

Dossier d'économétrie des données de panel

BACHER QUENTIN M1 APE-DS2E

I. Introduction

- 1.1. Contexte et objectif de l'étude
- 1.2. Problématique et hypothèses
- 1.3. Présentation des variables et du modèle économétrique

II. Le cadre statistique et les estimations

- 2.1. Présentation des données (source, description)
- 2.2. Présentation des résultats économétriques (tableau des estimations, comparaison des modèles)

III. Analyse et discussion des résultats

- 3.1. Choix du modèle et synthèse des résultats
- 3.2. Interprétation des résultats

IV. Conclusion et perspectives

- 4.1. Limites et améliorations possibles
- 4.2. Références bibliographiques

Annexes

Code R utilisé

I. Introduction

1.1. Contexte et objectif de l'étude

La criminalité est un enjeu majeur aux États-Unis, suscitant de nombreuses recherches en économie et en sciences sociales. La compréhension des déterminants du taux de criminalité est essentielle pour orienter les politiques publiques et améliorer la sécurité des populations. Parmi les facteurs souvent mis en avant, la **densité de population** pourrait jouer un rôle clé. D'un côté, une forte densité peut favoriser les tensions sociales et accroître les opportunités criminelles. De l'autre, elle peut aussi permettre une meilleure surveillance et dissuader les comportements criminels.

L'objectif de cette étude est d'analyser l'impact de la densité de population sur le **taux de criminalité violente** en utilisant une approche économétrique sur **données de panel**. Pour cela, nous mobiliserons des techniques économétriques adaptées afin d'estimer les effets individuels (propres à chaque État) et les tendances temporelles.

1.2. Problématique et hypothèses

La question centrale de cette recherche est la suivante :

Les États américains les plus densément peuplés ont-ils des taux de criminalité plus élevés ?

Nous testerons trois hypothèses principales :

- **H1** : Une forte densité de population entraîne une augmentation du taux de criminalité en raison de tensions sociales accrues et d'un plus grand nombre d'opportunités criminelles.
- **H2** : Une forte densité réduit la criminalité en facilitant la surveillance policière et en améliorant l'accès aux services publics.
- **H3** : La densité de population n'a pas d'effet significatif sur la criminalité lorsqu'on contrôle d'autres variables clés telles que le revenu moyen, les politiques judiciaires et le nombre de prisonniers.

1.3. Présentation des variables et du modèle économétrique

$$\text{Violent}_{it} = \beta_0 + \beta_1 \text{Density}_{it} + \beta_2 \text{Income}_{it} + \beta_3 \text{Law}_{it} + \beta_4 \text{Afam}_{it} + \beta_5 \text{Prisoners}_{it} + \alpha_i + \lambda_t + \epsilon_{it}$$

Ce modèle permet de prendre en compte les différences spécifiques entre les États et les années, en capturant les effets fixes ou aléatoires. En utilisant des effets fixes, il est possible de contrôler les caractéristiques uniques de chaque État (comme les politiques locales, le système judiciaire ou les pratiques sociétales) qui peuvent influencer la criminalité, indépendamment des variables observées. De plus, les effets temporels (comme les changements économiques, législatifs ou sociaux) peuvent être pris en compte, ce qui permet d'analyser l'évolution des tendances criminelles au fil du temps.

En tenant compte de ces différences spécifiques, le modèle réduit le risque de biais dans l'estimation des relations entre les variables explicatives (comme la densité de population, le revenu ou la législation sur les armes) et la criminalité. Cela permet d'obtenir des résultats plus précis, en identifiant plus clairement les facteurs qui influencent la criminalité tout en contrôlant les effets non observés pouvant fausser les conclusions.

De plus comme le dit Wooldridge, même si un effet n'est pas significatif, cela ne "coûte" rien de le mettre dans le modèle

Variables utilisées :

Violent : Le taux de criminalité violente par État et par année.

Density : La densité de population (personnes par km²) par État et par année.

Income : Le revenu moyen par habitant pour chaque État et chaque année.

Law : L'indicateur de sévérité des lois par État et par année.

Afam : La proportion d'Afro-Américains dans la population par État.

Prisoners : Le nombre de prisonniers par habitant dans chaque État et année.

Variable	Définition	Effet attendu	Explication	Objectif
Violent	Taux de criminalité violente	Variable dépendante	Mesure l'impact des facteurs socio-économiques sur la criminalité violente.	Mesurer l'évolution de la criminalité violente.
Density	Densité de population	Incertain	La densité peut augmenter les tensions sociales et la criminalité, mais aussi favoriser la surveillance.	Vérifier si la densité urbaine influence la criminalité.
Income	Revenu moyen par habitant	Négatif	Un revenu plus élevé réduit les incitations à commettre des crimes, sauf si les inégalités sont élevées.	Tester si un revenu plus élevé réduit le crime.
Law	Indicateur de sévérité des lois	Négatif	Des lois strictes peuvent dissuader la criminalité, mais l'efficacité dépend de l'application et de la justice perçue.	Évaluer l'effet dissuasif des lois sur la criminalité.
Afam	Part d'Afro-Américains dans la population	Incertain	Influence des dynamiques socio-économiques et historiques, l'effet peut varier selon les contextes.	Contrôler les effets socio-économiques et historiques.
Prisoners	Nombre de prisonniers par habitant	Négatif	Un nombre élevé de prisonniers peut dissuader, mais trop d'incarcération peut signifier une politique inefficace.	Analyser l'effet dissuasif de l'incarcération.

II. Le cadre statistique et les estimations

2.1. Présentation des données

- **Source des données**

Les données utilisées proviennent de la base de données "**Guns**", disponible dans le package **AER** sous R. Cette base de données rassemble des informations sur la criminalité et divers facteurs socio-économiques aux États-Unis, couvrant la période de 1977 à 1999. Elle comprend un total de **1 173 observations** réparties sur **51 entités** (les 50 États américains ainsi que le District de Columbia) sur une période de **23 ans**. La base de données est utilisée pour explorer l'impact des variables économiques, démographiques et législatives sur la criminalité aux États-Unis.

Les études liées à cette base de données comprennent des travaux de recherche tels que Ayres et Donohue (2003) qui ont analysé la relation entre la législation sur les armes et la criminalité, ainsi que Stock et Watson (2007) dans leur ouvrage *Introduction to Econometrics*, qui servent de base pour l'analyse économétrique de cette base.

2.2. Présentation des résultats économétriques (tableau des estimations, comparaison des modèles)

Variable	OLS (MCO)	Between (Effets fixes)	Between Temporel (Années)	Within (Effets fixes)	Within Temporel (Effets fixes)	Within Ind + Temp (Effets fixes)	FGLS (Effet aléatoire)
Constante	450.30*** (50.12)	380.21 (112.45)	410.78 (98.67)	Non applicable	Non applicable	Non applicable	425.56*** (32.98)
Density	76.90*** (5.58)	-80.10 (51.66)	-76.18 (46.26)	-80.10 (51.66)	-76.18 (46.26)	91.10*** (17.37)	91.10*** (17.37)
Income	0.0086*** (0.0026)	0.0061* (0.0033)	0.0108** (0.0038)	0.0061* (0.0033)	0.0108** (0.0038)	0.0108** (0.0038)	0.0014 (0.0029)
Law	-131.70*** (14.04)	0.46 (10.88)	10.23 (10.14)	0.46 (10.88)	10.23 (10.14)	-18.61* (10.63)	-18.61* (10.63)
Afam	8.00*** (1.53)	7.60 (9.31)	-5.16 (8.40)	7.60 (9.31)	-5.16 (8.40)	8.49* (4.59)	8.49* (4.59)
Prisoners	0.83*** (0.045)	0.27*** (0.051)	0.22*** (0.052)	0.27*** (0.051)	0.22*** (0.052)	0.43*** (0.038)	0.43*** (0.038)
Effets	Non applicable	Individuels	Temporels	Individuels	Temporels	Individuels + Temp	Non applicable
R ² ajusté	0.6455	0.1311	0.3626	0.7752	0.7588	0.0886	0.2209

*** : p-value < 0.001 (très significatif)

** : p-value < 0.01 (significatif à 1%)

* : p-value < 0.05 (significatif à 5%)

III. Analyse et discussion des résultats

3.1. Choix du modèle et synthèse des résultats

Test de Hausman (p-value = $2.787e-11$) : La p-value est extrêmement faible (inférieure à 0.05), ce qui signifie que l'hypothèse nulle est rejetée. **Donc un modèle à effet fixe est préféré à celui à effet aléatoire**

Test de Breusch-Pagan pour l'homoscédasticité (p-value < $2.2e-16$) : La p-value étant très faible, l'hypothèse nulle d'homoscédasticité est rejetée. Cela suggère qu'il y a une **hétéroscédasticité** présente

Test de Wooldridge pour l'autocorrélation (p-value < $2.2e-16$) : La p-value très faible signifie que l'hypothèse nulle d'absence d'autocorrélation est rejetée. Le modèle souffre probablement **d'autocorrélation**

Je choisis le modèle à double effet car les variables sont légèrement plus significatives que les autres modèles within avec notamment une variable *density* significative, même si le R^2 renforcé est faible. Même si globalement tous les modèles sont plutôt mauvais pour savoir si les états les plus peuplés sont plus susceptibles à la criminalité.

3.2. Interprétation des résultats

Avec ce modèle, nous pouvons voir qu'effectivement la densité de population augmente le taux de criminalité. Nous pouvons aussi voir que conformément à ce que l'on pouvait s'attendre avec l'indicateur *law*, la criminalité baisse avec des lois plus strictes. Néanmoins *income* n'est pas conforme à ce que l'on attendait puisque l'on voit que lorsque le revenu moyen par habitant augmente, la violence aussi. De plus, ce modèle nous dit aussi que plus le nombre de prisonniers augmente, plus le taux de violence dans les états augmente, ce qui peut paraître surprenant. De plus, certaines variables ne sont pas significatives, comme la variable *Afam* et n'ont donc pas d'impact sur l'évolution de la criminalité en fonction de la densité de la population.

IV. Conclusion et perspectives

4.1. Limites et améliorations possibles

Le modèle ne permet pas de bien savoir si les états les plus peuplés sont plus susceptibles de la criminalité et comporte plusieurs problèmes mettant à mal son bon fonctionnement. Tout d'abord certaines variables explicatives ne sont que peu significatives (Ex: *afam* non significative à 10%), et le R^2 renforcé du modèle à double effet est faible. De plus, le test de Wooldridge et Brush Pagan montre des signes de corrélation et d'hétéroélasticité.

Il faudrait donc remplacer certaines variables inutiles et ajouter de nouvelles variables variables qui permettraient d'avoir peut être de meilleurs résultats (Ex : nombre de force de l'ordre dans chaque état) afin d'augmenter la précision du modèle et d'expliquer mieux la variable dépendante. Il serait aussi pertinent de renforcer le modèle avec la méthode de White.

4.2. Références bibliographiques

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*.

Ayres, I., and Donohue, J.J. (2003). Shooting Down the 'More Guns Less Crime' Hypothesis. *Stanford Law Review*, 55, 1193–1312

Stock, J.H. and Watson, M.W. (2007). *Introduction to Econometrics*, 2nd ed. Boston: Addison Wesley.

Annexes

```
# Modèle OLS (MCO)
ols_model <- lm(violent ~ density + income + law + afam + prisoners, data = Guns)
summary(ols_model)

# Modèle à effets fixes individuels (within)
fe_ind_model <- plm(violent ~ density + income + law + afam + prisoners,
                    data = Guns, index = c("state", "year"), model = "within")

# Modèle à effets fixes temporels
fe_temp_model <- plm(violent ~ density + income + law + afam + prisoners + factor(year),
                    data = Guns, index = c("state", "year"), model = "within")

# Modèle à effets fixes individuels et temporels
fe_both_model <- plm(violent ~ density + income + law + afam + prisoners + factor(year),
                    data = Guns, index = c("state", "year"), model = "within", effect = "twoways")

# Modèle Between (comparaison entre États ou entre années)
be_ind_model <- plm(violent ~ density + income + law + afam + prisoners,
                    data = Guns, index = c("state", "year"), model = "between", effect = "individual")

be_temp_model <- plm(violent ~ density + income + law + afam + prisoners,
                    data = Guns, index = c("state", "year"), model = "between", effect = "time")

# Modèle à effets aléatoires (RE)
re_model <- plm(violent ~ density + income + law + afam + prisoners,
                data = Guns, index = c("state", "year"), model = "random")
```

```
# Test de Hausman (choix entre effets fixes et effets aléatoires)
hausman_test <- phptest(fe_both_model, re_model)
print(hausman_test)

# Test de Breusch-Pagan pour l'homoscédasticité
bptest(ols_model)

# Test de Wooldridge pour l'autocorrélation
pbgtest(fe_both_model)
```