'on ait (11). En effet, en reprenant le précédent calcul, on obtient, pour tous $(s, t) \in \{1, \dots, T\}^2$,

$$\begin{aligned}\mathbb{E}[X_{is}(Y_{it} - X_{it}'\beta_0)] &= \mathbb{E}[X_{is}\nu_{it}] \text{ (par le modèle structurel)} \\ &= \mathbb{E}[X_{is}\alpha_i] + \mathbb{E}[X_{is}\varepsilon_{it}] \text{ (en décomposant le terme d'erreur)} \\ &= \mathbb{E}[X_{is}\alpha_i] + 0 \text{ (par l'hypothèse d'exogénéité stricte)} \\ &\neq 0 \text{ (par l'hypothèse "\textit{fixed effects}").}\end{aligned}$$

Sous les hypothèses faites dans cette question, on n'a donc *pas* $\forall s < t, \mathbb{E}[X_s(Y_t - X_t'\beta_0)] = 0$ mais $\forall (s, t), \mathbb{E}[X_s(Y_t - X_t'\beta_0)] \neq 0$. Dit autrement, pour toute paire $(s, t) \in \{1, \dots, T\}^2$, β_0 n'est pas une racine / un zéro de la fonction $\beta \mapsto \mathbb{E}[X_s(Y_t - X_t'\beta)]$ et on ne peut donc pas utiliser les conditions de moment (11) pour construire un estimateur GMM consistant de β_0.[18]

La réponse 2 est donc fausse. Elle était là pour faire ces quelques rappels sur les GMM et comme un piège par analogie avec les conditions de moment utilisées dans la troisième section du cours : "*fixed effects*" et exogénéité *faible* seulement (et non stricte), qui sont (voir équation (5), slide 32 du Chapitre 2 d'EM2)

$$\forall (s,t) \in \{1,\dots,T\}^2 \text{ tels que } s < t,\ \mathbb{E}[X_s(\Delta Y_t - \Delta X_t'\beta_0)] = \mathbb{E}[X_s \Delta\varepsilon_t] = 0. \qquad (12)$$

Ces conditions de moment (12) sont par contre bien correctes sous l'hypothèse d'exogénéité faible : on a bien $\mathbb{E}[X_s(\Delta Y_t - \Delta X_t'\beta_0)] = 0$ pour tout couple $(s,t) \in \{1,\dots,T\}^2$ avec $s < t$. C'est justement ce qui permet d'estimer de façon consistante β_0 dans le cadre de la troisième section du cours : soit par l'estimateur GMM utilisant les conditions de moment (12) ou bien par un estimateur 2MC sur les données empilées en différences premières avec un instrument dont l'exogénéité repose sur une partie de ces conditions de moment (12).

1. l'estimateur des MCO empilés – **Faux**.
2. l'estimateur GMM utilisant les conditions de moment[19] : $\forall s < t, \mathbb{E}[X_s(Y_t - X_t'\beta_0)] = 0$ – **Faux**.
3. l'estimateur des différences premières (first-difference) ou l'estimateur within, car les estimateurs précédents, MCO empilés et GMM de la proposition 2, ne sont pas consistants – **Vrai**.
4. aucun des estimateurs proposés précédemment n'est un estimateur consistant de β_0 et il faut recourir à des variables instrumentales – **Faux**, c'est dans la troisième section du cours avec exogénéité faible et non stricte qu'on est amené à utiliser des variables instrumentales ; ici, avec l'exogénéité stricte, on n'en a pas besoin.

Question 6 (modèle dynamique)

On considère le modèle de panel suivant

$$Y_{it} = \alpha_i + \beta_0 Y_{it-1} + \varepsilon_{it},$$

où $i = 1,\dots,n$ indice les individus et $t = 1,\dots,T$ indice les dates.

Pour tout individu i, on suppose $(\varepsilon_{i1},\dots,\varepsilon_{iT})$ i.i.d. et également que, pour toute date t, $\mathbb{E}[\varepsilon_{it}] = 0$ et $\mathbb{V}[\varepsilon_{it}] > 0$.

Alors

Il s'agit de l'exemple 1 du slide 28 du Chapitre 2 : *modèle dynamique*, simplifié car on ne considère pas d'éventuelles variables explicatives W_{it} additionnelles[20].

Dans ce modèle dynamique, pour reprendre les notations du modèle standard de panel étudié au Chapitre 2, on a donc $X_{it} = Y_{it-1}$. Or, d'après le modèle structurel, en l'écrivant à la date $t-1$, on a

$$Y_{it-1} = \alpha_i + \beta_0 Y_{it-2} + \varepsilon_{it-1}.$$

18. Vous pouvez revoir en particulier slide 26, Chapitre 1, d'EM2, deuxième bullet-point pour plus de détails sur ce raisonnement : l'hypothèse d'identification, c'est-à-dire le fait que les conditions de moment soient "correctes" au sens où le paramètre d'intérêt β_0 est bien l'unique paramètre qui annule ces conditions de moment, est nécessaire pour que l'estimateur des GMM obtenu à partir de ces conditions soit bien un estimateur consistant et asymptotiquement normal de β_0.

19. Écrites en omettant l'indice i des unités car elles sont supposées i.i.d. comme toujours.

20. Ce qui d'ailleurs ne changerait rien au raisonnement qui suit à condition de supposer ces contrôles W_{it} faiblement exogènes : $\mathbb{E}[W_{it}\varepsilon_{it'}] = 0$, $\forall t \leq t'$.

En conséquence, par construction du modèle, Y_{it-1} et ε_{it-1} sont a priori corrélés puisque

$$\begin{aligned}\mathbb{C}\mathrm{ov}(Y_{it-1}, \varepsilon_{it-1}) &= \mathbb{C}\mathrm{ov}(\alpha_i + \beta_0 Y_{it-2} + \varepsilon_{it-1}, \varepsilon_{it-1}) \\ &= \mathbb{C}\mathrm{ov}(\alpha_i, \varepsilon_{it-1}) + \beta_0 \mathbb{C}\mathrm{ov}(Y_{it-2}, \varepsilon_{it-1}) + \mathbb{V}[\varepsilon_{it-1}]\end{aligned}$$

contient au moins[21] la variance de ε_{it-1} non nulle.

Ainsi, le régresseur $X_{it} = Y_{it-1}$ est corrélé avec un choc idiosyncratique *passé* ε_{it-1} (on aurait de même avec ε_{it-s} pour $s > 1$). Par conséquent, l'exogénéité stricte n'est pas satisfaite.

Par contre, si les $(\varepsilon_{it})_t$ ne sont pas corrélés entre eux (inter-t : $\mathbb{E}[\varepsilon_{it}\varepsilon_{it'}] = 0, \forall t \neq t'$) et ne sont pas corrélés à α_i et à Y_{i0} (Y_{i0} : "condition initiale" pour la variable expliquée, avant la première date $t = 1$ apparaissant dans les données), alors l'exogénéité faible est vérifiée (voir la slide 28 du cours pour des détails).

L'exogénéité faible est vérifiée mais non la stricte $\longrightarrow$ on est dans le cadre de la troisième section du cours **"$\mathbb{E}[X_{it}\alpha_i] \neq 0$ mais *exogénéité faible* "$\mathbb{E}[X_{it}\varepsilon_{it'}] = 0, \forall t' \geq t$"** (voir résumé et structure du Chapitre 2 en début de document).

Les deux premières réponses proposées sont liées à l'estimateur des différences premières. En effet, le modèle différencié s'écrit ici, pour tout $t \in \{2, \ldots, T\}$,

$$Y_{it} - Y_{it-1} = \beta_0(Y_{it-1} - Y_{it-2}) + (\varepsilon_{it-1} - \varepsilon_{it-2}),$$

qu'on note, en utilisant l'opérateur Δ de différence première (évolution temporelle),

$$\Delta Y_{it} = \beta_0 \Delta Y_{it-1} + \Delta\varepsilon_{it}. \tag{13}$$

Il ne faut donc pas mettre de constante effectivement, comme d'habitude pour les estimateurs $\widehat{\beta}_{\mathrm{FD}}$ et $\widehat{\beta}_{\mathrm{W}}$ $\longrightarrow$ voir le premier point de la slide 14 du Chapitre 2 et la modification de β_0 qui ne contient plus la constante (implicitement si ce n'est pas précisé ; la slide 14 explicite justement cela) dès lorsqu'on parle d'estimateurs des différences premières ou within. Cela exclut donc d'emblée la deuxième réponse.

Mais surtout, puisque l'exogénéité stricte n'est pas satisfaite, *l'estimateur des différences premières* qui est exactement l'estimateur proposé par la réponse 1, c'est-à-dire *l'estimateur des MCO* (sous-entendu empilés dans ce cadre de données de panel, c'est-à-dire pour toutes les observations $i = 1, \ldots, n$ et $t = 2, \ldots, T$) *sur les données différenciées* sans la constante (soit ici les MCO empilés de $\Delta Y_{it} = Y_{it} - Y_{it-1}$ sur $\Delta Y_{it-1} = Y_{it-1} - Y_{it-2}$ sans la constante) n'est pas un estimateur consistant de β_0. La première réponse est donc fausse. Les deux premiers estimateurs proposés ne sont pas des estimateurs consistants de β_0.

D'ailleurs, l'estimateur within ne le serait pas non plus puisqu'il nécessite également l'hypothèse d'exogénéité stricte.

Exogénéité faible seulement et "*fixed effects*" (troisième section du cours) $\longrightarrow$ il faut donc utiliser des variables instrumentales.

On va instrumenter le régresseur endogène $\Delta Y_{it-1} = Y_{it-1} - Y_{it-2}$ dans le modèle en différences premières (13). En appliquant le cours[22] avec ici[23] $X_t = Y_{t-1}$, on peut instrumenter ΔY_{t-1} dans (13) par toute combinaison linéaire des régresseurs *passés*, c'est-à-dire des X_s avec $s < t$, soit ici des $X_s = Y_{s-1}$.

21. Le résultat du calcul dépend des hypothèses faites. Dans le cas favorable (au sens où il donne justement l'exogénéité faible – voir plus bas) où l'on suppose les chocs idiosyncratiques $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ non corrélés entre eux et non corrélés avec α_i ni avec Y_{i0}, alors on obtient $\mathbb{C}\mathrm{ov}(Y_{it-1}, \varepsilon_{it-1}) = \mathbb{V}[\varepsilon_{it-1}] > 0$.

22. Voir le résumé en début de document et également en particulier l'équation (5) du slide 32 du Chapitre 2.

23. En omettant l'indice i car, comme toujours, échantillonnage supposé i.i.d inter-unité $i \in \{1, \ldots, n\}$.

La réponse 3. ne fonctionne donc pas puisqu'on propose comme instrument $Y_{t-1} = X_t$ qui n'est pas un régresseur X_s passé mais le régresseur *contemporain* ($s = t$).

La réponse 4. est la bonne : on instrumente bien par un régresseur passé : $Y_{t-2} = X_{t-1}$ ($s = t - 1 < t$). On peut remarquer que ce choix correspond à la première possibilité [24] (on instrumente par le régresseur retardé en niveau) évoquée dans le cours parmi les deux choix classiques d'instruments au slide 33 du Chapitre 2.

Cette correction devrait expliquer le troisième point (N.B.) sur les modèles dynamiques à la slide 32 du Chapitre 2.

1. L'estimateur des MCO empilés de $Y_{it} - Y_{it-1}$ sur $Y_{it-1} - Y_{it-2}$ sans la constante est un estimateur consistant de β_0 – **Faux**, c'est exactement l'estimateur des différences premières $\widehat{\beta}_{\text{FD}}$ mais il requiert l'exogénéité stricte pour être consistant.
2. L'estimateur des MCO empilés de $Y_{it} - Y_{it-1}$ sur $Y_{it-1} - Y_{it-2}$ et la constante est un estimateur consistant de β_0 – **Faux**, d'une part il ne faut inclure de constante, d'autre part, sans constante, ce serait $\widehat{\beta}_{\text{FD}}$ mais il n'est pas consistant ici.
3. Les deux estimateurs précédents (1. et 2.) ne sont pas convergents. Il faut instrumenter $Y_{it-1} - Y_{it-2}$ par Y_{it-1} dans la régression de $Y_{it} - Y_{it-1}$ sur $Y_{it-1} - Y_{it-2}$ – **Faux**, Y_{it-1} n'est pas un instrument valide car il n'est pas exogène.
4. Les deux estimateurs précédents (1. et 2.) ne sont pas convergents. Il faut instrumenter $Y_{it-1} - Y_{it-2}$ par Y_{it-2} dans la régression de $Y_{it} - Y_{it-1}$ sur $Y_{it-1} - Y_{it-2}$ – **Vrai**.

Question 7 (avec une nouvelle condition d'exogénéité)

On considère le modèle de panel standard étudié au Chapitre 2, où l'on a décomposé le terme d'erreur comme dans le cours :

$$Y_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it},$$

où $i = 1, \ldots, n$ indice les individus et $t = 1, \ldots, T$ indice les dates, avec $T \geq 2$. On cherche à estimer le paramètre d'intérêt β_0.

Comme dans la seconde partie du Chapitre 2, on ne suppose *pas* $\mathbb{E}[X_{it}\alpha_i] = 0$: X_{it} et α_i peuvent être corrélées. Mais on suppose la condition suivante : pour tous individu i et date t, on a $\mathbb{E}[\varepsilon_{it} \,|\, X_{i1}, \ldots, X_{iT}] = 0$.

Alors

La condition d'exogénéité sur les chocs idiosyncratiques supposée ici est

$$\forall i \in \{1, \ldots, n\}, \forall t \in \{1, \ldots, T\}, \ \mathbb{E}[\varepsilon_{it} \,|\, X_{i1}, \ldots, X_{iT}] = 0 \, (= \mathbb{E}[\varepsilon_{it}]), \tag{14}$$

c'est-à-dire que les chocs idiosyncratiques $(\varepsilon_{it})_t$ sont “indépendants en moyenne/espérance” [25] de tous les régresseurs, passés, contemporains, ou futurs.

Or, l'indépendance en espérance (“*mean-independence*”) implique la non-corrélation par la loi des espérances itérées. On a en effet, pour toutes dates t et t' dans $\{1, \ldots, T\}$,

$$\begin{aligned}
\mathbb{E}[X_{it}\varepsilon_{it'}] &= \mathbb{E}[\mathbb{E}[X_{it}\varepsilon_{it'} \,|\, X_{i1}, \ldots, X_{iT}]] \quad \text{(loi des espérances itérées)} \\
&= \mathbb{E}[X_{it}\mathbb{E}[\varepsilon_{it'} \,|\, X_{i1}, \ldots, X_{iT}]] \quad \text{(linéarité de l'espérance conditionnelle)} \\
&= \mathbb{E}[X_{it} \times 0] = 0 \quad \text{(hypothèse (14))}.
\end{aligned}$$

24. D'où une question bonus : écrivez l'instrument utilisé si on suivait le deuxième choix dans le cadre de ce modèle dynamique. Combien de dates minimales observées seraient alors nécessaires ?

25. La terminologie anglais est “*mean-independent*” (voir par exemple la page Wikipedia *ici*) mais je ne connais pas de traduction française classique.

La condition (14) de cette question implique donc l'hypothèse d'exogénéité stricte du cours : $\forall (t,t') \in \{1,\ldots,T\}^2$, $\mathbb{E}[X_{it}\varepsilon_{it'}] = 0$.

Ainsi est-on dans le cadre de la deuxième section du cours **"$\mathbb{E}[\boldsymbol{X_{it}\alpha_i}] \neq \mathbf{0}$ mais *exogénéité stricte* $\mathbb{E}[\boldsymbol{X_{it}\varepsilon_{it'}}] = \mathbf{0}$, $\forall (\boldsymbol{t,t'})$"**. On peut alors appliquer les différents résultats du cours de cette section (voir le résumé en début de document) pour déterminer si les réponses proposées sont vraies ou fausses.

— On sait que, sous l'hypothèse d'exogénéité stricte, les estimateurs within $\widehat{\beta}_{\mathrm{W}}$ et des différences premières $\widehat{\beta}_{\mathrm{FD}}$ sont des estimateurs consistants de β_0. Ils convergent donc en probabilité vers la même valeur, $\beta_0 \longrightarrow$ la réponse 3. est fausse.

— Cette propriété de consistance, et également la propriété de normalité asymptotique d'ailleurs, ne requiert aucune hypothèse touchant à l'homoscédasticité des chocs idiosyncratiques. Autrement dit, sous exogénéité stricte, on a la consistance et la normalité asymptotique de $\widehat{\beta}_{\mathrm{FD}}$ et de $\widehat{\beta}_{\mathrm{W}}$ y compris si le modèle est hétéroscédastique au sens où la variance des chocs dépend de la valeur des régresseurs $\longrightarrow$ la réponse 2. est fausse.

  Cette condition n'intervient qu'au second ordre à propos de l'efficacité asymptotique des estimateurs (voir notamment les slides 18 et 22 du Chapitre 2) ; l'ordre un étant d'abord la consistance et la normalité asymptotique de l'estimateur avant de se préoccuper de sa précision / optimalité / efficacité asymptotique.

— les estimateurs within $\widehat{\beta}_{\mathrm{W}}$ et des différences premières $\widehat{\beta}_{\mathrm{FD}}$ sont les mêmes pour $T = 2$ (voir slide 24 du Chapitre 2). Par contre, en général, ce qui est le cas ici où l'on ne suppose pas $T = 2$, ils sont différents $\longrightarrow$ la réponse 1. est fausse.

Par élimination (une unique bonne réponse), on obtient que la réponse 2 est correcte.

Pour une justification positive de cette réponse 2, l'idée est simplement d'appliquer le résultat d'absence de biais à distance finie / non-asymptotique de l'estimateur MCO lorsque le résidus est "*mean-independent*" des régresseurs (voir TD1 d'EM1, question 2).

Ici, le modèle en différences premières est

$$\Delta Y_{it} = \Delta X_{it}'\beta_0 + \Delta\varepsilon_{it}$$

et, pour toute unité i et date t, on a

$$\mathbb{E}[\Delta\varepsilon_{it} \,|\, \Delta X_{it}] = \mathbb{E}[\varepsilon_{it} \,|\, \Delta X_{it}] - \mathbb{E}[\varepsilon_{it-1} \,|\, \Delta X_{it}] = 0 - 0 = 0.$$

En effet, pour n'importe quelle fonction mesurable g des $(X_{i1},\ldots,X_{iT})$, et ainsi en particulier pour $g(X_{i1},\ldots,X_{iT}) = \Delta X_{it} = X_{it} - X_{it-1}$, et pour toute date t, on a [26]

$$\begin{aligned}\mathbb{E}[\varepsilon_{it} \,|\, g(X_{i1},\ldots,X_{iT})] &= \mathbb{E}\left(\mathbb{E}[\varepsilon_{it} \,|\, X_{i1},\ldots,X_{iT}] \,\middle|\, g(X_{i1},\ldots,X_{iT})\right) \quad \text{(composition des projections)}\\ &= \mathbb{E}[0 \,|\, g(X_{i1},\ldots,X_{iT})] = 0 \quad \text{(hypothèse (14))}.\end{aligned}$$

Ainsi, le résidus $\Delta\varepsilon_{it}$ est "*mean-independent*" du régresseur ΔX_{it} dans le modèle en différences premières. Or, par définition, *l'estimateur des différences premières* est l'estimateur MCO (empilés) du modèle en différences premières. Donc, en appliquant le résultat du TD1 d'EM1 (question 2), il est sans biais à distance finie / non-asymptotiquement.

26. *Composition des projections* : généralisation de la loi des espérances itérées en un sens. Intuitivement, en voyant l'espérance conditionnelle comme une projection, si on projette plusieurs fois successivement, on se retrouve sur le "plus petit" espace. Exemple : si Y, X_1 et X_2 sont des variables aléatoires (potentiellement des vecteurs aléatoires) telles que les espérances considérées sont bien définies, on a $\mathbb{E}[Y \,|\, X_1] = \mathbb{E}(\mathbb{E}[Y \,|\, X_1, X_2] \,|\, X_1)$, X_1 est une fonction de (X_1, X_2). Plus largement, pour un vecteur aléatoire X, si $g(X)$ est une fonction mesurable de X, on a $\mathbb{E}[Y \,|\, g(X)] = \mathbb{E}\{\mathbb{E}[Y \,|\, X] \,|\, g(X)\}$. Remarque : si $X_1 = 1$, la constante, on retombe sur la loi des espérances itérées : projeter sur l'espace des constantes, c'est juste prendre l'espérance : $\mathbb{E}[Y \,|\, 1] = \mathbb{E}[Y]$.

1. L'estimateur within coïncide toujours avec l'estimateur des différences premières – **Faux**, c'est le cas lorsque que $T = 2$ mais non en général pour $T > 2$.
2. L'estimateur des différences premières est sans biais à distance finie [27] – **Vrai**.
3. Les estimateurs within et des différences premières convergent en probabilité vers des limites différentes – **Faux**, ce serait le cas en l'absence d'exogénéité stricte.
4. Les estimateurs within et des différences premières ne sont pas des estimateurs consistants de β_0 si le modèle est hétéroscédastique, c'est-à-dire si $\mathbb{V}[\varepsilon_{it} \mid X_{i1}, \ldots, X_{iT}]$ dépend de $(X_{i1}, \ldots, X_{iT})$ – **Faux**, sous l'hypothèse d'exogénéité stricte, les estimateurs within et des différences premières sont consistants et asymptotiquement normaux qu'importe l'hétéroscédasticité ou l'homoscédasticité des chocs.

27. Au sens du biais statistique à distance finie (ou non-asymptotique) de Statistique 1 : pour toute taille d'échantillon, l'espérance de l'estimateur est égale au paramètre qu'on cherche à estimer.