Introduction aux modèles statiques en données de panel

Avec application au lien climat–rendement agricole

Cours synthétique pour débutant

13 décembre 2025

Table des matières

1 Introduction générale

En économétrie, on distingue plusieurs grands types de données :

- **Données transversales** : un seul moment dans le temps, plusieurs unités (ménages, pays, entreprises, ...).
- **Séries temporelles** : une seule unité observée sur plusieurs périodes (PIB du Sénégal de 1980 à 2023, par exemple).
- **Données de panel** (ou longitudinales) : plusieurs unités *et* plusieurs périodes pour chaque unité.

Les **données de panel** combinent donc information *transversale* et *temporelle*. Elles permettent de suivre un même individu (ou pays) au cours du temps, et sont très utiles pour :

- contrôler les caractéristiques inobservées mais fixes dans le temps,
- mieux identifier des relations causales,
- modéliser des dynamiques (lag de la variable dépendante, etc., mais ici on reste au modèle statique).

Dans ce cours, on se concentre sur les **modèles statiques de panel** et les **tests de spécification de Hsiao (2003)**, en les illustrant sur un exemple lié à ton projet : l'impact du climat (température, précipitations, jours chauds) sur le rendement ou le PIB agricole.

2 Données de panel : notation et structure

On note généralement :

- $i = 1, \ldots, N$: les unités (pays, ménages, entreprises, etc.) ;
- $t = 1, \ldots, T$: les périodes (années, trimestres, etc.) ;
- Y_{it} : la variable dépendante (ici, par exemple, log du PIB agricole du pays i à l'année t) ;
- X_{it} : un vecteur de K variables explicatives pour l'unité i à la période t (ici : degrés-jours, précipitations, jours très chauds, etc.).

Exemple de structure de panel climat–agriculture

Supposons que l'on dispose des données suivantes :

- Pays d'Afrique : Sénégal, Mali, Burkina Faso, Niger, Tchad, Madagascar, Mauritanie, Soudan ;
- Période : 2010–2023 ;
- Variables :
  - Y_{it} : log du PIB agricole (ou du rendement moyen) du pays i à l'année t ;
  - degdays_{it} : degrés-jours de croissance (chaleur utile) ;
  - rain_{it} : précipitations cumulées en saison de croissance ;
  - hotdays_{it} : nombre de jours très chauds (au-dessus d'un seuil).

Un extrait fictif de la base pourrait ressembler au tableau 1.

TABLE 1 – Exemple simplifié de structure de panel climat–PIB agricole

Pays (i)	Année (t)	degdays_{it}	rain_{it}	hotdays_{it}	$\log(\text{PIB agri}_{it})$
Sénégal	2010	1250	650	32	10.8
Sénégal	2011	1300	620	34	10.9
...	...	...	...	...	...
Mali	2010	1400	590	36	10.5
Mali	2011	1380	610	35	10.6
...	...	...	...	...	...

3 Modèles statiques en données de panel

L'idée de base est de relier Y_{it} aux variables explicatives X_{it} par un modèle linéaire. Dans ce cours, on utilise la forme suivante :

$$Y_{it} = \alpha_i + \gamma_t + \beta' X_{it} + \varepsilon_{it}, \tag{1}$$

où :
- α_i : effet spécifique à l'unité (pays) i, constant dans le temps ;
- γ_t : effet spécifique à l'année t (choc commun à tous les pays) ;
- β : vecteur de coefficients (effet moyen de X_{it} sur Y_{it}) ;
- ε_{it} : terme d'erreur idiosyncratique.

Selon la façon dont on traite α_i (et éventuellement γ_t), on obtient différents modèles : **pooling**, **effets fixes**, **effets aléatoires**.

3.1 Modèle pooling (régression classique)

Le modèle le plus simple ignore complètement la structure de panel et suppose :

$$Y_{it} = \alpha + \beta' X_{it} + \varepsilon_{it}, \tag{2}$$

avec un *même* intercept α et les mêmes coefficients β pour tous les pays et toutes les périodes.

C'est une simple régression OLS sur toutes les observations empilées, comme si l'on avait une grande base de données transversale.

Limite principale : toutes les différences permanentes entre pays (qualité des sols, institutions, etc.) sont noyées dans l'erreur. Si ces caractéristiques sont corrélées avec les variables explicatives, on a un **biais d'omission**.

3.2 Modèle à effets fixes (Fixed Effects, FE)

Dans le modèle à **effets fixes individuels**, on autorise un intercept spécifique à chaque pays :

$$Y_{it} = \alpha_i + \beta' X_{it} + \varepsilon_{it}, \tag{3}$$

où α_i capte toutes les caractéristiques inobservées mais *constantes dans le temps* du pays i : qualité moyenne des terres, institutions, niveau structurel de productivité agricole, etc.

On peut également ajouter des effets fixes temporels :

$$Y_{it} = \alpha_i + \gamma_t + \beta' X_{it} + \varepsilon_{it}, \tag{4}$$

ce qui permet de contrôler les chocs communs à tous les pays (prix mondiaux des céréales, tendance globale du climat, etc.).

Idée clé : le modèle FE exploite les *variations dans le temps à l'intérieur d'un même pays.* Les coefficients β s'interprètent comme l'effet d'un changement de climat (par exemple une augmentation des degrés-jours) sur Y_{it}, *une fois que l'on tient compte du pays considéré.*

3.3 Modèle à effets aléatoires (Random Effects, RE)

Dans le modèle à **effets aléatoires**, on suppose :

$$Y_{it} = \alpha + u_i + \beta' X_{it} + \varepsilon_{it}, \tag{5}$$

où u_i est un effet spécifique au pays i mais traité comme une variable aléatoire, avec :

$$\mathbb{E}[u_i] = 0, \quad \mathrm{Var}(u_i) = \sigma_u^2,$$

et, surtout, on suppose :

$$\mathrm{Cov}(u_i, X_{it}) = 0 \quad \forall i, t.$$

Autrement dit, les caractéristiques inobservées propres au pays i ne sont pas corrélées aux variables explicatives X_{it}. C'est une hypothèse forte et souvent discutée en pratique.

Conséquence :
- Si l'hypothèse $\mathrm{Cov}(u_i, X_{it}) = 0$ est vraie, le modèle RE est plus efficace (plus précis) que FE.
- Si elle est fausse, les estimations RE sont biaisées, et il vaut mieux utiliser FE.

3.4 Illustration graphique : effets fixes vs effets aléatoires

La figure 1 illustre schématiquement la différence entre un monde « effets fixes » et un monde où les pentes/intercepts varient aléatoirement.

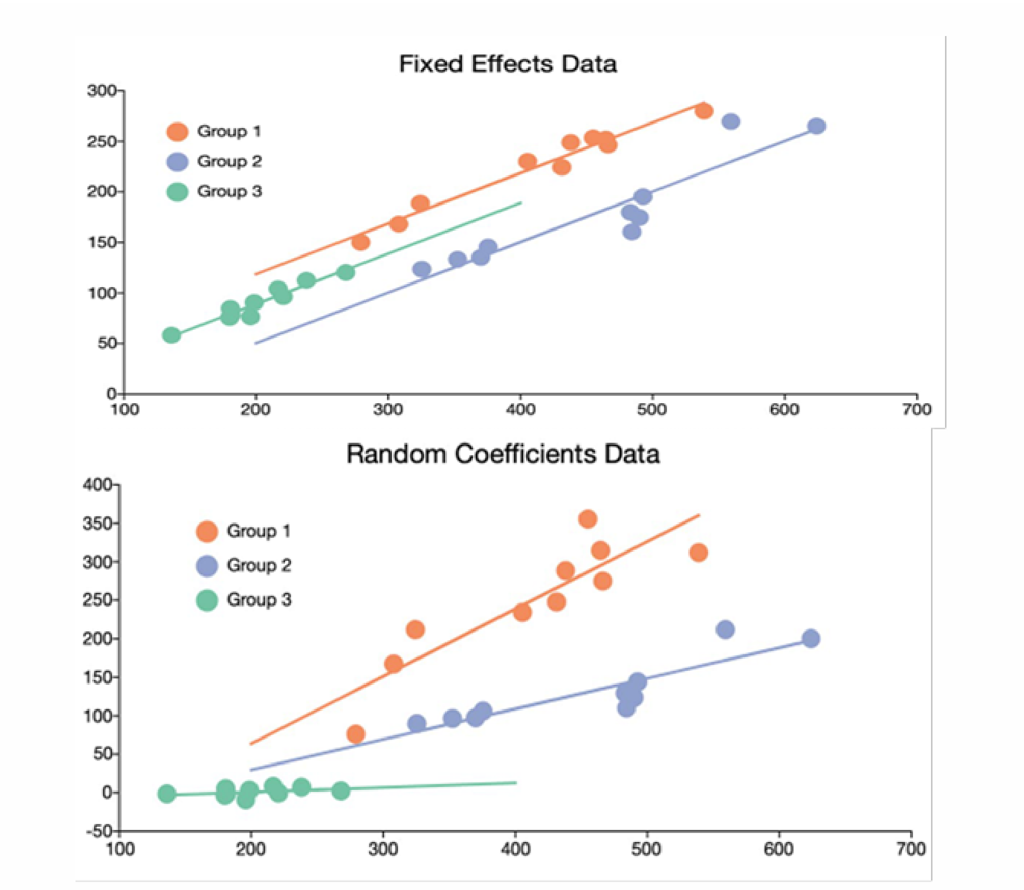

Figure 1 – Illustration : données compatibles avec un modèle à effets fixes (haut) ou à coefficients aléatoires (bas).

Dans la partie supérieure, les droites des différents groupes sont « parallèles » : les pentes sont communes, seuls les intercepts changent. C'est l'intuition du modèle à effets fixes individuels avec pente commune.

4 Tests de spécification de Hsiao (2003)

La question pratique est : *quel modèle choisir ?* Pooling ? Effets fixes ? Effets aléatoires ? Avec pentes communes ou spécifiques ?

Plusieurs critères peuvent guider le choix :
— arguments théoriques et économiques,
— inspection graphique,
— tests économétriques : ici, les tests de **Hsiao (2003)**.

4.1 Logique générale et arbre de décision

Hsiao (2003) propose une séquence de tests F pour comparer plusieurs spécifications imbriquées. On considère, pour simplifier, le cas d'une seule variable explicative X_{it} et trois grandes familles de modèles :

$$\begin{aligned} \text{(1) Pooled :} \quad & Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}, \\ \text{(2) Effets fixes individuels :} \quad & Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}, \\ \text{(3) Pentes et intercepts spécifiques :} \quad & Y_{it} = \alpha_i + \beta_i X_{it} + \varepsilon_{it}. \end{aligned}$$

Les tests reposent sur la comparaison des **sommes des carrés des résidus** (SCR) entre modèles contraints et non contraints.

4.2 Étape 1 : homogénéité complète (test F_1)

On compare :
— **Modèle non contraint (Mnc)** : $Y_{it} = \alpha_i + \beta_i X_{it} + \varepsilon_{it}$;
— **Modèle contraint (Mc)** : $Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}$.
Hypothèses :

$$H_0^1 : \alpha_i = \alpha \text{ et } \beta_i = \beta \ \forall i \quad \text{(homogénéité complète)},$$

$$H_1^1 : \exists (i,j) \text{ tels que } \alpha_i \neq \alpha_j \text{ ou } \beta_i \neq \beta_j.$$

Notons :
— SCR_1 : somme des carrés des résidus du modèle non contraint (Mnc) ;
— $SCR_{1,c}$: somme des carrés des résidus du modèle contraint (Mc).
La statistique de test est :

$$F_1 = \frac{(SCR_{1,c} - SCR_1)/[(n-1)(K+1)]}{SCR_1/[nT - n(K+1)]}, \tag{6}$$

où n est le nombre d'unités (pays), T le nombre de périodes et K le nombre de variables explicatives.

On rejette H_0^1 si F_1 est trop grand par rapport à la loi F avec $[(n-1)(K+1)]$ et $[nT-n(K+1)]$ degrés de liberté.

4.3 Étape 2 : pentes communes ou non ? (test F_2)

Si H_0^1 est rejetée, on sait que le modèle pooled est trop restrictif. On se demande alors si au moins les **pentes** peuvent être communes.

On compare :
— **Modèle non contraint (Mnc)** : $Y_{it} = \alpha_i + \beta_i X_{it} + \varepsilon_{it}$;
— **Modèle contraint (Mc)** : $Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$ (effets fixes individuels, pente commune).

Hypothèses :

$$H_0^2 : \beta_i = \beta \ \forall i \quad \text{(pentes communes)},$$
$$H_1^2 : \exists (i,j) \text{ tels que } \beta_i \neq \beta_j.$$

Notons :
— SCR_1 : SCR du modèle non contraint (le même que pour F_1) ;
— $SCR_{1,c'}$: SCR du modèle contraint avec pentes communes et intercepts spécifiques.
La statistique est :

$$F_2 = \frac{(SCR_{1,c'} - SCR_1)/[(n-1)K]}{SCR_1/[nT - n(K+1)]}. \tag{7}$$

On rejette H_0^2 si F_2 est trop grand.

4.4 Étape 3 : intercepts communs ou effets fixes ? (test F_3)

Supposons maintenant que les pentes communes ne sont pas rejetées (H_0^2 vraie). On compare alors :
— **Modèle non contraint (Mnc)** : $Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$;
— **Modèle contraint (Mc)** : $Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}$.
Hypothèses :

$$H_0^3 : \alpha_i = \alpha \ \forall i \quad \text{(pas besoin d'effets fixes)},$$
$$H_1^3 : \exists (i,j) \text{ tels que } \alpha_i \neq \alpha_j.$$

On note :
— SCR_1 : SCR du modèle non contraint (effets fixes individuels) ;
— $SCR_{1,c''}$: SCR du modèle contraint (pooled avec pente commune).
La statistique est :

$$F_3 = \frac{(SCR_{1,c''} - SCR_1)/(n-1)}{SCR_1/[n(T-1) - K]}. \tag{8}$$

4.5 Illustration du processus

Si H_0^3 est rejetée, on conclut que les effets fixes individuels sont nécessaires : le modèle retenu est alors, en général,

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}.$$

5 Application : climat et PIB agricole

On applique maintenant ces idées à un exemple stylisé inspiré de ton projet.

5.1 Spécification économétrique

On considère le modèle suivant :

$$\log(\text{PIB agri}_{it}) = \alpha_i + \gamma_t + \beta_1 \,\text{degdays}_{it} + \beta_2 \,\text{rain}_{it} + \beta_3 \,\text{hotdays}_{it} + \varepsilon_{it}, \tag{9}$$

où :
— α_i : niveau moyen de productivité agricole propre au pays i ;
— γ_t : choc commun à tous les pays en année t (prix mondiaux, tendance technologique, etc.) ;
— β_1 : effet marginal de la chaleur utile (degrés-jours) sur le PIB agricole ;
— β_2 : effet des précipitations de saison de croissance ;
— β_3 : effet du nombre de jours très chauds (attendu négatif).
On peut procéder comme suit :

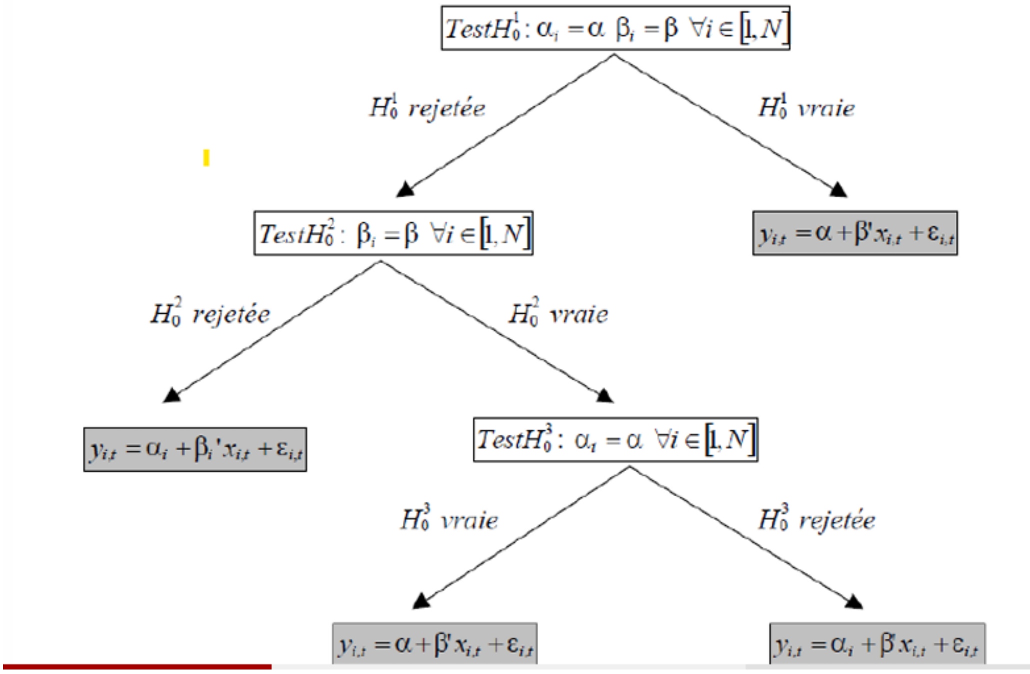

FIGURE 2 – Illustration du processus.

1. Estimer un modèle pooled (sans α_i ni γ_t) et un modèle à pentes et intercepts spécifiques pour une seule variable (par exemple degdays$_{it}$) afin de calculer les tests F_1, F_2, F_3 à la Hsiao.
2. En pratique, sur des données climatiques et agricoles, on s'attend à ce que :
   — le pooled soit rejeté (F_1 significatif),
   — les pentes communes ne soient pas forcément rejetées (les pays réagissent de façon similaire aux variations climatiques),
   — les intercepts communs soient rejetés (F_3 significatif), ce qui justifie l'utilisation d'un **modèle à effets fixes individuels**.
3. On retient alors le modèle (9) estimé par **effets fixes pays + effets fixes temporels**.

5.2 Interprétation des coefficients dans ce contexte

Dans le modèle (9) estimé en FE :

— β_1 mesure l'effet d'une augmentation de la chaleur utile (par exemple +100 degrés-jours) sur le PIB agricole, *à pays donné* et *année donnée*. Si $\beta_1 > 0$, une hausse modérée des degrés-jours est bénéfique à la production.

— β_2 traduit l'effet d'une augmentation des précipitations en saison de croissance. Trop peu de pluie est défavorable, mais trop de pluie peut aussi être nuisible : on peut donc envisager un terme quadratique si besoin.

— β_3 capte l'effet des journées *trop chaudes*. On s'attend souvent à un coefficient négatif : plus de jours dépassant un seuil critique réduit les rendements.

Les effets fixes α_i ne sont pas interprétés individuellement, mais ils permettent de neutraliser des facteurs structurels inobservés (sols, institutions, politiques agricoles de long terme, etc.).

5.3 Implémentation pratique en R (avec `plm`)

Voici un exemple de code R (à adapter à ta vraie base de données) pour estimer le modèle FE et réaliser un test de Hausman entre FE et RE.

```r
library(plm)

# On suppose que 'panel' contient :
# country, year, degdays, rain, hotdays, agri_gdp

panel$log_agri_gdp <- log(panel$agri_gdp)

# D claration du panel
pdata <- pdata.frame(panel, index = c("country", "year"))

# Mod le pooled
pool_mod <- plm(
  log_agri_gdp ~ degdays + rain + hotdays,
  data  = pdata,
  model = "pooling"
)

# Mod le  effets fixes pays + ann es
fe_tw <- plm(
  log_agri_gdp ~ degdays + rain + hotdays,
  data  = pdata,
  model = "within",
  effect = "twoways"
)

# Mod le  effets al atoires
re_mod <- plm(
  log_agri_gdp ~ degdays + rain + hotdays,
  data  = pdata,
  model = "random",
  effect = "individual"
)

summary(fe_tw)

# Test de Hausman : FE vs RE
phtest(fe_tw, re_mod)
```

Le test de Hausman fournit une autre manière de choisir entre FE et RE : si la p-value est faible, on rejette RE au profit de FE, ce qui est cohérent avec l'idée que les effets pays sont corrélés aux variables climatiques.

Conclusion

Ce cours a introduit les concepts de base des modèles statiques en données de panel : pooling, effets fixes, effets aléatoires, ainsi que les tests de spécification de Hsiao (2003). L'exemple climat–PIB agricole illustre comment ces outils peuvent être utilisés pour analyser l'impact des degrés-jours, des précipitations et des journées très chaudes sur l'activité agricole de plusieurs pays au cours du temps.

Dans un travail appliqué, il est important de combiner :

— une réflexion économique (quelle forme fonctionnelle est plausible ?) ;
— une inspection des données (graphes, statistiques descriptives) ;
— et des tests économétriques (Hsiao, Hausman, tests de spécification, tests de résidus).