

Econométrie Panel - TD2

Fadli Aaron - Université de Tours — Master 1 (2025-2026)

Exercice 1

L'objectif de ce TD est de comparer les résultats des estimations de différents modèles possibles avec des données de panel.

On utilise la base de données **WAGEPAN**, qui contient des informations sur les salaires et les caractéristiques de **545 hommes** sur la période **1980–1987**.

Nous considérons l'équation de salaire suivante :

$$\log(pwage_{it}) = \alpha + \beta_1 educ_i + \beta_2 exper_{it} + \beta_3 expersq_{it} + \beta_4 married_{it} + \beta_5 union_{it} + \varepsilon_{it}$$

```
library(tidyverse)
library(haven)
library(plm)
library(lmtest)
library(stargazer)
library(sandwich)
```

```
WagePan <- read.csv("data/Wagedata.csv")
```

1. Intérêt d'avoir à la fois *exper* et *expersq* en variables explicatives

Inclure simultanément **exper** (l'expérience) et **expersq** (l'expérience au carré) permet de modéliser une relation non linéaire entre l'expérience et le salaire.

En effet :

exper capte l'effet marginal initialement positif de l'expérience sur le salaire : un an d'expérience supplémentaire augmente le salaire.

expersq, qui est généralement négatif, permet d'introduire une courbure dans la relation : l'effet marginal de l'expérience décroît au fur et à mesure que l'expérience augmente.

Autrement dit, on modélise l'idée que :

- au début de la carrière, chaque année d'expérience augmente fortement le salaire ;
- mais plus tard, ces gains deviennent plus faibles : rendements décroissants.

Sans **expersq**, on imposerait une relation strictement linéaire, ce qui serait irréaliste dans la plupart des modèles de capital humain.

2. Signe attendu de chaque coefficient

On se place dans le cadre "classique" de l'équation de salaire, les signes attendus sont les suivants :

- Pour β_1 : $\beta_1 > 0 \rightarrow$ Plus d'années d'éducation augmentent le salaire (capital humain plus élevé).

- Pour β_2 : $\beta_2 > 0 \rightarrow$ L'expérience augmente la productivité, donc le salaire.
- Pour β_3 : $\beta_3 < 0 \rightarrow$ Rendements décroissants de l'expérience : le salaire continue d'augmenter avec l'expérience, mais de moins en moins vite (concavité).
- Pour β_4 : $\beta_4 > 0 \rightarrow$ Les hommes mariés ont en moyenne un salaire plus élevé (effet de spécialisation, stabilité, sélection, etc.).
- Pour β_5 : $\beta_5 > 0 \rightarrow$ Prime syndicale : les travailleurs syndiqués obtiennent en général de meilleurs salaires (pouvoir de négociation).

3. Estimation du modèle par les MCO

```
data("WagePan", package = "plm")
pooled <- lm(lwage ~ educ + exper + expersq + married + union, data = WagePan)

#summary(pooled)
#coeftest(pooled)

stargazer(pooled,
           float = TRUE,
           table.placement = "H",
           title = "Pooled Model")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: dim., janv. 04, 2026 - 16:36:39

Table 1: Pooled Model

	<i>Dependent variable:</i>
	lwage
educ	0.099*** (0.005)
exper	0.086*** (0.010)
expersq	-0.003*** (0.001)
married	0.123*** (0.016)
union	0.169*** (0.017)
Constant	-0.034 (0.063)
Observations	4,360
R ²	0.179
Adjusted R ²	0.178
Residual Std. Error	0.483 (df = 4354)
F Statistic	189.857*** (df = 5; 4354)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

L'estimation du modèle pooled OLS montre que toutes les variables explicatives ont l'effet attendu sur le salaire. Le coefficient de l'éducation est positif et significatif : une année supplémentaire d'études augmente le salaire horaire d'environ 9,9 %, ce qui confirme l'importance du capital humain. L'expérience a également un effet positif (environ 8,6 % par année), mais son carré est négatif, ce qui traduit des rendements décroissants : les premières années d'expérience augmentent fortement le salaire, puis l'effet marginal diminue avec le temps.

Les caractéristiques individuelles jouent aussi un rôle important. Être marié est associé à un salaire environ 12,3 % plus élevé, ce qui peut s'expliquer par des effets de spécialisation ou de sélection. De même, les travailleurs syndiqués perçoivent un salaire environ 16,9 % plus élevé que les non-syndiqués, ce qui reflète une prime syndicale liée au pouvoir de négociation.

Dans l'ensemble, les résultats sont cohérents avec les prédictions théoriques : l'éducation, l'expérience et certaines caractéristiques personnelles contribuent significativement à la formation des salaires.

4. Si les erreurs ne sont pas sphériques...

Si les erreurs ne sont pas sphériques, cela signifie que certaines hypothèses fondamentales des MCO ne sont pas respectées. En particulier, les erreurs peuvent présenter de l'hétéroscédasticité (variance non constante) ou de l'autocorrélation, ce qui est très probable dans un contexte de données de panel.

Dans ce cas, les estimateurs des coefficients restent non biaisés et convergents, mais leurs erreurs-types ne sont plus valides : elles sont incorrectement estimées et conduisent à des tests statistiques trompeurs.

Concrètement, les statistiques t ne sont plus fiables, les intervalles de confiance sont mal calculés et l'on risque de conclure à la significativité de certains coefficients alors qu'ils ne le sont pas, ou l'inverse.

Pour corriger ce problème, on utilise des estimateurs d'erreurs-types robustes, comme ceux proposés par White ou, dans le cas de données de panel, ceux de type Arellano, qui tiennent compte à la fois de l'hétéroscédasticité et de la corrélation intra-individu. Ces corrections permettent de retrouver des tests statistiques valides sans modifier les coefficients estimés.

5. Correction des écarts-types

En corrigeant les erreurs-types :

```
# Etape 1 : Estimation du modèle à effets poolé
# =====
pooled_corr <- plm(lwage ~ educ + exper + expersq + married + union,
                  data = WagePan,
                  model = "pooling",
                  index = c("nr", "year"))

#Etape 2 : Matrice de variance-covariance robuste
# =====
cov1 <- vcovHC(pooled_corr, method = "arellano")

#Etape 3 : Extraction des erreurs standard robustes
# =====
robust_se1 <- sqrt(diag(cov1))

#Etape 4 : Tableau de résultats avec stargazer
# =====
stargazer(pooled_corr,
          title = "Pooled Model with standard error correction",
          float = TRUE,
          table.placement = "H",
          se = list(robust_se1))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: dim., janv. 04, 2026 - 16:36:39

Table 2: Pooled Model with standard error correction

	<i>Dependent variable:</i>
	lwage
educ	0.099*** (0.009)
exper	0.086*** (0.013)
expersq	-0.003*** (0.001)
married	0.123*** (0.026)
union	0.169*** (0.028)
Constant	-0.034 (0.115)
Observations	4,360
R ²	0.179
Adjusted R ²	0.178
F Statistic	189.857*** (df = 5; 4354)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Après application de la correction des erreurs-types selon la méthode d’Arellano, on observe que les coefficients du modèle restent rigoureusement identiques à ceux obtenus avec les erreurs-types classiques. Cela est normal : la correction robuste n’affecte pas les estimateurs eux-mêmes, mais uniquement l’estimation de leur variabilité.

En revanche, les erreurs-types augmentent de manière notable pour plusieurs variables, notamment pour l’expérience, le statut marital et l’appartenance syndicale. Cette hausse reflète le fait que la méthode d’Arellano corrige à la fois l’hétéroscédasticité et la corrélation intra-individu, qui sont très fréquentes dans les données de panel. En conséquence, les intervalles de confiance deviennent légèrement plus larges.

Malgré cette correction plus exigeante, tous les coefficients restent hautement significatifs. Cela signifie que les relations mises en évidence précédemment — effet positif de l’éducation et de l’expérience (avec rendements décroissants), prime salariale pour les individus mariés et pour les travailleurs syndiqués — demeurent robustes même lorsque l’on tient compte de la structure de panel et des éventuelles violations des hypothèses classiques des MCO.

En résumé, la correction robuste confirme la solidité statistique des résultats obtenus : les interprétations économiques ne changent pas et les conclusions du modèle sont stables.

6. Modèle à erreurs composées

Dans le cadre des données de panel, le modèle à erreurs composées décompose l’erreur en deux parties :

$$\varepsilon_{it} = u_i + v_{it},$$

où u_i désigne l'hétérogénéité individuelle non observée, constante dans le temps, et v_{it} la composante idiosyncratique qui varie selon les individus et les périodes. Cette structure permet de prendre en compte le fait que chaque individu possède des caractéristiques propres non observées qui influencent son salaire (talent, motivation, capacité, caractéristiques familiales, etc.).

Ce point est particulièrement important pour la variable **married**, qui est susceptible d'être corrélée avec cette hétérogénéité individuelle non observée. Par exemple, les individus qui ont une plus grande stabilité professionnelle, une meilleure aptitude au travail ou un niveau d'organisation plus élevé sont à la fois davantage susceptibles de se marier et d'obtenir des salaires plus élevés. Ces caractéristiques non observées seraient alors capturées par la composante u_i .

Si la variable married est corrélée avec u_i , alors le modèle Pooled OLS ou le modèle à effets aléatoires ne sont plus appropriés, car ils reposent sur l'hypothèse que les variables explicatives sont indépendantes de u_i . Dans ce cas, l'estimation est biaisée et l'effet estimé du mariage intègre non seulement son impact propre, mais aussi des caractéristiques individuelles non observées.

Ainsi, le modèle à erreurs composées met en évidence un enjeu important : la variable married peut être **endogène** au sens où elle est liée aux effets individuels non observés. Cela explique pourquoi les estimations Pooled ou Random Effects doivent être interprétées avec prudence pour cette variable. En revanche, un modèle à effets fixes permettrait de neutraliser u_i et d'identifier uniquement l'effet du changement de statut marital au cours du temps.

7. Effets aléatoires

```
# Etape 1 : Estimation du modèle à effets aléatoires
# =====
re_model <- plm(lwage ~ educ + exper + expersq + married + union,
               data = WagePan, model = "random", effect = "individual",
               index = c("nr", "year"))

#Etape 2 : Matrice de variance-covariance robuste
# =====
cov2 <- vcovHC(re_model, method = "arellano")

#Etape 3 : Extraction des erreurs standard robustes
# =====
robut_se2 <- sqrt(diag(cov2))

#Etape 4 : Tableau de résultats avec stargazer
# =====
stargazer(re_model,
          float = TRUE,
          table.placement = "H",
          title = "Random Effect Model with standard error correction",
          se = list(robut_se2))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: dim., janv. 04, 2026 - 16:36:40

Table 3: Random Effect Model with standard error correction

	<i>Dependent variable:</i>
	lwage
educ	0.101*** (0.009)
exper	0.111*** (0.011)
expersq	-0.004*** (0.001)
married	0.067*** (0.019)
union	0.104*** (0.021)
Constant	-0.119 (0.110)
Observations	4,360
R ²	0.176
Adjusted R ²	0.175
F Statistic	932.127***
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

L'estimation du modèle à effets aléatoires avec erreurs-types robustes confirme dans l'ensemble les résultats obtenus précédemment. Les coefficients restent très proches du modèle non corrigé, ce qui est normal car la correction robuste n'affecte pas les estimateurs, seulement la précision de leur estimation.

L'éducation exerce toujours un effet positif et significatif sur le salaire : une année supplémentaire d'études augmente le salaire horaire d'environ 10,1 %. L'expérience a également un impact positif d'environ 11,1 %, tandis que son carré est négatif, ce qui met en évidence la présence de rendements décroissants de l'expérience, comme attendu dans les modèles de capital humain.

Les caractéristiques individuelles jouent aussi un rôle important. Être marié est associé à un salaire environ 6,7 % plus élevé, tandis que l'appartenance syndicale conduit à un salaire supérieur d'environ 10,4 %. Ces effets sont tous significatifs avec les erreurs-types robustes, ce qui confirme leur robustesse même en tenant compte de l'hétéroscédasticité et de la corrélation intra-individu.

Dans l'ensemble, la correction robuste ne modifie pas les conclusions économiques : les effets estimés sont cohérents, significatifs et conformes aux prédictions théoriques. Le modèle à effets aléatoires apparaît donc stable et fiable après correction des erreurs-types.

8. Comparaison RE / OLS

9. Effets fixes

```
# Etape 1 : Estimation du modèle à effets fixes
# =====
fe_model <- plm(lwage ~ educ + exper + expersq + married + union,
```

```

data = WagePan, model = "within", effect = "individual",
index = c("nr", "year"))

#Etape 2 : Matrice de variance-covariance robuste
# =====
cov3 <- vcovHC(fe_model, method = "arellano")

#Etape 3 : Extraction des erreurs standard robustes
# =====
robut_se3 <- sqrt(diag(cov3))

#Etape 4 : Tableau de résultats avec stargazer
# =====
stargazer(fe_model,
float = TRUE,
table.placement = "H",
title = "Fixed Effect Model with standard error correction",
se = list(robut_se3))

```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: dim., janv. 04, 2026 - 16:36:40

Table 4: Fixed Effect Model with standard error correction

	<i>Dependent variable:</i>
	lwage
exper	0.117*** (0.011)
expersq	-0.004*** (0.001)
married	0.045** (0.021)
union	0.082*** (0.023)
Observations	4,360
R ²	0.178
Adjusted R ²	0.060
F Statistic	206.375*** (df = 4; 3811)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

10. Coefficients non estimés ?

Dans le modèle à effets fixes, certains coefficients ne peuvent pas être estimés car les variables qui ne varient pas dans le temps pour un individu sont parfaitement colinéaires avec les effets individuels. Les effets fixes capturent en effet toute la composante invariante dans le temps, de sorte que toute variable explicative qui ne change pas d'une période à l'autre ne peut plus être identifiée séparément.

C'est exactement ce qui se produit pour la variable educ : comme le niveau d'éducation d'un individu est

constant sur toute la période, il est absorbé par l'effet individuel fixe u_i . Le modèle ne dispose alors d'aucune variation intra-individu pour identifier son impact, et le coefficient associé n'est donc pas estimé.

Plus généralement, toute variable strictement constante pour chaque individu — qu'il s'agisse du sexe, de l'origine ethnique, du niveau d'éducation ou de toute autre caractéristique fixe — ne peut pas apparaître dans un modèle within. Cela ne signifie pas que ces variables n'ont aucun effet sur le salaire, mais simplement que l'estimation par effets fixes ne permet pas de les distinguer de l'effet individuel spécifique.

11. Test de Hausman

```
phptest(fe_model, re_model, data = WagePan, method = "aux", vcov = vcovHC)
```

Hausman Test

data: lwage ~ educ + exper + expersq + married + union chisq = 33.712, df = 4, p-value = 8.537e-07
alternative hypothesis: one model is inconsistent