

## Les techniques de modélisation des données de panel Application à l'impact de la dette publique sur les recettes fiscales: cas des pays du Maghreb Arabe

### Panel Data Modeling Techniques Application to the impact of public debt on tax revenue: case Arab Maghreb countries

تطبيق تقنيات نمذجة البيانات لأثر الدين العام على الإيرادات الجبائية: حالة دول المغرب العربي

Dr. Talem Ali<sup>1</sup>

Université Ibn-Khaldoun -Tiaret- Algérie

ali.talem@univ-tiaret.dz

Dr. Boulabbas Mokhtar

Université Ibn-Khaldoun -Tiaret- Algérie

mokhtar.boulabbas@univ-tiaret.dz

Received: 14/12 /2022

Accepted: 09/06/2023

Published: 11/06/2023

#### Résumé

L'objectif principal de cette étude est de présenter les différents tests de spécification utilisés pour déterminer le modèle théorique approprié pour modéliser les données de panel. Ces tests ont été appliqués sur les données relatives aux dettes publiques et recettes fiscales des trois pays du Maghreb Arabe, afin de savoir si l'accumulation de la dette publique exerce un effet sur la mobilisation des recettes fiscales dans ces trois pays.

L'étude a abouti que le modèle à effets individuels fixes est le modèle approprié, l'estimation de ce dernier montre l'impact négatif de l'endettement sur les recettes fiscales de ces pays..

**Mots clés :** données de panel, données longitudinales, modèle à effets individuels fixes, modèle à effets individuels aléatoires, homogénéité, hétérogénéité.

#### Abstract

The objective of the study is to present the different specification tests used to determine the appropriate model for modeling panel data. These tests were applied to data on public debt and tax revenue from the three countries of the Arab Maghreb, in order to determine whether the accumulation of public debt has an effect on the mobilization of tax revenue in those countries.

The study concluded that the fixed individual effects model is the appropriate one, the estimate of the latter shows a negative impact of indebtedness on the tax revenues of these countries.

**Key words:** panel data, longitudinal data, fixed effects model, random effects model, homogeneity, heterogeneity.

ملخص:

موضوع هذه الدراسة هو تقديم مختلف اختبارات التحديد المستخدمة لتحديد النموذج النظري الملائم و المناسب لنمذجة معطيات بانل. هذه الاختبارات تم تطبيقها على المعطيات المتعلقة بالديون العمومية و الإيرادات الجبائية للدول المركزية للمغرب العربي وهذا بغرض معرفة فيما اذا كان لتراكم الديون العمومية اثر على الإيرادات الجبائية على مستوى هذه الدول الثلاث.

الدراسة توصلت الى ان نموذج التأثيرات الفردية الثابتة هو النموذج المناسب و الملائم، تقدير هذا الاخير اوضح وجود علاقة عكسية بين كل من الديون العمومية و الإيرادات الجبائية لهذه الدول.

<sup>1</sup> - Corresponding author: Talem Ali :ali.talem@univ-tiaret.dz

## 1. Introduction

Les données statistiques sont présentées sous trois formes principales, soit sous la forme de séries chronologiques (temporelles) qui possèdent une dimension temporelle notée  $t$ , et décrivent le comportement d'un seul individu au cours d'un ensemble de périodes. Donc elles négligent l'effet du changement de comportement de la variable étudiée d'un individu à l'autre. Elles considèrent ainsi que tous les individus se comportent de la même manière vis-à-vis d'un événement, c-à-d. que les individus ont le même comportement, ce qui laisse supposer leur homogénéité. La deuxième forme est les données en coupe instantanée (transversales), qui possèdent une dimension individuelle notée  $i$ , et elles décrivent le comportement d'un ensemble d'individus à une date donnée, donc elles négligent l'effet de la variation des valeurs des variables d'un instant à l'autre et supposent que le comportement des individus ne change pas dans le temps. Enfin la troisième forme est celle qui possède les deux dimensions spatiales et temporelle, c-à-d. une combinaison des données en séries temporelles et des données transversales. Ce troisième type de données, décrit le comportement d'un ensemble d'individus au cours d'un ensemble de périodes de temps, c'est-à-dire il combine à la fois les caractéristiques et les spécificités des deux formes de données citées précédemment, et c'est là que réside l'importance d'utiliser ce type de données, car elles contiennent des informations nécessaires qui prennent en compte la dynamique du temps et la multiplicité des individus. Ces données contiennent alors les deux effets. Il est à noter que les appellations de ce type de données sont multiples et identiques: les données croisées, les données combinées, les données longitudinales, En utilisant les données de panel, on pourra exploiter les deux sources de variation de l'information statistique: Temporelle ou variabilité intra-individuelle et individuelle ou variabilité inter-individuelles. Comme les autres type de données, celles de panel seront utilisées dans le processus de modélisation, donc la question principale qui se pose et qui forme la problématique de ce travail est la suivante :

### **Quelles sont les différentes techniques utilisées pour déterminer le modèle théorique approprié pour modéliser les données de panel ?**

La réponse à cette question est l'objectif de ce travail, qui est subdivisé en cinq sections. A travers la première, nous avons présenté les principaux modèles des données de panel. Dans la seconde section nous avons exposé une synthèse des principaux tests de spécifications des effets individuels (Hsiao et Hausman). A travers la troisième section nous avons appliqué ces tests de spécification sur les données de panel relatives aux dettes publiques et aux recettes fiscales des trois pays du Maghreb Arabe (l'Algérie, le Maroc et la Tunisie), afin de savoir si l'accumulation de la dette publique a un impact sur la mobilisation des recettes fiscales dans ces pays. Cela nous amène à discuter les résultats à partir de la quatrième section et tirer quelques remarques et conclusions.

## 2. Les principaux modèles des données de panel

Si nous disposons d'un échantillon de données statistiques de taille  $n = N \times T$  sur  $N$  individus ou unités statistique pendant et durant  $T$  périodes de temps, et afin de prendre en compte la différence et l'hétérogénéité dans le temps et entre les individus de l'échantillon, le modèle en données de panel s'écrit comme un modèle à double indice qui prend la forme suivante: <sup>1</sup>

$$Y_{it} = \beta_{0it} + \beta_{1it} X_{1it} + \beta_{2it} X_{2it} + \beta_{3it} X_{3it} + \dots + \beta_{jit} X_{jit} + \dots + \beta_{kit} X_{kit} + \varepsilon_{it}$$
$$i = 1, \dots, N, \quad t = 1, \dots, T$$

$Y_{it}$  : représente la variable endogène.

$X_{jit}$  : les variables explicatives  $j = 1, \dots, k$

$\beta_{0it}$  : l'intercepte ou le terme constant

$\beta_{jit}$  : les coefficients des variables explicatives ou les coefficients de régression ou les coefficients de pente

$\varepsilon_{it}$  : le terme aléatoire

On remarque que les coefficients  $\beta_{0it}$  et  $\beta_{jit}$  avec  $j = 1, \dots, k$  changent en même temps par rapport au temps et entre individus, c'est-à-dire le comportement de ces individus n'est pas homogène, en plus il change à travers le temps. Le nombre de coefficients du modèle ci-dessus est égal à  $NT(K+1)$  est supérieur au nombre d'équations, qui est égal à  $NT$ , donc il est nécessaire de mettre et d'imposer un ensemble de restrictions et contraintes sur ces coefficients afin de pouvoir les estimer. Ces contraintes sont que le terme constant du modèle  $\beta_{0it}$  et les coefficients des variables explicatives  $\beta_{jit}$  ne changent plus à travers le temps,<sup>2</sup> ceci se traduit par l'écriture suivante :

$$\beta_{0it} = \beta_{0i} \text{ et } \beta_{jit} = \beta_{ji}$$

Suivant les contraintes imposées, le modèle ci-dessus prendras la forme suivante:

$$Y_{it} = \beta_{0i} + \beta_{1i}X_{1it} + \beta_{2i}X_{2it} + \beta_{3i}X_{3it} + \dots + \beta_{ji}X_{jit} + \dots + \beta_{ki}X_{kit} + \varepsilon_{it},$$

Ce modèle représente la forme générale du modèle linéaire simple des données de panel. A partir de cette forme nous distinguons quatre types de modèles que nous pouvons utiliser pour modéliser les données de panel.

1. Modèle à effets individuels fixes
2. Modèle à effets individuels aléatoires
3. Modèle multiple pour la totalité des individus
4. Modèle multiple pour chaque individu

### 2.1. Le Modèle à effets individuels fixes :

Le modèle à effets individuels connu aussi sous le nom de LSDV (Least Squares Dummy Variables) suppose que le terme constant est différent d'un individu à l'autre, en plus il repose sur l'hypothèse que les coefficients des variables explicatives sont identiques et égaux entre les individus  $\beta_{ji} = \beta_j \forall i, j$ , c'est-à-dire l'effet marginal des variables explicatives sur la variable dépendante est constant pour chaque unité au sein de la section transversale, cela implique que les relations entre la variable endogène et les variables exogènes sont identiques pour tous les individus quel que soit la période de temps.<sup>3</sup> Si l'on dispose de  $N$  individus observés sur  $T$  périodes de temps et  $k$  variables explicatives, le modèle à effets individuels fixes s'écrit alors :

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \dots + \beta_j X_{jit} + \dots + \beta_k X_{kit} + \varepsilon_{it},$$

Avec  $\beta_{0i}$  est le terme constant spécifique à l'individu  $i$ , il représente la spécification individuelle supposée fixe à travers le temps. Nous signalons que le modèle à effets individuels s'agit en fait d'un modèle classique avec variables indicatrices individuelles.

### 2.2. Le Modèle à effets individuels aléatoires

Le modèle à effets individuels aléatoires connu aussi sous le nom de modèle à erreurs composées ou à composantes d'erreur, suppose que la spécificité individuelle est sous forme aléatoire, c-a-d le terme constant spécifique à l'individu  $i$  est stochastique. Ce terme se décompose en un terme fixe noté  $\beta_0$  qui désigne la composante fixe, plus un terme aléatoire spécifique à l'individu noté  $v_i$  qui désigne la composante aléatoire individuelle, permettant de contrôler l'hétérogénéité individuelle.<sup>4</sup> En regroupant les termes du modèle, nous obtenons une structure à erreurs composées. Le modèle à effets individuels aléatoire s'écrit sous la forme suivante :

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \dots + \beta_j X_{jit} + \dots + \beta_k X_{kit} + (v_i + \varepsilon_{it}),$$

### 3. Les tests de spécification

Lorsqu'on dispose d'un échantillon de données de panel, et avant de procéder à la modélisation, il apparaît nécessaire d'appliquer les tests de spécification pour déterminer le modèle théorique approprié, en examinant la présence du phénomène d'homogénéité ou non dans le processus générateur de ces données. Le phénomène d'homogénéité signifie que le comportement des individus est identique, c'est-à-dire la similitude et la congruence des comportements des individus de l'échantillon, ce qui signifie que le modèle théorique étudié est le même pour tous ces individus. Cela est du point de vue économique. Du point de vue économétrique, l'homogénéité signifie que les coefficients du modèle théorique étant égaux dans sa dimension transversale ou individuelle. A partir de la forme générale du modèle linéaire simple des données de Panel, on peut distinguer quatre cas possibles:<sup>5</sup>

1. Homogénéité totale.
2. Hétérogénéité totale
3. Homogénéité des termes constants et hétérogénéité des coefficients des variables explicatives
4. Hétérogénéité des termes constants et homogénéité des coefficients des variables explicatives

**Cas n°1: Homogénéité totale :** Dans ce cas nous avons:

- Les  $N$  constantes sont identiques pour tous les individus, c'est-à-dire la constante  $\beta_{0i}$  est identique et la même pour tous les individus

$$\beta_{01} = \beta_{02} = \beta_{03} = \dots = \beta_{0i} = \dots = \beta_{0N} = \beta_0 \Leftrightarrow \beta_{0i} = \beta_0, \quad i = 1, \dots, N$$

- Les coefficients des variables explicatives  $\beta_{ji}$  sont identiques et égaux pour tous les individus, c'est-à-dire ne changent plus d'un individu à l'autre

$$\beta_{ji} = \beta_j, \quad j = 1, \dots, k, \quad i = 1, \dots, N$$

Sous les deux hypothèses citées ci-dessus on, qualifie alors le panel de panel homogène et le modèle général des données de panel ci-dessus prend la forme suivante:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \dots + \beta_j X_{jit} + \dots + \beta_k X_{kit} + \varepsilon_{it},$$

Dans ce cas on estime un modèle de régression linéaire multiple (le modèle pooled) avec  $n = N \times T$  observations empilées.

**Cas n°2: Hétérogénéité totale:** Dans ce cas nous avons

- Les N constantes sont tous différents pour tous les individus, c'est-à-dire la constante  $\beta_{0i}$  n'est pas identique pour tous les individus, elle change d'un individu à l'autre, c'est-à-dire chaque individu  $i$  a sa propre constante  $\beta_{0i}$ .
- Les coefficients des variables explicatives  $\beta_{ji}$  sont tous différents pour tous les individus, et changent d'un individu à l'autre

Sous les deux hypothèses citées ci-dessus, on rejette la structure panel, et on a donc  $N$  modèles différents c'est-à-dire, chaque individu a son propre modèle, et le modèle général des données de panel ci-dessus prend la forme suivante :

$$Y_{it} = \beta_{0i} + \beta_{1i}X_{1it} + \beta_{2i}X_{2it} + \beta_{3i}X_{3it} + \dots + \beta_{ji}X_{jit} + \dots + \beta_{ki}X_{kit} + \varepsilon_{it} ,$$

Dans ce cas on estime N modèles de régression linéaire multiple, chaque modèle avec  $n = T$  observations.

**Cas n°3: Homogénéité des termes constants et hétérogénéité des coefficients de régression:**

Dans ce cas nous avons

- Les N constantes sont identiques pour tous les individus, c'est-à-dire la constante  $\beta_{0i}$  est identique et la même pour tous les individus  

$$\beta_{01} = \beta_{02} = \beta_{03} = \dots = \beta_{0i} = \dots = \beta_{0N} = \beta_0 \Leftrightarrow \beta_{0i} = \beta_0 , i = 1, \dots, N$$
- Les coefficients des variables explicatives  $\beta_{ji}$  sont tous différents pour tous les individus, et changent d'un individu à l'autre

Sous les deux hypothèses citées ci-dessus, tous les coefficients du modèle, à l'exception des constantes, sont différents selon les individus. On a donc N modèles différents c'est-à-dire pour chaque individu on a son propre modèle et le modèle général des données de panel ci-dessus prend la forme suivante :

$$Y_{it} = \beta_0 + \beta_{1i}X_{1it} + \beta_{2i}X_{2it} + \beta_{3i}X_{3it} + \dots + \beta_{ji}X_{jit} + \dots + \beta_{ki}X_{kit} + \varepsilon_{it} ,$$

Dans ce cas on estime N modèles de régression linéaire multiple, chaque modèle avec un nombre d'observations égal  $n = T$ .

**Cas n°4: Hétérogénéité des termes constants et homogénéité des coefficients de régression: le modèle à effets individuels** Dans ce cas nous avons:

- Les N constantes sont tous différents pour tous les individus, c'est-à-dire la constante  $\beta_{0i}$  n'est pas identique pour tous les individus, elle change d'un individu à l'autre, c'est-à-dire chaque individu  $i$  a sa propre constante  $\beta_{0i}$ .

- Les coefficients des variables explicatives  $\beta_{ji}$  sont identiques pour tous les individus, c'est-à-dire ne changent plus d'un individu à l'autre

$$\beta_{ji} = \beta_j, j = 1, \dots, k, i = 1, \dots, N$$

Sous les deux hypothèses citées ci-dessus on obtient un modèle à effets individuels. et le modèle général des données de panel ci-dessus prend la forme suivante: <sup>6</sup>

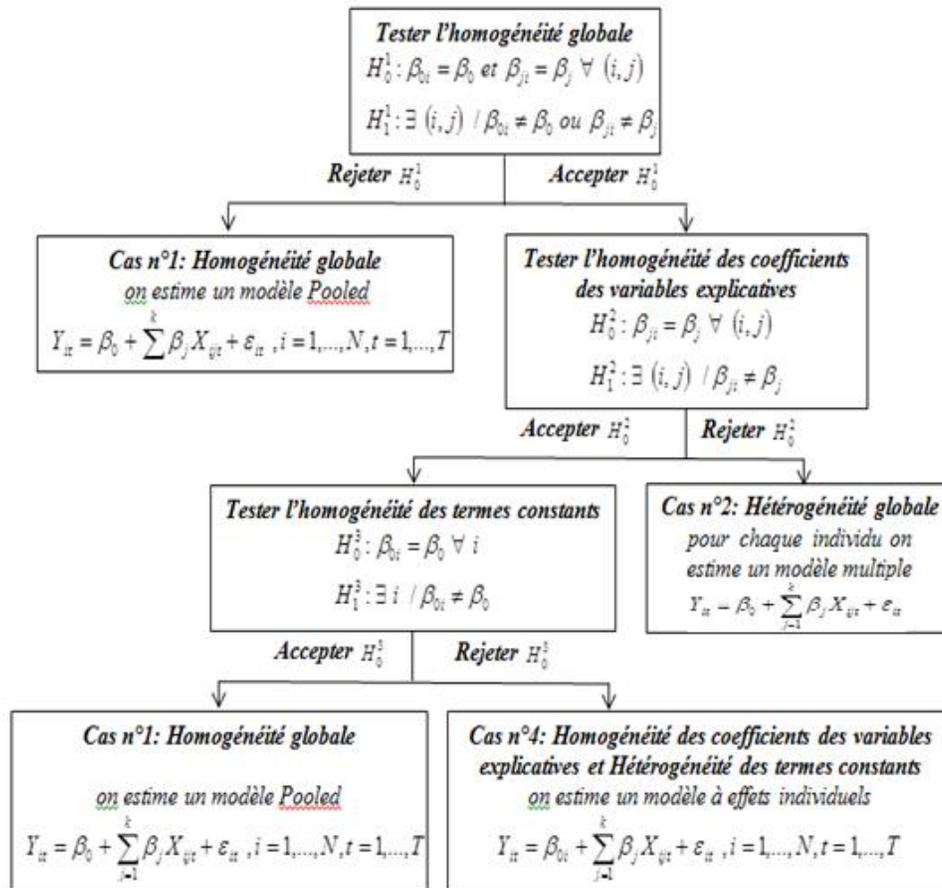
$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \dots + \beta_j X_{jit} + \dots + \beta_k X_{kit} + \varepsilon_{it},$$

Dans ce cas on estime un seul modèle avec  $n = N \times T$  observations.

### 3.1. Le test d'homogénéité de Hsiao

Afin de distinguer les différents cas cités au-dessus et déterminer le modèle théorique de panel approprié pour modéliser les données de panel, nous avons recours au test d'homogénéité proposé par Hsiao (1986). Ce test est une méthode séquentielle basée sur un ensemble d'étapes illustrées et reportées sur la Figure 1 ci-dessous:

Figure N° 1. Stratégie de test d'homogénéité de Hsiao



Source: <sup>7</sup>

Selon la figure 1 ci-dessus, les étapes du test d'homogénéité de Hsiao sont les suivantes

#### a. Première étape : test d'homogénéité complète des comportements des individus

A travers cette étape on teste l'hypothèse nulle dont le contenu est l'homogénéité totale du comportement des individus, contre l'hypothèse alternative l'hétérogénéité globale du comportement de ces individus. Cela se traduit par :

$H_0^1$  : Homogénéité complète des comportements

$H_1^1$  : Hétérogénéité complète des comportements

L'homogénéité complète et parfaite du comportement des individus signifie que les constantes  $\beta_{0i}$  sont égaux et identiques pour tous les individus  $\beta_{0i} = \beta_0, \forall i$ , en plus l'égalité des coefficients des variables explicatives  $\beta_{ji}$  de ces individus  $\beta_{ji} = \beta_j \forall (i, j)$ . Par conséquent les deux hypothèses ci-dessus peuvent être reformulées comme suit:

$H_0^1$  :  $\beta_{0i} = \beta_0$  et  $\beta_{ji} = \beta_j \forall (i, j)$

$H_1^1$  :  $\exists (i, j) / \beta_{0i} \neq \beta_0$  ou  $\beta_{ji} \neq \beta_j$

Pour effectuer le test l'homogénéité totale on utilise un test de Fisher, la statistique de test associée s'écrit sous la forme suivante et suit la loi de Fisher avec  $(k+1) \times (N-1)$  et  $NT - N(k+1)$  degrés de libertés :

$$F_{Cal}^1 = \frac{SCR_{H_0^1} - SCR}{(k+1) \times (N-1)} \bigg/ \frac{SCR}{NT - N(k+1)}$$

Tel que :

- $SCR_{H_0^1}$  : représente la somme des carrés des résidus du modèle contraint par l'hypothèse nulle, c'est-à-dire le modèle linéaire multiple ou le modèle pooled avec un nombre d'observations égal à  $n = N \times T$ .
- $SCR$  : représente la somme des carrés des résidus du modèle non contraint. Elle est égale à la somme des  $N$  sommes des carrés des résidus des  $N$  modèles individuels estimés, soit  $SCR = \sum_{i=1}^N SCR_i$ , c'est-à-dire pour chaque individu on estime un modèle linéaire multiple avec un nombre d'observations égale à  $T$ . Et par conséquent on obtient  $N$  modèle estimés, et pour chacun on a la sommes des carrés des résidus  $SCR_i$ .

La règle de décision et les conclusions de ce test sont les suivantes:

- Si la réalisation de la statistique de test de Fisher calculée  $F_{Cal}^1$  est inférieure ou égale à celle de Fisher tabulée c-à-d.  $F_{Cal}^1 \leq F_{(k+1) \times (N-1), NT - N(k+1)}^{\alpha\%}$ , on accepte l'hypothèse nulle  $H_0^1$  d'homogénéité. On obtient alors un modèle de pooled totalement et globalement homogène.
- Si  $F_{Cal}^1 > F_{(k+1) \times (N-1), NT - N(k+1)}^{\alpha\%}$ , on rejette l'hypothèse nulle, et on passe à la deuxième phase qui consiste à déterminer si l'hétérogénéité provient des coefficients des variables explicatives  $\beta_{ji}$ .

**b. Seconde étape: test d'homogénéité des coefficients des variables explicatives  $\beta_{ji}$**

Cette phase consiste à tester l'homogénéité et l'égalité pour tous les individus des  $k$  coefficients des variables explicatives. Pour cela on teste l'hypothèse nulle  $H_0^2$  contre l'hypothèse alternative  $H_1^2$

$$\begin{cases} H_0^2 : \beta_{ji} = \beta_j \quad \forall (i, j) \\ H_1^2 : \exists (i, j) / \beta_{ji} \neq \beta_j \end{cases}$$

La statistique de test associée s'écrit sous la forme suivante et suit la loi de Fisher avec  $k(N-1)$  et  $NT - N(k+1)$  degrés de libertés :

$$F_{Cal}^2 = \frac{SCR_{H_0^2} - SCR}{SCR} \cdot \frac{k(N-1)}{NT - N(k+1)}$$

Tel que :

- $SCR_{H_0^2}$  : représente la somme des carrés des résidus du modèle contraint par l'hypothèse nulle  $H_0^2$ , c'est-à-dire le modèle à effets individuels fixes.

La règle de décision et les conclusions de ce test sont les suivantes:

- Si  $F_{Cal}^2 > F_{k(N-1), NT-N(k+1)}^{\alpha\%}$ , on rejette l'hypothèse nulle  $H_0^2$  d'homogénéité des coefficients des variables explicatives, et on rejette alors la structure de panel, puisque seules les termes constants  $\beta_{0i}$  peuvent être homogènes et identiques.
- Si  $F_{Cal}^2 \leq F_{k(N-1), NT-N(k+1)}^{\alpha\%}$ , on accepte l'hypothèse nulle, alors on retient la structure de panel dans ce cas, et on passe à la troisième et dernière phase qui consiste à tester l'homogénéité des termes constants  $\beta_{0i}$ .

**c. Troisième étape : test d'homogénéité des termes constants  $\beta_{0i}$**

Cette dernière phase de la procédure de test de Hsiao consiste à tester l'égalité des  $N$  termes constants individuels, sous l'hypothèse d'homogénéité des coefficients des variables explicatives à tous les individus.

$$\begin{cases} H_0^3 : \beta_{0i} = \beta_0 \quad \forall i \\ H_1^3 : \exists i / \beta_{0i} \neq \beta_0 \end{cases}$$

La statistique de test associée s'écrit sous la forme suivante et suit la loi de Fisher avec  $(N-1)$  et  $N(T-1) - k$  degrés de libertés :

$$F_{Cal}^3 = \frac{SCR_{H_0^1} - SCR_{H_0^2} / (N-1)}{SCR_{H_0^2} / N(T-1) - k}$$

La règle de décision et les conclusions de ce test sont les suivantes:

- Si  $F_{Cal}^3 \leq F_{(N-1), N(T-1)-k}^{\alpha\%}$ , on accepte l'hypothèse nulle  $H_0^3$  d'homogénéité des termes constants  $\beta_{0i}$ , alors on retrouve une structure de panel totalement homogène, c'est-à-dire on obtient alors un modèle pooled. Dans ce cas le test de  $H_0^3$  sert à confirmer ou infirmer les résultats et les conclusions obtenus du test de  $H_0^1$ .
- Si  $F_{Cal}^3 > F_{(N-1), N(T-1)-k}^{\alpha\%}$ , on rejette l'hypothèse nulle, alors on obtient un modèle de panel avec effets individuels.

### 3.2. Le test de Hausman : effets individuels fixes ou effets individuels aléatoires

Supposons qu'à travers les tests de Hsiao, l'existence et la présence des spécificités et des effets individuels a été confirmée, c'est-à-dire que chaque individu des individus de l'échantion étudiée a ses propres caractéristiques et spécificités qui lui sont liées, l'étape suivante consiste à connaître et savoir la nature de ces caractéristiques et ces effets individuels, c'est-à-dire, si ces effets individuels sont de type effets fixes ou, au contraire, sont de type effets aléatoires. Afin de répondre à cette question, Hausman (1978) a présenté un test connu sous le nom de test de Hausman, également appelé test de H. Ce test permet de savoir la nature des effets individuels, ce qui permet de choisir entre le modèle à effets individuels fixes et le modèle à effets individuels aléatoires afin de déterminer le modèle théorique final approprié pour modéliser les données de panel. Le test de Hausman repose sur les hypothèses suivantes:<sup>8</sup>

$H_0$  : il n'y a pas de différence systématique des coefficients

$H_1$  : il existe une différence entre les coefficients

Ces deux hypothèses sont équivalentes et similaires aux hypothèses suivantes :

$H_0$  : le modèle à effets individuels aléatoires est le modèle approprié

$H_1$  : le modèle à effets individuels fixes est le modèle approprié

La statistique du test d'Hausman notée  $H$  appliqué au test de la spécification des effets individuels est la suivante :

$$H = (\hat{\beta}_{FEM} - \hat{\beta}_{REM})' [Var(\hat{\beta}_{FEM}) - Var(\hat{\beta}_{REM})]^{-1} (\hat{\beta}_{FEM} - \hat{\beta}_{REM})$$

Tel que :

- $\hat{\beta}_{FEM}$  : les estimateurs des paramètres du modèle à effets individuels fixes
- $\hat{\beta}_{REM}$  : les estimateurs des paramètres du modèle à effets individuels aléatoires

- $Var(\hat{\beta}_{FEM})$ : vecteur des variances des estimateurs des paramètres du modèle à effets individuels fixes
- $Var(\hat{\beta}_{REM})$ : vecteur des variances des estimateurs des paramètres du modèle à effets individuels aléatoires

La règle de décision est la suivante :

- Si la réalisation de la statistique  $H$  est supérieure à la valeur critique  $\chi_{(k)}$ , on rejette l'hypothèse nulle et l'on privilégie l'adoption du modèle à effets individuels fixes.
- Si la réalisation de la statistique  $H$  est inférieure ou égale à la valeur critique  $\chi_{(k)}$ , on accepte l'hypothèse nulle et l'on privilégie l'adoption du modèle à effets individuels aléatoires.

#### 4. Application:

Afin de répondre concrètement à la question émise à travers la problématique, qui est de savoir les différentes techniques utilisées pour déterminer le modèle théorique approprié pour modéliser les données de panel, une application économétrique s'impose. A travers cette partie nous allons appliquer les différents tests de spécification cités plus haut, sur des données de panel relatives aux dettes publiques et aux recettes fiscales des trois pays du Maghreb Arabe (l'Algérie, la Tunisie et le Maroc), afin de savoir si l'accumulation de la dette publique exerce un effet sur la mobilisation des recettes fiscales dans ces trois pays.

##### 4.1 Spécification du modèle: Les tests de spécification:

A travers cette étude nous adapterons un modèle de régression simple retenant une variable endogène, les recettes fiscales notée  $Rec\_Fis$  expliquée par une variable exogène, la dette publique notée  $Det\_Pub$ .

$$Rec\_Fis = f(Det\_Pub)$$

$$Rec\_Fis_{it} = \beta_{0i} + \beta_{1i} Det\_Pub_{it} + \varepsilon_{it}$$

Pour atteindre le modèle économétrique approprié à nos données d'étude, nous faisons recours aux tests de spécification, l'application de ces tests va nous permettre de connaître avec exactitude le modèle théorique le mieux adapté à nos données d'étude. Parmi ces tests, on compte le test d'homogénéité de Hsiao et le test d'Hausman.

##### 4.1.1 Le test de Hsiao:

Par le test d'homogénéité de Hsiao, on cherche à vérifier la présence et l'existence des spécificités et des effets individuels, c'est-à-dire que chaque pays des pays de l'échantillon étudiée a ses propres caractéristiques et spécificités qui lui sont liées.

Pour établir le test d'homogénéité de Hsiao nous utilisons un programme formé d'un ensemble de commandes `evIEWS` qui permet de calculer les valeurs des différentes statistiques de test  $F_1$ ,  $F_2$ ,  $F_3$  et les P-Values associées.

```

scalar r2sum=0
for %i _DZA _TUN _MAR
    equation eqindi.ls rec_fis%i c det_pub%i
    scalar r2sum=r2sum+@ssr
next
panel.ls rec_fis? det_pub?
    
```

```

scalar r2pooled=panel.@ssr
scalar Kreg=panel.@ncoef
panel.ls(f) rec_fis? det_pub?
scalar r2EI=panel.@ssr
scalar Nnobs=panel.@ncross
scalar Tnobs=panel.@regobs/panel.@ncross
scalar F1=(r2pooled-R2sum)/r2sum/(((Nnobs-1)*(Kreg+1))*(Nnobs*Tnobs-Nnobs*(Kreg+1)))
scalar pvalueF1=@fdist(F1,(((Nnobs-1)*(Kreg+1)),(Nnobs*Tnobs-Nnobs*(Kreg+1))))
scalar F2=(r2ei-R2sum)/r2sum/(((Nnobs-1)*Kreg)*(Nnobs*Tnobs-Nnobs*(Kreg+1)))
scalar pvalueF2=@fdist(F2,(((Nnobs-1)*Kreg),(Nnobs*Tnobs-Nnobs*(Kreg+1))))
scalar F3=(r2pooled-r2ei)/r2ei/((Nnobs-1)*(Nnobs*(Tnobs-1)-Kreg))
scalar pvalueF3=@fdist(F3,(Nnobs-1),(Nnobs*(Tnobs-1)-Kreg))
    
```

Suivant les résultats du test d'homogénéité de Hsiao figurants dans l'annexe I on a :

- Concernant la première phase du test de Hsiao, on a la valeur de la statistique de test est supérieure à la valeur critique (tabulée)  $F_{Cal}^1 = 699.8694 > F_{4,57}^{0.05} = 2.5335$  et aussi la probabilité de la statistique de test  $1.03E-47$  est inférieure au seuil de 5%, donc l'hypothèse nulle  $H_0^1$  relative à l'hétérogénéité totale des coefficients est rejetée, donc on passe à la deuxième phase du test
- Concernant la seconde phase on remarque que la valeur de la statistique de test est inférieure à la valeur critique,  $F_{Cal}^2 = 2.2608 < F_{2,57}^{0.05} = 3.1588$  et aussi la probabilité de la statistique de test 0.08 est supérieure au dit seuil 0.05 donc l'hypothèse nulle  $H_0^2$  relative à homogénéité des coefficients des régressions est acceptée, donc on passe à la troisième phase du test.
- Concernant cette dernière phase, la valeur de la statistique de test est supérieure à la valeur critique  $F_{Cal}^3 = 1218.4870 > F_{4,59}^{0.05} = 3.1531$  et aussi la probabilité de la statistique de test  $1.05E-48$  est inférieure au dit seuil 5% donc l'hypothèse nulle  $H_0^3$  relative à l'homogénéité des termes constants est rejetée et l'hypothèse alternative  $H_1^3$  relative à l'hétérogénéité des termes constants est acceptée.

L'acceptation de l'hypothèse nulle  $H_0^2$  (homogénéité des coefficients des régressions) et de l'hypothèse alternative  $H_1^3$  (hétérogénéité des termes constants), nous mène à déduire la présence et l'existence des spécificités et des effets individuels de chaque pays. Pour savoir la nature de ces effets et ces caractéristiques nous faisons recours au test de Hausman.

#### 4.1.2 Le test de Hausman:

La présence et l'existence des effets individuels a été confirmée par le test d'homogénéité de Hsiao, c'est-à-dire que chaque pays des pays du Maghreb Arabe a ses propres caractéristiques et spécificités qui lui sont liées, la phase suivante des tests de spécifications consiste à utiliser le test de Hausman qui permet de savoir la nature des effets individuels, ce qui permet de choisir entre le modèle à effets individuels fixes et le modèle à effets individuels aléatoires afin de déterminer le modèle théorique final approprié. Le test de Hausman repose sur les hypothèses suivantes :

$H_0$  : le modèle à effets individuels aléatoires est le modèle approprié

$H_1$  : le modèle à effets individuels fixes est le modèle approprié

Les résultats du test d'Hausman figurants à l'annexe III, montrent que la réalisation de la statistique de test d'hausman  $H$  est de 4.504. Etant donné que le modèle comporte une variable explicative ( $K=1$ ), cette statistique suit un chi deux à un degré de liberté. A 95%, le seuil est de 3.841, en plus la p-value nulle de la statistique de test  $H$ , donc ici on rejette l'hypothèse nulle et on accepte l'hypothèse alternative que les effets individuels sont du type fixes.

Suivant les résultats des tests de Hsiao et de Hausman, on doit privilégier l'adoption d'un modèle à effets individuels fixes. Donc nous allons estimer l'impact de l'endettement sur les recettes fiscales, sur un panel de trois pays du Maghreb Arabe, observé sur la période qui s'étale de 2000 à 2020. Le modèle à effets individuels fixes théorique à estimer s'écrit comme suit:

$$Rec\_Fis_{it} = \beta_{0i} + \beta_{1i} Det\_Pub_{it} + \varepsilon_{it}$$

Les résultats d'estimation sont présentés en annexe II, alors le modèle estimé s'écrit comme suit :

$$Rec\_Fis_{it} = 18.60692 - 3.28E-10 Det\_Pub_{it}$$

## 5. Résultats et discussion.

Les résultats de l'estimation montrent que :

- Le coefficient de détermination ajusté  $\bar{R}^2 = 0.8757$ , cela signifie que 87.57% des variations et des fluctuations des recettes fiscales des pays d'étude sont expliquées par les dettes publiques. 12.43% de ces fluctuations sont expliquées par d'autres variables non prises en considération dans le modèle mais sont insérées dans le terme aléatoire  $\varepsilon_{it}$  du modèle. Donc le modèle estimé à un pouvoir de prédiction globalement acceptable.
- dû à la nullité de sa p-value, Le terme constant du modèle estimé présente une signification statistique.
- Le coefficient associé à la variable explicative relative à l'accumulation de la dette publique, est statistiquement significatif car sa p-value est de 0.0074 inférieure au seuil de signification 0.05.
- Suivant la valeur négative du coefficient de la variable explicative, cette dernière est négativement liée à la variable expliquée, c'est-à-dire la dette publique affecte négativement les recettes fiscales. Ce résultat est conforme au fondement théorique de la relation entre ces deux variables. Donc La variable explicative de la dette publique à un impact significatif sur la variable endogène des recettes fiscales.
- Une augmentation d'une unité de la dette publique engendre une diminution des recettes fiscales de 3.28E-10. toutes choses égales par ailleurs.

## 6. Conclusion

A travers cette étude, nous avons pour objectif de présenter les différentes techniques utilisées pour déterminer le modèle théorique approprié pour modéliser les données de panel. L'atteinte de cet objectif a nécessité une application économétrique. Cette dernière a été menée sur les données de panel relatives aux dettes publiques et aux recettes fiscales des trois pays du Maghreb Arabe (l'Algérie, la Tunisie et le Maroc), afin de savoir si l'accumulation de la dette publique a un impact sur la mobilisation des recettes fiscales dans ces trois pays.

Après l'utilisation des différents tests de spécification et de validation des modèles (Hsiao, Hausmann), il ressort que le modèle approprié et le mieux adapté pour ces types de données des trois pays du Maghreb Arabe est le modèle à effets individuels fixes.

Les résultats d'estimation du modèle à effets individuels fixes nous a permis d'établir conformément à la littérature existante, l'impact négatif de l'accumulation de la dette publique sur la mobilisation des recettes fiscales. Cela signifie que l'endettement affecte négativement les recettes fiscales, telles que une augmentation d'un point du taux d'endettement public entraîne et implique une baisse et une diminution très légère de 3.28E-8 % du taux de recettes fiscales, toutes choses égales par ailleurs. Ce résultat contribue à l'explication de la dépendance des pays du Maghreb Arabe (l'Algérie, le Maroc et la Tunisie) à la dette et à l'aide publique au développement.

Vu que l'endettement à un impact négatif sur les recettes fiscales des trois pays du Maghreb Arabe, il serait opportun de ramener le seuil de ce ratio à un niveau moins élevé et d'accroître celui de la pression fiscale de sorte à réduire et minimiser au minimum et à auto financer le décollage

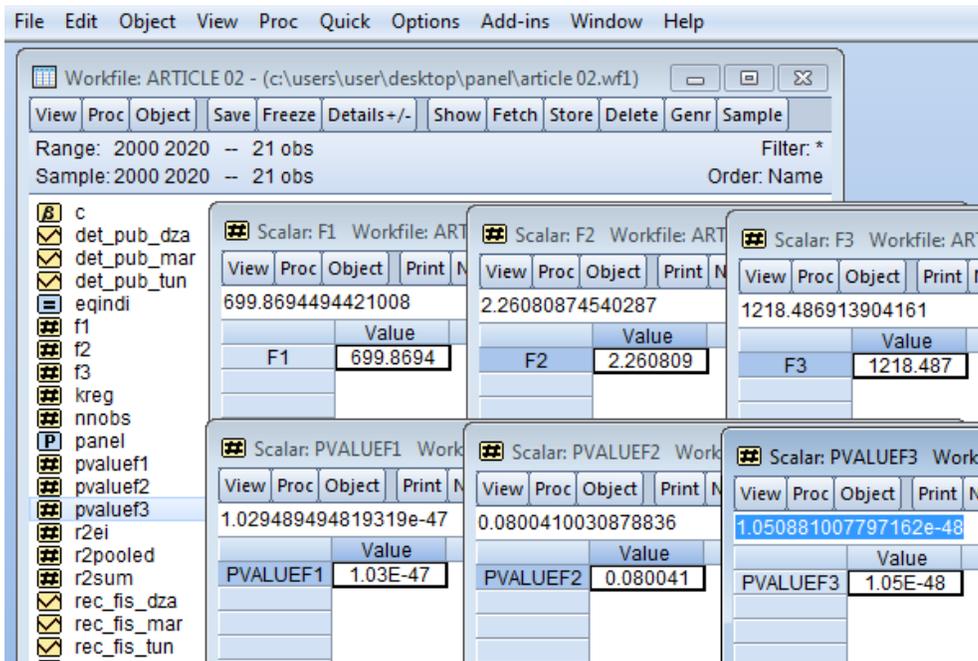
Il est sûr que ce travail n'est pas parfait, et présente des limites. C'est pour cela nous recommandons de réestimer le modèle théorique proposé en insérant d'autres variables de contrôle additionnellement à la variable d'intérêt, et élargissant la période d'étude.

## 7. Citations:

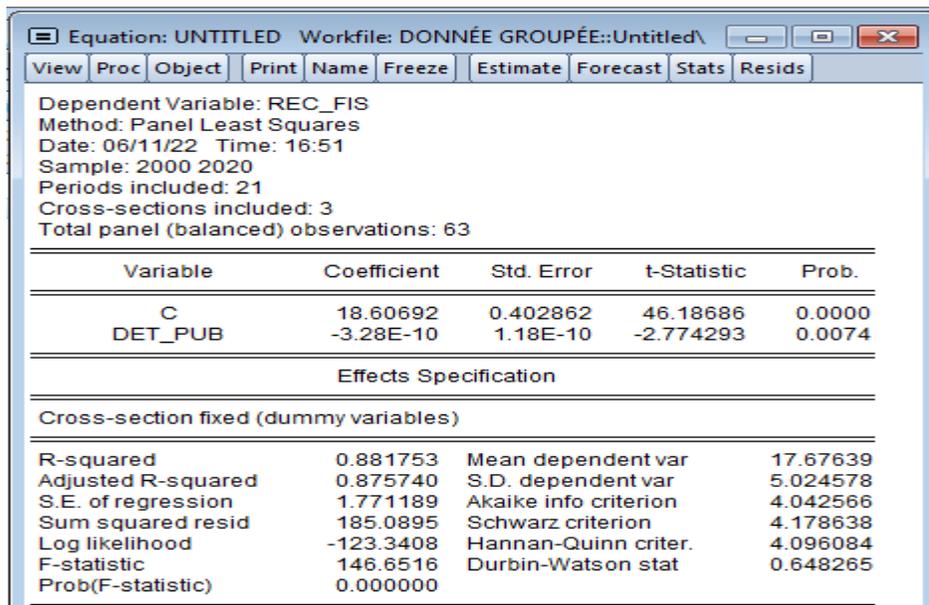
- 
- <sup>1</sup> Bouhena K. & El-Arbi M. (2020). *Institutional Quality and Economic Growth in North African Countries: An Empirical Study Using Panel-ARDL Model During The Period (1990- 2019)*, *Journa of North African*, Volume 16, Issue 02, P.8
  - <sup>2</sup> Droesbeke, J.-J., & Saporta, G. (2010). *Analyse statistique des données longitudinales*. Paris, France: Technip.
  - <sup>3</sup> Matyas, L., & Sevestre, P. (2008). *The Econometrics of Panel Data : Fundamentals and Recent Developments*. Springer.
  - <sup>4</sup> Frees, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Royaume Uni, Royaume Uni: Cambridge University Press.
  - <sup>5</sup> Green, W. H. (2006). *Econometric Analysis*. New York, USA: PEARSON.
  - <sup>6</sup> Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, England: The MIT Press Cambridge.
  - <sup>7</sup> Bourbonnais, R. (2009). *Econométrie : Cours et exercices corrigés*. Paris, France: Dunod.
  - <sup>8</sup> Madala, G. (1993). *The econometrics for panel data*. New York, U.S.A: Brookfield: Eglar.

8. Annexes :

Annexe I: Résultats du test d'homogénéité de Hsiao



Annexe II: Résultats d'estimation du modèle à effets individuels fixes



Cross-section Fixed Effects				
	CROSSID	Effect		
1	1	-6.636858		
2	2	2.330397		
3	3	4.306461		

Annexe III: Résultats du test d'hausman

Correlated Random Effects - Hausman Test  
Equation: Untitled  
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	4.504906	1	0.0000