

FACULTE DE SCIENCES ECONOMIQUES

DÉPARTEMENT MIASHS

Projet d'Econométrie et modélisation économique

Formation : Master mention Mathématiques appliquées, statistique

Parcours : Science des données, statistique et économétrique

*Facteurs explicatifs des disparités de revenus entre les
départements de France*

Rédigé par :

Conrad S. AGOSSOU

&

Uriel C. ADOTANOU

Responsable de formation :

Isabelle CADORET,

Professeur,

Membre permanent du CREM

Enseignant :

Véronique Thelen,

Professeur Assistant Faculté des Sciences

Economiques

40ème promotion

Sommaire

Résumé	iv
Abstract	v
Introduction	1
1 Cadre théorique et méthodologique de l'étude	2
2 Résultats, analyses et interprétations	8
Conclusion	19
Références	20

Liste des tableaux

1.1	Description des variables utilisées dans l'analyse	7
2.1	Comparaison des R^2 ajustés des trois modèles	14
2.2	Comparaison des estimations MCO, MCO White, et MCO Newey pour le modèle log-log	15
2.3	Vérification des hypothèses	18

Table des figures

1.1	Organigramme de la modélisation	6
2.1	Analyse descriptive univariée : Distribution des variables par boxplots	8
2.2	Relations initiales entre la médiane du niveau de vie et les variables explicatives (Unités légales, Prestations sociales, Densité de population)	9
2.3	Analyse bivariée : Comparaison des variables selon le type de zone (rural vs urbain)	10
2.4	Relation entre le logarithme du nombre d'unités légales et la médiane du niveau de vie	11
2.5	Relation entre le logarithme des prestations sociales et la médiane du niveau de vie .	11
2.6	Relation entre le logarithme de la densité de population et la médiane du niveau de vie	11
2.7	Matrice de corrélation des variables explicatives	12

Résumé

Cette étude a pour but de déterminer les facteurs explicatifs des disparités de revenus entre les départements de France. Elle s'appuie sur des données socio-économiques issues de sources publiques pour analyser les écarts de revenus mesurés par la médiane du niveau de vie. À cet effet, une modélisation économétrique (régression multiple) a été réalisée à l'aide des logiciels R et Excel. Les analyses ont été effectuées après une description univariée et bivariée des données, suivie de l'estimation des modèles.

De l'analyse descriptive, il ressort que les disparités de revenus sont marquées par des différences importantes entre départements. Les résultats obtenus des analyses économétriques indiquent que la proportion de ménages imposés, la dépendance aux prestations sociales, le taux de pauvreté, et la structure socio-professionnelle (notamment la proportion de cadres et d'ouvriers) sont les principaux facteurs explicatifs des écarts de revenus. En particulier, il apparaît que les départements avec une forte proportion de ménages imposés présentent des revenus médians plus élevés, tandis que ceux ayant une forte dépendance aux prestations sociales ou des taux de pauvreté élevés affichent des revenus médians significativement plus faibles.

Les résultats des modèles économétriques suggèrent également que le taux de chômage, bien qu'associé négativement à l'activité économique, peut paradoxalement être corrélé à des revenus médians plus élevés dans certains départements, probablement en raison de l'effet des dispositifs de compensation sociale.

Au regard de ces résultats, nous suggérons aux décideurs publics de cibler les politiques de réduction des disparités sur les départements les plus vulnérables en encourageant l'activité économique locale, en réduisant la pauvreté, et en limitant la dépendance aux prestations sociales.

Abstract

This study aims to identify the key factors explaining income disparities across French departments. It relies on socio-economic data from public sources to analyze income differences measured by the median standard of living. To this end, an econometric modeling approach (multiple regression) was conducted using R and Excel software. The analyses were carried out following a univariate and bivariate description of the data, followed by model estimation.

From the descriptive analysis, it was found that income disparities are significantly marked by major differences between departments. Econometric results reveal that the proportion of taxable households, dependency on social benefits, poverty rate, and socio-professional structure (particularly the proportion of executives and workers) are the main factors explaining income disparities. Specifically, departments with a higher proportion of taxable households tend to have higher median incomes, while those with high dependency on social benefits or elevated poverty rates exhibit significantly lower median incomes.

Econometric model results also suggest that the unemployment rate, although generally associated with lower economic activity, paradoxically correlates with higher median incomes in some departments, likely due to the effect of social compensation mechanisms.

Based on these results, we recommend that public policymakers focus on reducing disparities by targeting the most vulnerable departments, promoting local economic activity, reducing poverty, and limiting dependency on social benefits.

Introduction

Les disparités de revenus représentent un enjeu majeur dans les sociétés contemporaines, révélant des écarts significatifs dans la répartition des richesses entre les individus ou les groupes sociaux. Ces écarts ne se limitent pas à un simple indicateur économique : ils témoignent également des inégalités d'accès aux ressources, aux opportunités et aux droits fondamentaux. Si, dans une certaine mesure, les différences de revenus peuvent être perçues comme un levier de motivation et de récompense pour les efforts et les compétences, leur amplification excessive entraîne des déséquilibres sociaux et économiques préoccupants.

Plusieurs facteurs contribuent à ces inégalités. L'éducation, par exemple, joue un rôle déterminant : l'accès à une formation de qualité conditionne souvent les opportunités professionnelles et les perspectives salariales. De même, le genre et l'origine ethnique continuent de peser lourdement sur la rémunération, malgré les avancées en matière d'égalité des droits. À cela s'ajoutent des dynamiques géographiques, avec des disparités marquées entre les zones urbaines et rurales, ainsi qu'entre les pays développés et les économies émergentes. Enfin, les transformations économiques, telles que la mondialisation et la digitalisation, exacerbent parfois ces écarts en creusant le fossé entre les travailleurs hautement qualifiés et les moins qualifiés.

Les conséquences de ces disparités de revenus vont au-delà de l'individu. Elles affectent la cohésion sociale en alimentant les tensions entre classes sociales, réduisent la mobilité économique et limitent la capacité de nombreuses personnes à accéder à une vie digne. Sur le plan macroéconomique, elles freinent la croissance en concentrant les richesses dans les mains de quelques-uns, ce qui restreint la consommation et l'investissement de la majorité.

Face à ces défis, les solutions envisagées sont multiples et nécessitent une approche holistique. Parmi elles figurent la mise en place de politiques fiscales redistributives, l'amélioration de l'accès à l'éducation et à la formation, la promotion de l'égalité des chances dans le monde du travail, et le développement de systèmes de protection sociale inclusifs. Ces mesures, bien qu'ambitieuses, sont essentielles pour réduire les écarts de revenus et favoriser des sociétés plus justes et plus résilientes.

Ainsi, les disparités de revenus ne sont pas une fatalité : elles constituent un enjeu complexe mais surmontable, à condition de mobiliser des efforts concertés à l'échelle locale, nationale et internationale.

En vue de mesurer au premier abord des différents facteurs socio-économiques, participer à l'accumulation des connaissances au sujet des disparités et mesurer l'impact de ces effets, nous avons choisi d'aborder le thème : « Expliquer les facteurs qui influencent les disparités de revenus entre les départements ». Le développement de ce thème s'articule autour de trois chapitres dont le premier propose une présentation de notre sujet. Le deuxième chapitre décrit le cadre théorique et la méthodologie de l'étude alors que le troisième est relatif à la présentation des résultats ainsi qu'aux implications de politiques.

1 | Cadre théorique et méthodologique de l'étude

Ce chapitre expose une réflexion théorique sur les facteurs explicatifs des disparités de revenus. Il définit la démarche méthodologique adoptée et montre essentiellement l'intérêt scientifique du sujet, en posant le problème, les objectifs et hypothèses de recherche appuyés par les études déjà réalisées dans ce sens.

1.1 Problématique et intérêt de l'étude

Depuis plusieurs décennies, la répartition des revenus suscite l'intérêt des économistes. Depuis les années 1990, les inégalités de revenus et de richesses ont fait l'objet d'un regain d'attention, à la fois dans les travaux théoriques et les études empiriques (Tura et Assefa, 2017).

Le choix des variables dans cette analyse s'appuie sur des recherches antérieures, bien que limitées, qui explorent les inégalités socio-économiques. Ces inégalités, caractérisées par des écarts significatifs entre les revenus des individus ou des ménages, demeurent un enjeu crucial pour nos sociétés. Malgré les progrès liés à la croissance économique et à l'amélioration générale des conditions de vie, cette prospérité n'a pas été répartie équitablement. En conséquence, de nombreux pays, qu'ils soient développés ou en développement, doivent faire face à des inégalités croissantes menaçant la cohésion sociale et la stabilité économique et politique.

Les disparités de revenus trouvent leurs origines dans divers facteurs : l'accès inégal à l'éducation, les écarts de rémunération entre secteurs, les effets de la mondialisation et les mutations technologiques. Les travailleurs qualifiés dans des secteurs porteurs, comme la technologie ou la finance, jouissent souvent de revenus bien supérieurs à ceux des moins qualifiés, actifs dans des secteurs en déclin. Ces inégalités sont exacerbées par des facteurs sociaux tels que les discriminations liées au genre, à l'origine ethnique ou au lieu de résidence.

Les conséquences de ces écarts sont profondes. Économiquement, elles peuvent freiner la croissance en réduisant la consommation des ménages à faibles revenus, tandis que les ménages aisés accumulent une épargne importante, accentuant les déséquilibres économiques. Socialement, ces inégalités alimentent des tensions, nuisent à la mobilité sociale, et perpétuent des cycles intergénérationnels de pauvreté.

Face à ces constats, une question essentielle se pose : quels facteurs socio-économiques et démographiques influencent les disparités de revenus, et quelles politiques publiques pourraient les réduire tout en soutenant une croissance durable et inclusive ?

Cette problématique appelle à examiner les causes profondes des inégalités, à évaluer leur impact sur différents aspects de la vie sociale et économique, et à envisager des solutions concrètes. Il s'agit notamment d'identifier des actions gouvernementales, des réformes éducatives, ou des initiatives privées capables de favoriser une répartition plus équitable des richesses. Le défi réside dans

la recherche d'un équilibre entre efficacité économique et justice sociale pour bâtir un avenir plus équitable.

Plusieurs travaux ont exploré les multiples dimensions de ces inégalités. Par exemple, le Conseil d'Analyse Économique dans son rapport de 2001, met en lumière des facteurs explicatifs tels que le chômage, le taux d'emploi, l'âge, la composition des ménages, le patrimoine hérité et les politiques fiscales. Une étude au Sénégal (Ndeye Fatou Faye, Moussa Sall, François Affholder et Françoise Gérard) révèle également l'impact du milieu de vie (rural ou urbain) et de la densité de la population sur les écarts de revenus.

Sur une perspective historique, Morrisson Christian et Wayne Snyder analysent les inégalités en France entre le XVIII^e siècle et 1985, en intégrant la notion de classes sociales, allant des ouvriers à la bourgeoisie, et en étudiant leur évolution en nombre et en revenus. Dans un autre registre, Marc Guérand , dans « Économie de l'éducation » insiste sur le rôle central de l'éducation dans la détermination des salaires. Il affirme que les inégalités éducatives influencent fortement la répartition des revenus et que l'éducation constitue à la fois un facteur de reproduction sociale et un levier pour réduire ces écarts.

Enfin, dans « Inégalité des revenus et protection sociale » (Förster et Thévenot), les prestations sociales (pensions, aides, etc.) sont mises en avant comme des instruments essentiels pour réduire les disparités et garantir une meilleure cohésion sociale. Ainsi, une compréhension approfondie des inégalités de revenus et des leviers d'action est indispensable pour concevoir des politiques publiques capables de concilier justice sociale et dynamisme économique. La présente recherche vise à combler ces vides. Elle apporte des éléments de réponses aux questions suivantes : Quels sont les effets de la profession sur le niveau de vie des populations en France ? Quels sont les effets des facteurs socio-démographiques sur le ? C'est dans le souci de répondre à ces questions que nous avons choisi mener notre étude sur le thème : « Expliquer les facteurs qui influencent les disparités de revenus entre les départements ».

1.2 Objectifs et hypothèses de l'étude

1.2.1 Objectifs de l'étude

L'objectif général de cette recherche est « Expliquer les facteurs qui influencent les disparités de revenus entre les départements en France ». De façon spécifique, il s'agira de :

- ✓ Identifier les principaux facteurs socio-économiques qui influencent les disparités de la médiane du niveau de vie entre les départements français ;
- ✓ Analyser l'impact des variables liées aux dynamiques démographiques (densité de population, type de zone) et aux politiques sociales (prestations sociales, taux de pauvreté) sur le revenu médian des départements.

1.2.2 Hypothèses

Pour atteindre ces objectifs, deux hypothèses sont formulées :

- ✓ Les départements avec une forte concentration d'activités économiques (mesurée par le nombre d'unités légales) et une proportion élevée de cadres ont un revenu médian significativement

plus élevé ;

- ✓ Les départements fortement dépendants des prestations sociales et ayant un taux de pauvreté élevé présentent un revenu médian significativement plus faible.

1.3 Méthodologie

Sous cette rubrique, nous allons présenter l'approche méthodologique ayant permis d'analyser les données collectées dans le cadre de cette étude. Il sera question de décrire notre méthodologie d'analyse, de spécifier nos modèles et enfin d'exposer notre procédure d'estimation et d'analyse.

1.3.1 Nature et sources des données

Les données utilisées dans le cadre de notre recherche sont issues de la base de données de l'Institut National de la Statistique et des Etudes Economiques. L'étude est réalisée en s'appuyant sur un échantillon de 98 départements.

1.3.2 Méthodes et outils d'analyse

1.3.2.1 Analyse descriptive

Les méthodes descriptives regroupent des techniques d'analyse univariée et bivariée. L'analyse univariée permet de dresser un portrait de la population étudiée à partir de sa répartition suivant la variable d'intérêt de l'étude et les principales variables explicatives retenues. L'analyse bivariée a pour but d'identifier les associations entre deux variables et d'apprécier la significativité statistique de cette association à partir d'un seuil donné. Dans le cadre de cette étude, elle permettra de voir les variations du phénomène selon ces dernières.

1.3.2.2 Analyse économétrique

Modèle de régression linéaire

En tenant compte de notre analyse, nous avons choisi d'utiliser le modèle de régression linéaire multiple. Il s'agit d'un modèle permettant de ressortir les liens entre une variable qu'on doit expliquer (dépendante) et plusieurs variables explicatives (indépendantes).

Afin d'estimer le modèle linéaire donné par l'équation ci-dessus, nous utilisons la méthode de Moindres Carrés Ordinaires (MCO). Théoriquement, les estimateurs des MCO sont BLUE, c'est-à-dire qu'ils sont sans biais, mais aussi convergents et efficaces. Nous utilisons le logiciel R-studio pour effectuer nos estimations et pour notre étude.

Détection des valeurs anormales

La détection de valeurs anormales dans les données est cruciale car elles peuvent avoir une influence négative dans les estimations et, à fortiori, dans les prévisions (effet levier de la fonction de régression). Les méthodes les plus utilisées sont la méthode des résidus normalisés de Pearson et le critère des distances de Cook. Dans le cas de notre étude, nous allons utiliser le critère des distances de Cook et le graphique des résidus standardisés selon leur levier.

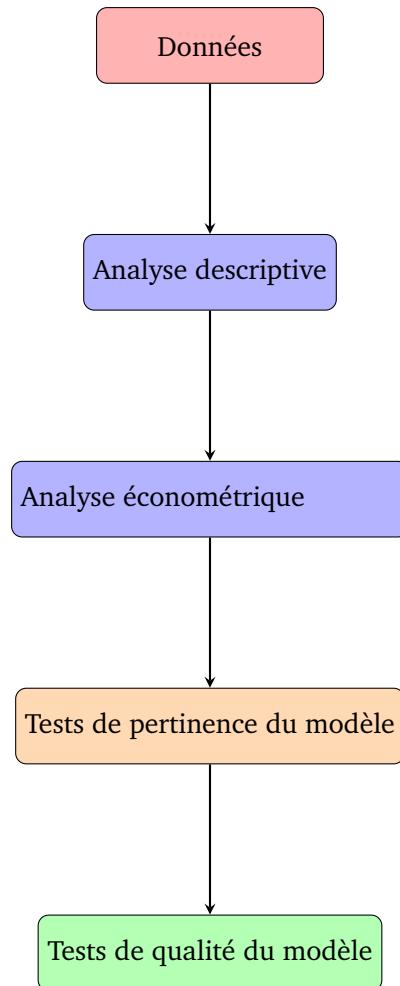
Qualité du modèle :

1. **Statistique de Fisher** : Permet de tester la significativité globale du modèle. Une p -valeur associée inférieure au seuil de 5% ($\alpha = 0.05$) indique que les variables explicatives expliquent significativement les variations de la variable dépendante.
2. **R^2 ajusté** : Mesure la proportion de variance de la variable dépendante expliquée par le modèle, tout en ajustant pour le nombre de variables explicatives. Une valeur proche de 1 indique une bonne capacité explicative.
3. **Hétéroscédasticité** : Teste si la variance des erreurs résiduelles est constante. Le test de White est utilisé, avec une p -valeur inférieure à 5% ($\alpha = 0.05$) indiquant une hétéroscédasticité significative. En cas de présence, des erreurs standards robustes (comme celles de White) seront appliquées.
4. **Autocorrélation** : Vérifie si les erreurs résiduelles sont indépendantes à l'aide du test de Durbin-Watson. Une p -valeur inférieure au seuil de 5% ($\alpha = 0.05$) indique une autocorrélation significative. Si détectée, des corrections comme celles de Newey-West doivent être appliquées pour garantir la fiabilité des résultats.

Sélection des variables du modèle

Dans le cadre de la modélisation, il est important que les variables choisies expliquent au mieux notre variable d'intérêt. Le choix des variables explicatives s'est fait à partir de la matrice de corrélation.

FIGURE 1.1 – Organigramme de la modélisation



1.4 Présentation des variables

TABLEAU 1.1 – Description des variables utilisées dans l’analyse

Libellé	Définition	Unité	Source	Traitemet
Code	Code départemental unique	-	INSEE	Aucun
Departement	Nom du département correspondant au code	-	INSEE	Aucun
Unites_legales	Nombre total d’unités légales (entreprises et structures enregistrées officiellement) dans le département	Nombre	INSEE	Aucun
Taux_activite	Proportion de la population active (en pourcentage) dans une tranche d’âge donnée	Pourcentage	INSEE	Aucun
Non_scolarisés_15_plus	Nombre de personnes âgées de 15 ans ou plus qui ne sont plus inscrites dans un établissement scolaire	Nombre	INSEE	Aucun
Menages_imposes	Part des ménages fiscaux soumis à l’impôt (en pourcentage)	Pourcentage	INSEE	Aucun
Mediane_niveau_vie	Médiane du niveau de vie annuel en euros	Euro	INSEE	Aucun
Prestations_sociales	Part des prestations sociales dans le revenu disponible (en pourcentage), mesurant la dépendance des revenus aux aides sociales	Pourcentage	INSEE	Aucun
Pensions_retraites	Part des pensions, retraites et rentes dans le revenu disponible (en pourcentage)	Pourcentage	INSEE	Aucun
Taux_pauvreté	Proportion des personnes dont le niveau de vie est inférieur au seuil de pauvreté (en pourcentage)	Pourcentage	INSEE	Aucun
Densité_population	Nombre d’habitants par kilomètre carré, reflétant la concentration de population dans le département	Hab./km ²	INSEE	Aucun
Cadres	Part des cadres et des professions intellectuelles supérieures parmi les emplois locaux (en pourcentage)	Pourcentage	INSEE	Aucun
Intermediaires	Part des professions intermédiaires parmi les emplois locaux (en pourcentage)	Pourcentage	INSEE	Aucun
Ouvriers	Part des ouvriers parmi les emplois locaux (en pourcentage)	Pourcentage	INSEE	Aucun
Taux_chomage	Taux annuel moyen de chômage (en pourcentage)	Pourcentage	INSEE	Aucun
Type_zone	Type du département (Rural ou Urbain), créé à partir de la densité de population	-	Conrad & Urié	Créé à partir de la densité de population

2 | Résultats, analyses et interprétations

Ce chapitre expose les résultats obtenus après le traitement, l'analyse des données et leurs interprétations. Il présente également les préconisations opérationnelles qui découlent de la présente étude.

2.1 Analyse descriptive

2.1.1 Analyse univariée

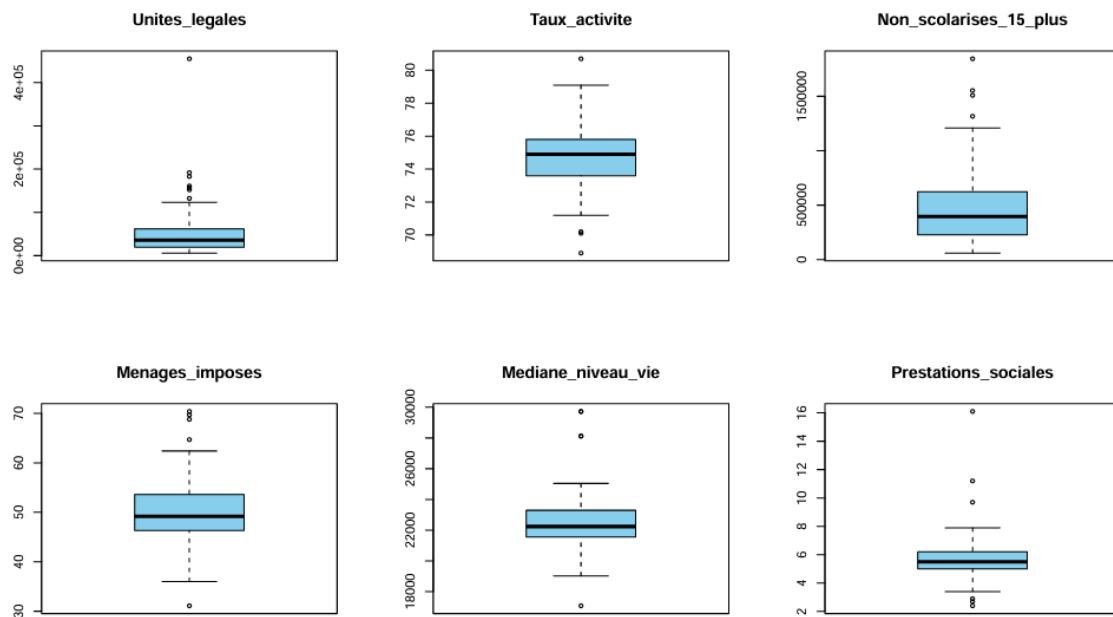


FIGURE 2.1 – Analyse descriptive univariée : Distribution des variables par boxplots

Interprétation des boxplots

— Unités légales

Ce boxplot montre une distribution fortement asymétrique. La majorité des départements ont un faible nombre d'unités légales, mais quelques départements affichent des valeurs extrêmement élevées (outliers). Les départements avec un nombre élevé d'unités légales pourraient correspondre à des pôles économiques majeurs (grandes agglomérations).

— Taux d'activité

Le taux d'activité est concentré autour de 72-78 %, avec peu de variabilité et quelques valeurs

atypiques en bas. Cela reflète une homogénéité générale, mais certains départements à faible taux d'activité pourraient être ruraux ou défavorisés économiquement.

— **Non-scolarisés de plus de 15 ans**

Distribution asymétrique, avec plusieurs départements ayant un nombre très élevé de non-scolarisés, en tant qu'outliers. Les départements avec de hauts niveaux de non-scolarisation pourraient refléter des problématiques sociales ou une population majoritairement âgée.

— **Ménages imposés**

Une majorité de départements se situe autour de 45-55 % de ménages imposés, avec des valeurs aberrantes vers le haut. Les outliers pourraient correspondre à des départements avec des revenus élevés ou une forte proportion de ménages fiscaux actifs.

— **Médiane du niveau de vie**

La médiane du niveau de vie est globalement homogène (autour de 22 000-26 000 €), avec quelques outliers élevés. Les départements avec des valeurs élevées pourraient être les plus riches ou concentrer des populations avec un fort pouvoir d'achat.

— **Prestations sociales**

Distribution avec quelques outliers importants dans les niveaux élevés. Les départements avec des prestations sociales élevées pourraient inclure ceux avec un fort taux de chômage ou un besoin accru d'aide sociale.

2.1.2 Analyse bivariée

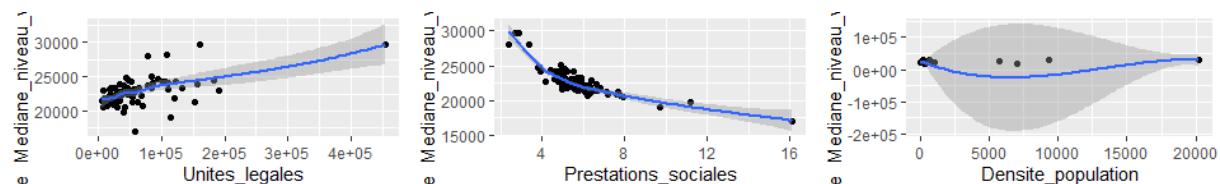


FIGURE 2.2 – Relations initiales entre la médiane du niveau de vie et les variables explicatives (Unités légales, Prestations sociales, Densité de population)

Transformation des variables

Dans l'analyse initiale (Figure 2.2), les nuages de points représentant les relations entre le revenu médian et les variables exogènes Unités légales, densité population, prestations sociales présentent des individus atypiques et la répartition des individus (trop concentrés) ne permet pas de distinguer clairement la nature de la relation entre ces exogènes et le revenu médian. Il reviendra donc pour améliorer la tendance et distinguer la nature de ces relations et obtenir une relation plus linéaire, d'appliquer l'opérateur « logarithme » (log).

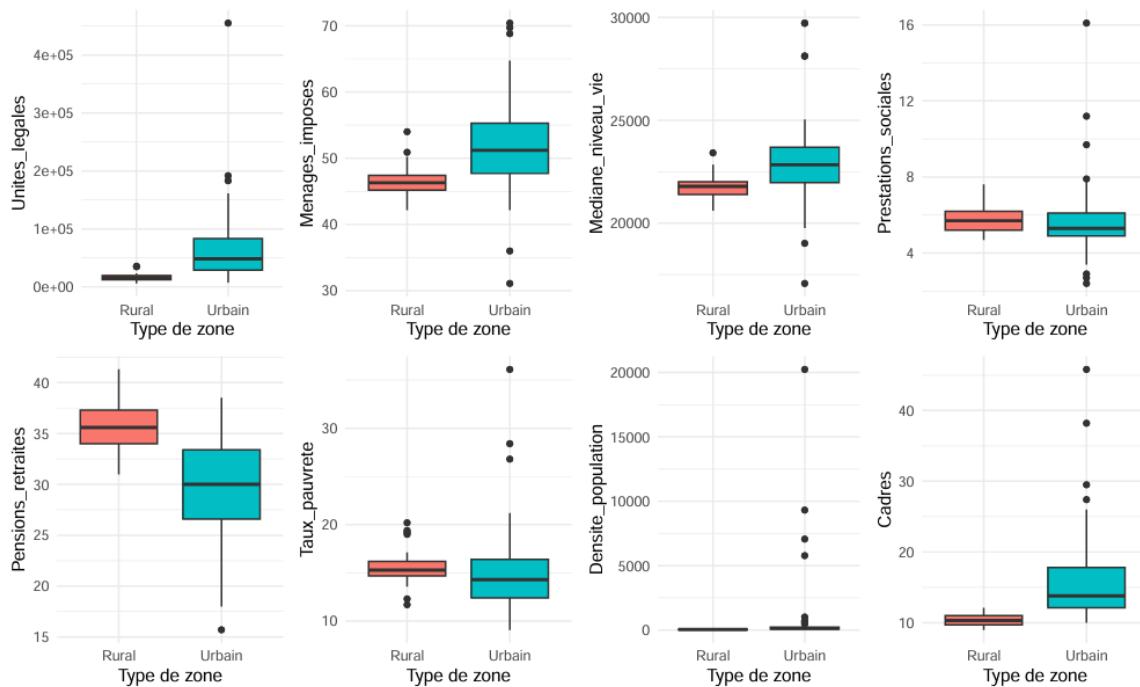


FIGURE 2.3 – Analyse bivariée : Comparaison des variables selon le type de zone (rural vs urbain)

Interprétation des boxplots

— Unités légales

Les zones urbaines présentent un nombre nettement plus élevé d'unités légales par rapport aux zones rurales. Cela reflète une concentration des activités économiques dans les zones urbaines.

— Ménages imposés

Le pourcentage de ménages imposés est généralement plus élevé dans les zones urbaines, mais avec une plus grande variabilité. Cela indique une meilleure capacité contributive dans ces zones.

— Médiane du niveau de vie

Les zones urbaines affichent une médiane du niveau de vie plus élevée que les zones rurales. Cela pourrait s'expliquer par des opportunités économiques plus importantes en milieu urbain.

— Prestations sociales

Les prestations sociales semblent légèrement plus élevées dans les zones urbaines, avec une variabilité plus marquée. Cela pourrait être lié à une plus forte densité de population en besoin d'aide sociale.

— Pensions de retraite

Les zones rurales montrent une proportion plus élevée de pensions de retraite. Cela reflète probablement une population vieillissante dans ces zones.

— Taux de pauvreté

Le taux de pauvreté est globalement plus bas dans les zones urbaines, mais il y a une plus grande dispersion dans les valeurs rurales.

— Densité de population

La densité de population est nettement plus élevée en milieu urbain, ce qui est attendu. Les zones rurales présentent une très faible densité, avec peu de variabilité.

— Cadres

La proportion de cadres est beaucoup plus importante en milieu urbain qu'en milieu rural, reflétant la localisation des emplois à forte qualification dans les villes.

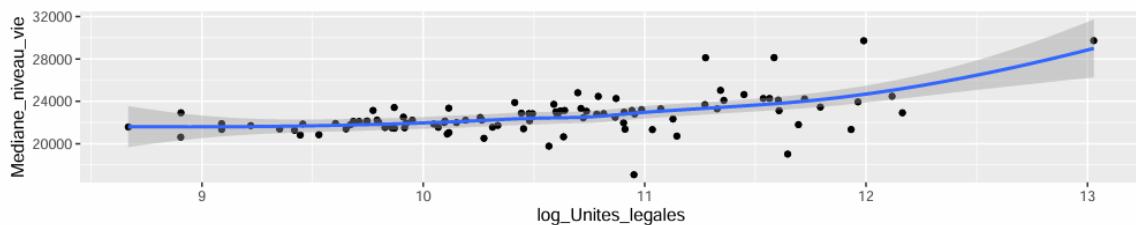


FIGURE 2.4 – Relation entre le logarithme du nombre d'unités légales et la médiane du niveau de vie

Interprétation :

La médiane du niveau de vie tend à augmenter avec le nombre d'unités légales (logarithme transformé). Cela suggère que les départements avec une activité économique plus dense (mesurée par les unités légales) ont généralement un niveau de vie plus élevé.

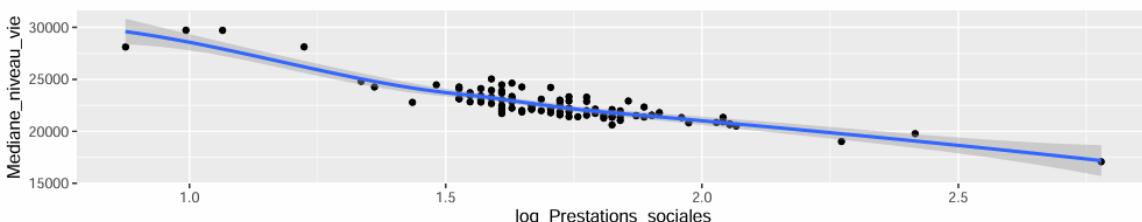


FIGURE 2.5 – Relation entre le logarithme des prestations sociales et la médiane du niveau de vie

Interprétation :

Une relation négative est observée entre le logarithme des prestations sociales et la médiane du niveau de vie. Cela indique que les départements où les prestations sociales sont plus élevées ont tendance à avoir un niveau de vie médian plus bas, probablement en raison d'une plus forte concentration de ménages en situation précaire.

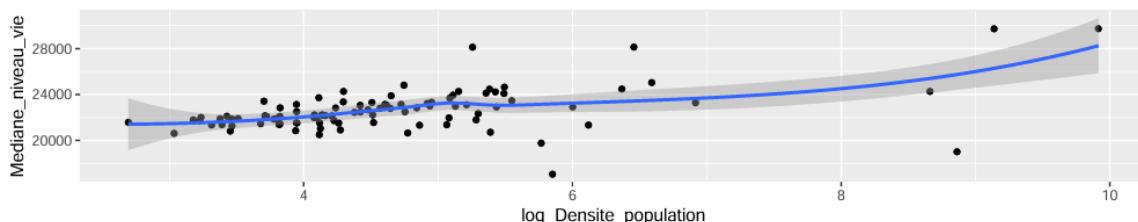


FIGURE 2.6 – Relation entre le logarithme de la densité de population et la médiane du niveau de vie

Interprétation :

La médiane du niveau de vie augmente avec la densité de population (logarithme transformé). Cela reflète une dynamique où les zones à forte densité, souvent urbaines, offrent des opportunités économiques et des salaires plus élevés, contribuant à un niveau de vie plus élevé.

2.2 Test de pré-estimation

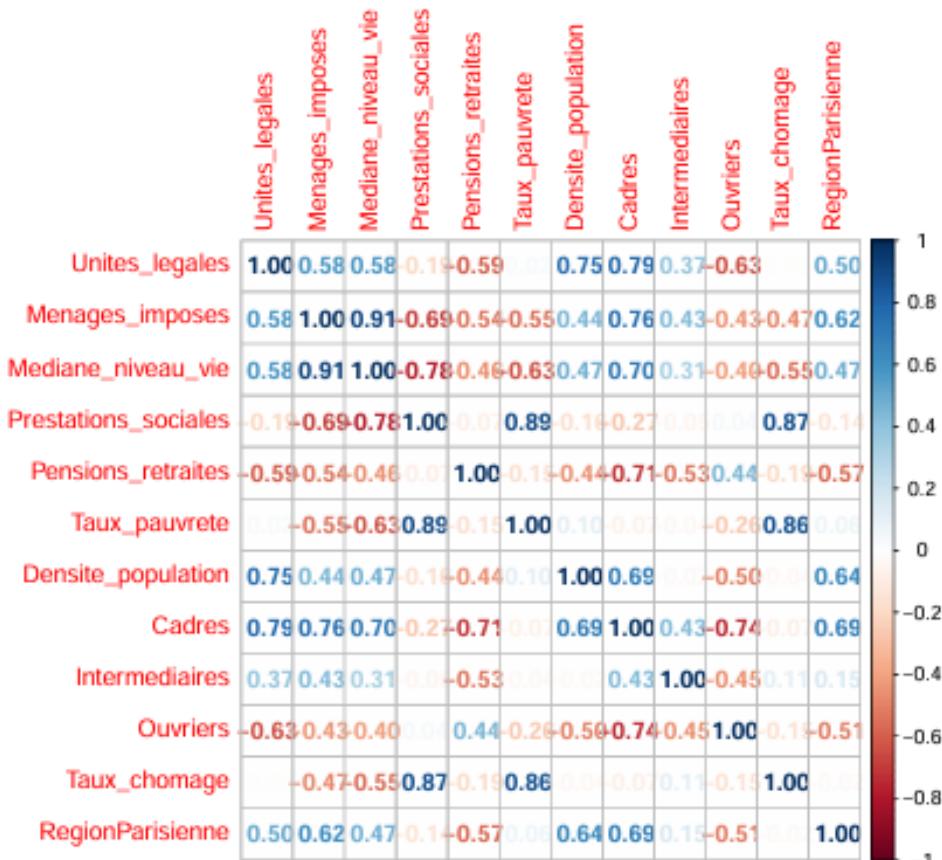


FIGURE 2.7 – Matrice de corrélation des variables explicatives

Interprétation

La matrice de corrélation (Figure 2.7) met en évidence les relations linéaires entre les variables explicatives. Voici les principales observations :

1. Corrélations fortes et positives :

- **Unites_legales et Non_scolarisés_15_plus** ($r = 0.85$) : Les deux variables montrent une forte corrélation positive, suggérant qu'une augmentation du nombre d'unités légales est liée à une proportion plus élevée de non-scolarisés de 15 ans et plus.
- **Menages_imposes et Mediane_niveau_vie** ($r = 0.91$) : Un lien très fort existe entre ces deux variables, ce qui est attendu, car les ménages imposés sont souvent liés à des niveaux de vie plus élevés.

- **Prestations_sociales et Taux_chomage** ($r = 0.87$) : Une forte corrélation entre ces deux variables montre que les prestations sociales augmentent dans les zones où le chômage est élevé.

2. Corrélations fortes et négatives :

- **Taux_pauvreté et Taux_activité** ($r = -0.72$) : Les zones ayant des taux d'activité élevés semblent avoir des taux de pauvreté plus faibles.
- **Prestations_sociales et Mediane_niveau_vie** ($r = -0.78$) : Dans les zones où la médiane du niveau de vie est élevée, les prestations sociales sont généralement plus faibles.
- **Cadres et Ouvriers** ($r = -0.74$) : Une forte relation négative, montrant qu'une proportion plus élevée de cadres est associée à une proportion plus faible d'ouvriers.
- On remarque qu'il existe une relation non négligeable entre le taux d'activité et le taux de chômage. Ces deux variables pourraient avoir les mêmes effets sur la variable endogène (les individus au chômage sont les individus qui ne sont pas en activité). Par la suite nous allons retirer une de ces variables de notre modèle, nous avons choisi la variable taux activité.

2.3 Analyse économétrique

2.3.0.1 Choix des variables explicatives

Le choix des variables explicatives s'est basé sur :

- **Matrice de corrélation** : Nous avons vérifié la corrélation entre les variables explicatives afin d'identifier et d'éliminer celles fortement corrélées entre elles, afin de limiter les problèmes de multicolinéarité.

2.3.0.2 Modèles estimés

Nous avons testé trois spécifications différentes pour expliquer la *médiane du niveau de vie* en fonction des variables explicatives sélectionnées. Les modèles estimés sont les suivants :

- **Modèle 1 (niveau-niveau) :**

$$\begin{aligned} \text{Mediane_niveau_vie}_i = & \beta_0 + \beta_1 \text{Unites_legales}_i + \beta_2 \text{Menages_imposes}_i + \beta_3 \text{Prestations_sociales}_i \\ & + \beta_4 \text{Pensions_retraites}_i + \beta_5 \text{Taux_pauvreté}_i + \beta_6 \text{Densite_population}_i + \beta_7 \text{Cadres}_i + \beta_8 \text{Intermediaires}_i \\ & + \beta_9 \text{Ouvriers}_i + \beta_{10} \text{Taux_chomage}_i + \beta_{11} \text{Type_zone}_i + \beta_{12} \text{RegionParisienne}_i + \epsilon_i \end{aligned}$$

- **Modèle 2 (niveau-log) :**

$$\begin{aligned} \text{Mediane_niveau_vie}_i = & \beta_0 + \beta_1 \log(\text{Unites_legales}_i) + \beta_2 \text{Menages_imposes}_i + \beta_3 \log(\text{Prestations_sociales}_i) \\ & + \beta_4 \text{Pensions_retraites}_i + \beta_5 \text{Taux_pauvreté}_i + \beta_6 \text{Densite_population}_i + \beta_7 \text{Cadres}_i + \beta_8 \text{Intermediaires}_i \\ & + \beta_9 \text{Ouvriers}_i + \beta_{10} \text{Taux_chomage}_i + \beta_{11} \text{Type_zone}_i + \beta_{12} \text{RegionParisienne}_i + \epsilon_i \end{aligned}$$

— **Modèle 3 (log-log) :**

$$\begin{aligned} \text{log}(\text{Mediane_niveau_vie}_i) = & \beta_0 + \beta_1 \text{log}(\text{Unites_legales}_i) + \beta_2 \text{Menages_imposes}_i + \beta_3 \text{log}(\text{Prestations_sociales}_i) \\ & + \beta_4 \text{Pensions_retraites}_i + \beta_5 \text{Taux_pauvreté}_i + \beta_6 \text{Densité_population}_i + \beta_7 \text{Cadres}_i + \beta_8 \text{Intermediaires}_i \\ & + \beta_9 \text{Ouvriers}_i + \beta_{10} \text{Taux_chomage}_i + \beta_{11} \text{Type_zone}_i + \beta_{12} \text{RegionParisienne}_i + \epsilon_i \end{aligned}$$

2.3.0.3 Critère de sélection des modèles

TABLEAU 2.1 – Comparaison des R^2 ajustés des trois modèles

Modèle	Forme du modèle	R^2 ajusté
Modèle 1	Médiane_niveau_vie ~ Unites_legales, Prestations_sociales, etc.	0.9379
Modèle 2	Médiane_niveau_vie ~ log(Unites_legales), log(Prestations_sociales), etc.	0.9593
Modèle 3	log(Médiane_niveau_vie) ~ log(Unites_legales), log(Prestations_sociales), etc.	0.9632

Le critère utilisé pour comparer les trois modèles était le **R^2 ajusté**. Parmi les trois modèles testés, le **modèle 3 (log-log)** s'est avéré être le plus performant en termes de capacité explicative des variations de la médiane du niveau de vie.

2.3.0.4 Conclusion

Le modèle final (log-log) corrigé pour l'hétérosécédasticité fournit une bonne capacité explicative des disparités de niveau de vie entre les départements de France, tout en respectant les hypothèses fondamentales de la régression linéaire.

2.3.0.5 Résultats de la régression

TABLEAU 2.2 – Comparaison des estimations MCO, MCO White, et MCO Newey pour le modèle log-log

Variable	MCO	MCO White	MCO Newey
log(Unites_legales)	-0.003 (0.004)	-0.003 (0.004)	-0.003 (0.004)
Menages_imposes	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
log(Prestations_sociales)	-0.161*** (0.018)	-0.161*** (0.021)	-0.161*** (0.021)
Pensions_retraites	-0.002*** (0.001)	-0.002*** (0.001)	-0.002*** (0.001)
Taux_pauvrete	-0.005*** (0.002)	-0.005*** (0.002)	-0.005*** (0.002)
log(Densite_population)	0.005 (0.005)	0.005 (0.006)	0.005 (0.006)
Cadres	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Intermediaires	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)
Ouvriers	-0.002** (0.001)	-0.002** (0.001)	-0.002** (0.001)
Taux_chomage	0.006*** (0.002)	0.006* (0.002)	0.006 (0.002)
RegionParisienne	-0.025** (0.012)	-0.025* (0.013)	-0.025 (0.015)
Type_zoneUrbain	-0.002 (0.005)	-0.002 (0.005)	-0.002 (0.005)
Constant	10.257*** (0.086)	10.257*** (0.109)	10.257*** (0.111)
Observations	98		
R²	0.968		
Adjusted R²	0.963		
Residual Std. Error		0.015 (df = 85)	
F-Statistic		212.33*** (df = 12; 85)	

Interprétation

La Table 2.2 compare les estimations obtenues avec les méthodes MCO, MCO White (robuste à l'hétérosclélasticité) et MCO Newey-West (robuste aux problèmes d'hétérosclélasticité et d'autocorrélation). Voici les points clés :

- **Variables significatives :**
 - **log(Prestations_sociales)** : La variable a un effet négatif significatif ($p < 0.01$) sur la médiane du niveau de vie, montrant qu'une forte dépendance aux prestations sociales est associée à des revenus médians plus bas.
 - **Menages_imposes** et **Taux_chomage** : Ces variables ont un effet positif significatif, reflétant une contribution importante au niveau de vie départemental.
 - **Pensions_retraites**, **Taux_pauvrete**, et **Ouvriers** : Ces variables ont un effet négatif significatif, indiquant leur lien avec des niveaux de vie médians plus faibles.
- **Stabilité des coefficients** : Les coefficients des variables significatives sont globalement stables entre les trois méthodes (MCO, White, Newey-West), témoignant de la robustesse du modèle.
- **Performances globales** : Avec un R^2 ajusté de 0.963, le modèle explique près de 96% des variations de la médiane du niveau de vie entre départements, soulignant sa forte capacité explicative.

2.3.0.6 Qualité du modèle

Les critères de qualité montrent que le modèle est statistiquement significatif et robuste. Voici les points essentiels :

1. **Statistique de Fisher** : La statistique de Fisher ($F = 212.33, p < 0.01$) confirme que le modèle est globalement significatif, c'est-à-dire que les variables explicatives expliquent une part importante de la variance de la médiane du niveau de vie.
2. **R^2 ajusté** : Le R^2 ajusté de 0.963 montre que le modèle explique **96.3%** des disparités de revenus, reflétant une très bonne capacité explicative.
3. **Hétéroscédasticité** : Le test de Breusch-Pagan a révélé la présence d'hétéroscédasticité dans les résidus, avec une statistique $BP = 47.319, df = 12$, et une p -value de 4.105×10^{-6} . Cette p -value étant inférieure au seuil de 5%, nous rejetons l'hypothèse nulle d'homoscédasticité des erreurs. De plus, le test de Goldfeld-Quandt a confirmé cette présence d'hétéroscédasticité ($GQ = 6.6288, p$ -value de 2.124×10^{-7}). Pour corriger ce problème, des erreurs standards robustes (méthodes White et Newey-West) ont été utilisées, garantissant ainsi la fiabilité des tests d'hypothèses.
4. **Autocorrélation** : Le test de Durbin-Watson a donné une statistique $DW = 1.9826$ avec une p -value de 0.3889, indiquant que l'autocorrélation des résidus n'est pas significative au seuil de 5%. Ainsi, nous ne rejetons pas l'hypothèse nulle d'absence d'autocorrélation. Malgré cela, la méthode de Newey-West a été appliquée par précaution pour garantir des estimations robustes en présence de tout potentiel d'autocorrélation non détectée.

Le modèle final est robuste et pertinent pour analyser les disparités de revenus entre départements en France.

2.3.0.7 Interprétation des résultats

1. Coefficients positifs significatifs :

- **Menages_imposes (+0.004)** : Une augmentation de 1% de la proportion de ménages imposés est associée à une augmentation de 0.4% de la médiane du niveau de vie. *Interprétation économique* : Cela reflète un lien direct entre la richesse des ménages et le niveau de vie départemental.
- **Taux_chomage (+0.006)** : Une augmentation de 1% du taux de chômage est associée à une hausse de 0.6% de la médiane du niveau de vie. *Interprétation économique* : Bien que contre-intuitif, ce résultat peut s'expliquer par le fait que certains départements riches offrent des mécanismes de compensation sociale élevés.

2. Coefficients négatifs significatifs :

- **log(Prestations_sociales (-0.161))** : Une augmentation de 1% des prestations sociales dans le revenu disponible est associée à une baisse de 16.1% de la médiane du niveau de vie. *Interprétation économique* : Une dépendance élevée aux prestations sociales reflète souvent des économies en difficulté.
- **Pensions_retraites (-0.002)** : Une augmentation de 1% des pensions et retraites dans les revenus disponibles est associée à une baisse de 0.2% de la médiane du niveau de vie.

Interprétation économique : Cela reflète une population vieillissante, souvent moins active économiquement.

- **Taux_pauvrete** (-0.005) : Une augmentation de 1% du taux de pauvreté est associée à une baisse de 0.5% de la médiane du niveau de vie. *Interprétation économique* : Une relation attendue : les départements pauvres affichent des niveaux de vie plus faibles.
- **Ouvriers** (-0.002) : Une augmentation de 1% de la proportion d'ouvriers parmi les emplois locaux est associée à une baisse de 0.2% de la médiane du niveau de vie. *Interprétation économique* : Les secteurs dominés par les ouvriers sont souvent moins rémunérateurs.

2.4 Vérification des hypothèses

TABLEAU 2.3 – Vérification des hypothèses

Hypothèses	Conclusion
H1 : Les départements avec une forte concentration d'activités économiques (mesurée par le nombre d'unités légales) et une proportion élevée de cadres ont un revenu médian significativement plus élevé.	Rejeté : La proportion de cadres montre un effet significatif et positif sur le revenu médian, mais le nombre d'unités légales n'est pas significatif.
H2 : Les départements fortement dépendants des prestations sociales et ayant un taux de pauvreté élevé présentent un revenu médian significativement plus faible.	Validée : Les prestations sociales et le taux de pauvreté ont un effet significatif et négatif sur le revenu médian.

Source : Auteurs, 2024

Conclusion

Cette étude avait pour objectif d'analyser les disparités de revenus entre les départements français et d'identifier les principaux facteurs explicatifs à travers une approche économétrique rigoureuse. Les résultats obtenus ont permis de dégager plusieurs enseignements majeurs tout en soulevant des questions sur l'interprétation des relations identifiées.

L'analyse économétrique a montré que les disparités de revenus départementales sont largement expliquées par des facteurs socio-économiques et démographiques. Parmi les principaux résultats :

- Une proportion élevée de ménages imposés est associée à un revenu médian plus élevé, confirmant le rôle central de la richesse économique locale.
- À l'inverse, une forte dépendance aux prestations sociales, un taux de pauvreté élevé, et une population vieillissante (pensions retraites) contribuent négativement au niveau de vie médian.
- Le taux de chômage, bien qu'habituellement perçu comme un indicateur négatif, montre un effet positif significatif, probablement dû à l'importance des mécanismes de compensation sociale dans certains départements.

Le modèle final (*log-log*), avec un R^2 ajusté de 0.963, explique une part importante des disparités observées, attestant de la pertinence des variables incluses.

Malgré leur robustesse, les résultats de cette étude doivent être interprétés avec prudence. En effet, les relations identifiées sont basées sur des corrélations et ne permettent pas d'établir une causalité directe. Par exemple :

- La forte dépendance aux prestations sociales peut être une conséquence d'un faible niveau de vie, et non sa cause. Cela soulève la question de l'endogénéité des variables.
- L'effet positif du taux de chômage sur le revenu médian pourrait résulter d'une corrélation indirecte liée aux mécanismes sociaux (indemnités chômage, structures économiques) plutôt qu'à un lien causal direct.

Ces limites montrent la nécessité d'approches méthodologiques complémentaires (comme des modèles structurels ou des analyses longitudinales) pour mieux cerner les dynamiques causales.

Cette étude offre une base solide pour orienter les politiques publiques visant à réduire les disparités de revenus entre départements. Les recommandations incluent :

- Encourager le développement économique local dans les départements à faible proportion de ménages imposés.
- Réduire la dépendance aux prestations sociales par des politiques favorisant l'emploi et l'autonomie économique.
- Prendre en compte les spécificités démographiques, notamment le vieillissement de la population, dans les stratégies de développement.

Enfin, cette étude appelle à des recherches futures intégrant des approches longitudinales, des données spatiales, ou des modèles d'analyse causale pour affiner la compréhension des dynamiques territoriales et leur impact sur les inégalités de revenus.

2.5 Limites de l'étude

Malgré les efforts déployés pour garantir la robustesse des résultats, certaines limites méritent d'être soulignées :

- **Définition du revenu médian :** La médiane du niveau de vie, utilisée comme indicateur central, peut ne pas capturer toutes les disparités économiques internes aux départements, notamment celles liées aux inégalités intradépartementales.
- **Agrégation géographique :** L'analyse repose sur des données agrégées au niveau départemental, ce qui peut masquer des disparités significatives au niveau communal ou individuel.
- **Influence des variables non observées :** Certains facteurs potentiellement déterminants, comme la qualité des infrastructures ou les dynamiques migratoires, n'ont pas été inclus, faute de données disponibles.
- **Temporalité des données :** Les données utilisées sont issues d'une année spécifique, ce qui limite la capacité à analyser les évolutions temporelles ou les tendances longues des disparités de revenus.

Ces limites ne remettent pas en cause les résultats obtenus, mais elles suggèrent des pistes d'amélioration pour des études futures, notamment en intégrant des données plus détaillées et en explorant des approches multi-échelles.

Références

- [1] Faye, N. F., Sall, M., Affholder, F., et al. (2019). « Inégalités de revenu en milieu rural dans le bassin arachidier du Sénégal ». *Papiers de recherche*, pp. 1-54.
- [2] Rodgers, G. (2022). « Changements de perspectives sur la pauvreté et les inégalités : l'apport de la Revue internationale du Travail ». *Revue internationale du Travail*, 161(4), e71-e82.
- [3] Sawadogo, D. (2023). « Croissance économique et inégalités de revenu dans les pays en développement : rôle des institutions ». *Revue Ouest Africaine de Sciences Economiques et de Gestion*, 16(2).
- [4] Fayat, H. (2022). *Socio-histoire des inégalités*. La Découverte.
- [5] Morrisson, C., et Snyder, W. (1998). « Les inégalités de revenu en France depuis le début du XVIIIe siècle ». Paris : Mimeo.
- [6] Union Européenne. (n.d.). « Inégalités de revenu en milieu rural dans le bassin arachidier du Sénégal ».

Table des matières

Résumé	iv
Abstract	v
Introduction	1
1 Cadre théorique et méthodologique de l'étude	2
1.1 Problématique et intérêt de l'étude	2
1.2 Objectifs et hypothèses de l'étude	3
1.2.1 Objectifs de l'étude	3
1.2.2 Hypothèses	3
1.3 Méthodologie	4
1.3.1 Nature et sources des données	4
1.3.2 Méthodes et outils d'analyse	4
1.4 Présentation des variables	7
2 Résultats, analyses et interprétations	8
2.1 Analyse descriptive	8
2.1.1 Analyse univariée	8
2.1.2 Analyse bivariée	9
2.2 Test de pré-estimation	12
2.3 Analyse économétrique	13
2.4 Vérification des hypothèses	18
Conclusion	19
2.5 Limites de l'étude	20
Références	20