

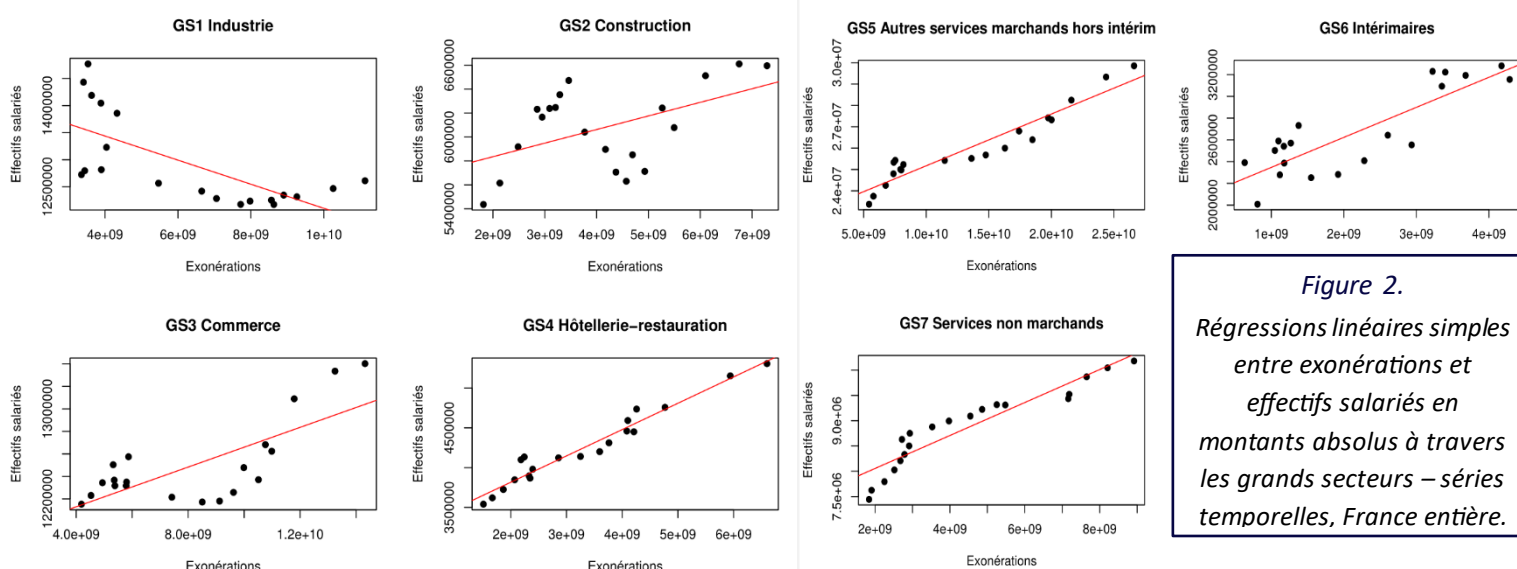
Explication de ma démarche.

I. Première analyse sur les secteurs d'activité, en valeurs absolues.

Disposant de séries de données temporelles d'exonérations et d'effectifs salariés finement discriminées par secteur et sous-secteur d'activité, j'ai souhaité dans un premier temps me concentrer sur la relation entre les allègements fiscaux et l'emploi dans chacun des sept grands secteurs économiques à travers le temps, à échelle nationale.

Pour ce faire, j'ai dû effectuer un travail préliminaire de traitement de données, en surface. Sous R, j'ai extrait les valeurs des indicateurs pertinents de chacune des deux bases de données concernées, en l'occurrence les montants absolus des effectifs salariés et d'exonérations, que j'ai agrégés par année et par secteur. Les données que je possédais ne m'ont permis d'étendre la période d'analyse que de 2004 à 2023, et ce tout au long de mon travail.

J'ai ainsi pu effectuer une première série de régressions simples, basées sur des modèles niveau-niveau, et prendre une première mesure des évolutions temporelles de mes deux indicateurs. En voici les graphes.



En dehors du secteur de l'industrie, une corrélation largement positive entre exonérations fiscales et effectifs salariés apparaissait pour chacun des grands secteurs de l'économie. Mais cette corrélation ne pouvait être appréciée telle quelle, du fait des limites que présente l'intégration de valeurs absolues, et puisqu'un biais lié à une tendance temporelle générale à la hausse pouvait exister pour chacune des variables. La seule conclusion significative à tirer de ces graphiques demeure selon moi la confirmation d'une rupture entre l'industrie et les autres secteurs, étant sans doute une matérialisation d'une forte tendance à la désindustrialisation dans le pays.

II. Deuxième analyse sur les anciennes régions, en valeurs relatives.

En suivant les conseils de mon enseignant, j'ai cherché à ajuster l'indicateur d'emploi en passant à une mesure relative, et en réduisant l'échelle géographique prise en compte. Cela devait permettre de corriger une partie de la tendance naturelle à la hausse des effectifs salariés, et améliorer la précision de la mesure des répercussions directes de l'exonération. Il est à noter que cette décision m'a contraint de mettre de côté la démarche de distinction des valeurs par secteur d'activité, du fait d'un manque de données relatives discriminant à la fois régions et secteurs.

Dès lors, j'ai entrepris de construire une base de données contenant la part de personnes employées au fil du temps, au sein de régions bien définies. Les nouvelles régions étant à mes yeux trop larges, j'ai préféré partir des données démographiques disponibles pour les *anciennes* régions, étant au nombre de vingt-six et me paraissant mieux adaptées à l'observation de tendances locales.

Bien loin de ce que j'avais imaginé, la compilation de données nécessaire à la confection de la base en question s'est révélée très fastidieuse, principalement du fait que les recensements soient conduits à une faible fréquence, et que les anciennes régions soient (c'est un pléonasme) caduques. Pour le dénominateur (i. e. la population totale par région), mon point de départ a été une base de l'Insee contenant les données de populations issues de *chaque* commune de France, mesurées à plusieurs années d'intervalle.

Pour la convertir en une base complète et exploitable, je devais réaliser deux manipulations, la première étant l'agrégation des données des communes par ancienne région, la seconde étant la prédiction des niveaux de population entre les années de recensement. Plusieurs étapes auront été nécessaires, en voici le détail :

1. Génération d'une liste de départements par ancienne région en CSV via une intelligence artificielle générative (*Claude.ai*).
2. Passage du format CSV en virgules au format en points-virgules avec Python.
3. Agrégation des données de population par année et par ancienne région via VBA (cf. annexe II.), et remplissage d'un tableau. *Observation : les totaux ne correspondent pas à la réalité.*
4. Ajustement des valeurs agrégées en multipliant par un coefficient obtenu en comparant les totaux empiriques et la population nationale officielle (en s'assurant que Mayotte ne soit jamais pris en compte car trop peu de données disponibles).
5. Vérification sur internet des données obtenues. *Conclusion : succès.*
6. Estimation des valeurs de populations entre les années de recensement par interpolation linéaire à l'aide de formules sur Excel.

Une fois ce travail accompli, il devenait possible de réaliser de nouvelles régressions linéaires entre exonérations et emploi à plus petite échelle et surtout, en prenant en compte la réalité démographique de chacun des indicateurs.

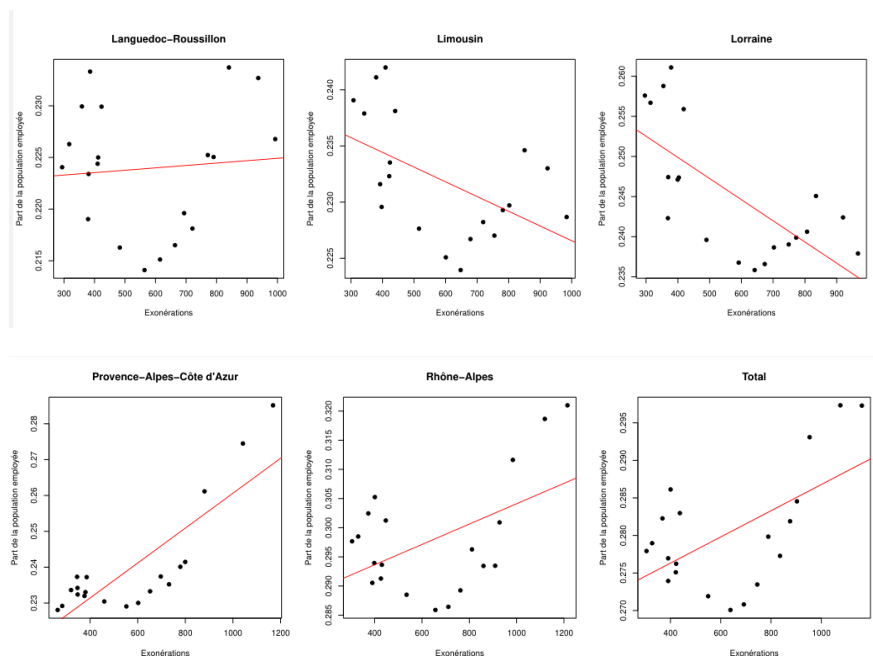
Ma démarche fut la suivante : j'ai effectué sous R un travail de remaniement des données similaire à celui effectué en première partie, cette fois en agrégeant les valeurs temporelles par ancienne région, et en ajoutant le dénominateur commun de population tant aux exonérations qu'aux effectifs salariés.

De ces manipulations ont pu découler la réalisation d'autant de modèles de régression linéaire que de régions (plus un, avec le total), et j'obtins un résultat pour le moins surprenant. A quelques exceptions près, l'ensemble des graphiques associés aux régressions affichaient une relation non linéaire entre les valeurs relatives d'exonérations et d'emploi (voir figure 3 et annexe II.).

Figure 3.

Régressions linéaires entre les montants d'exonérations fiscales aux entreprises par habitant et la part de la population étant employée. Séries temporelles – régions.

Observation : une distribution en 'U' apparaît sur la quasi-totalité des 27 graphiques.



Avant d'en tirer une conclusion hâtive sur la nature de la relation entre les indicateurs, il m'a semblé nécessaire d'ajuster mon travail selon deux observations importantes. D'une part, les distributions des couples exonérations - part employée étaient si semblables visuellement à travers plusieurs graphiques (cela est particulièrement manifeste ici entre la région Rhône-Alpes et le total) qu'il était vraisemblable qu'une tendance temporelle se soit immiscée dans les séries de données brutes, qu'il faudrait corriger. D'autre part, une rupture marquée entre les distributions associées aux régions en métropoles et celles des régions d'outre-mer laissait entendre qu'il serait pertinent de distinguer les tendances selon ces deux catégories.

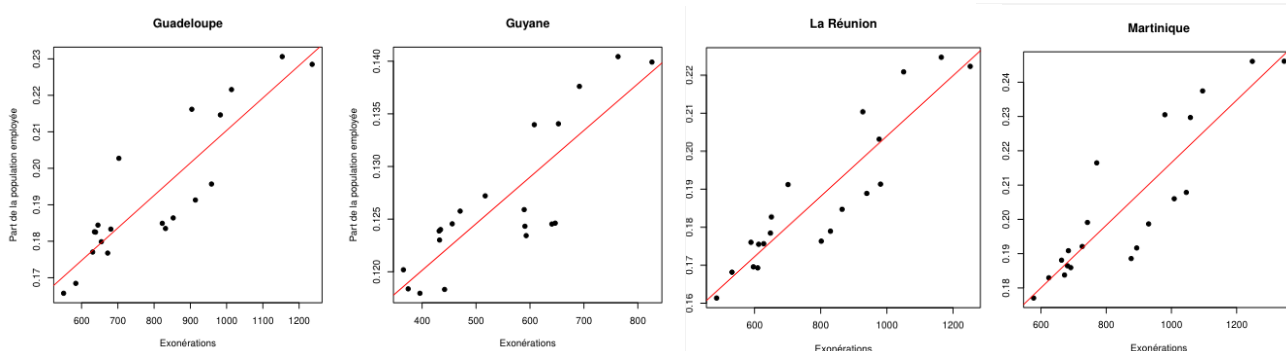


Figure 4.

(cf. figure 3)

Les régions d'O.-M se distinguent par la linéarité de leur tendance.

III. Détection des tendances temporelles.

Afin de confirmer ou d'infirmer mes impressions à l'égard de l'existence de tendances temporelles dans les séries de chacun de mes indicateurs, j'ai isolé mes deux variables et ai reporté leurs valeurs sur des graphiques chronologiques, sous R ; laissant place à une première phase de simple observation. Les résultats étaient très éloquentes : les exonérations par habitant connaissent une croissance linéaire invariable au fil des ans, sans exception tandis que la part de la population employée de toutes les régions à l'exception des régions d'outre-mer avait évolué de façon non-linéaire, de telle sorte qu'une distribution en 'U' apparaissait à nouveau (cf. figures 5 et 6 et annexe III.).

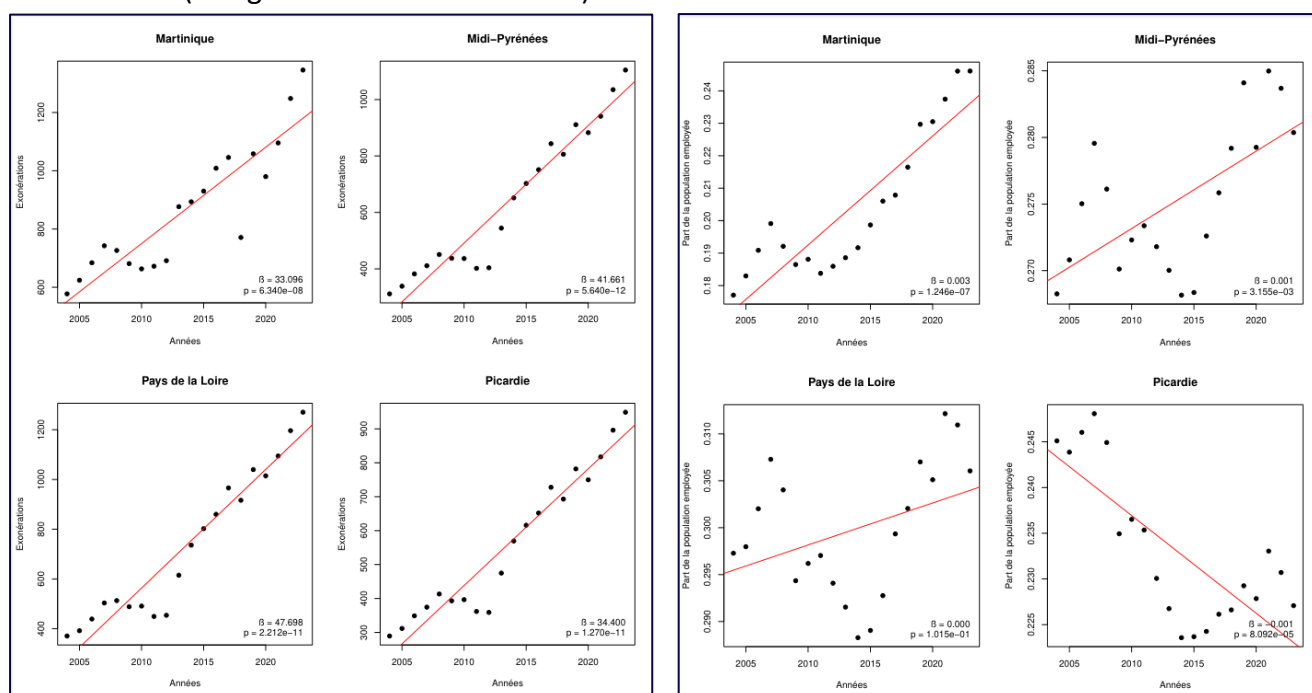


Figure 5 et 6.

A gauche : évolution chronologique des exonérations par habitant.

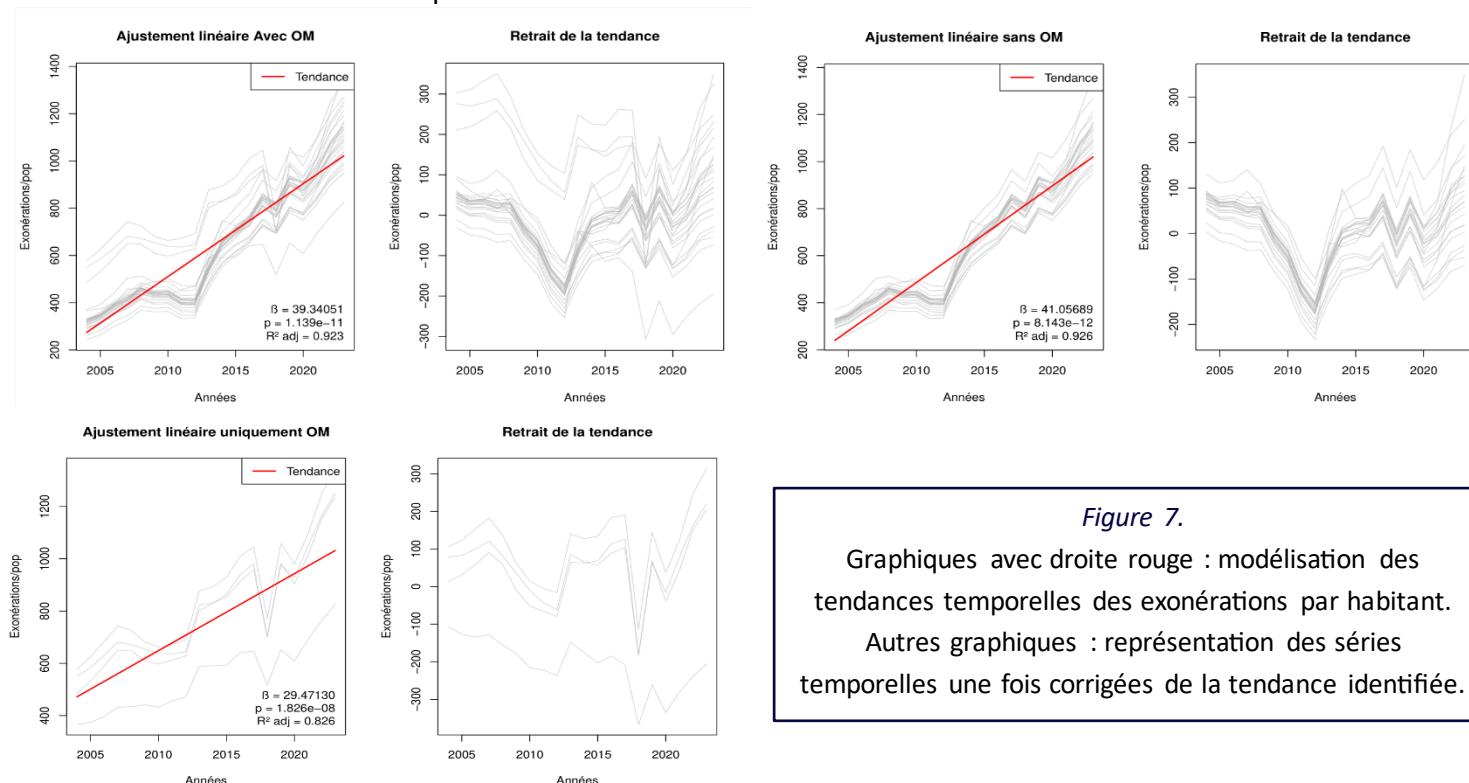
A droite : évolution chronologique de la part de la population employée.

Observation : les territoires d'O.-M. marquent la rupture par l'évolution linéaire de leur part employée.

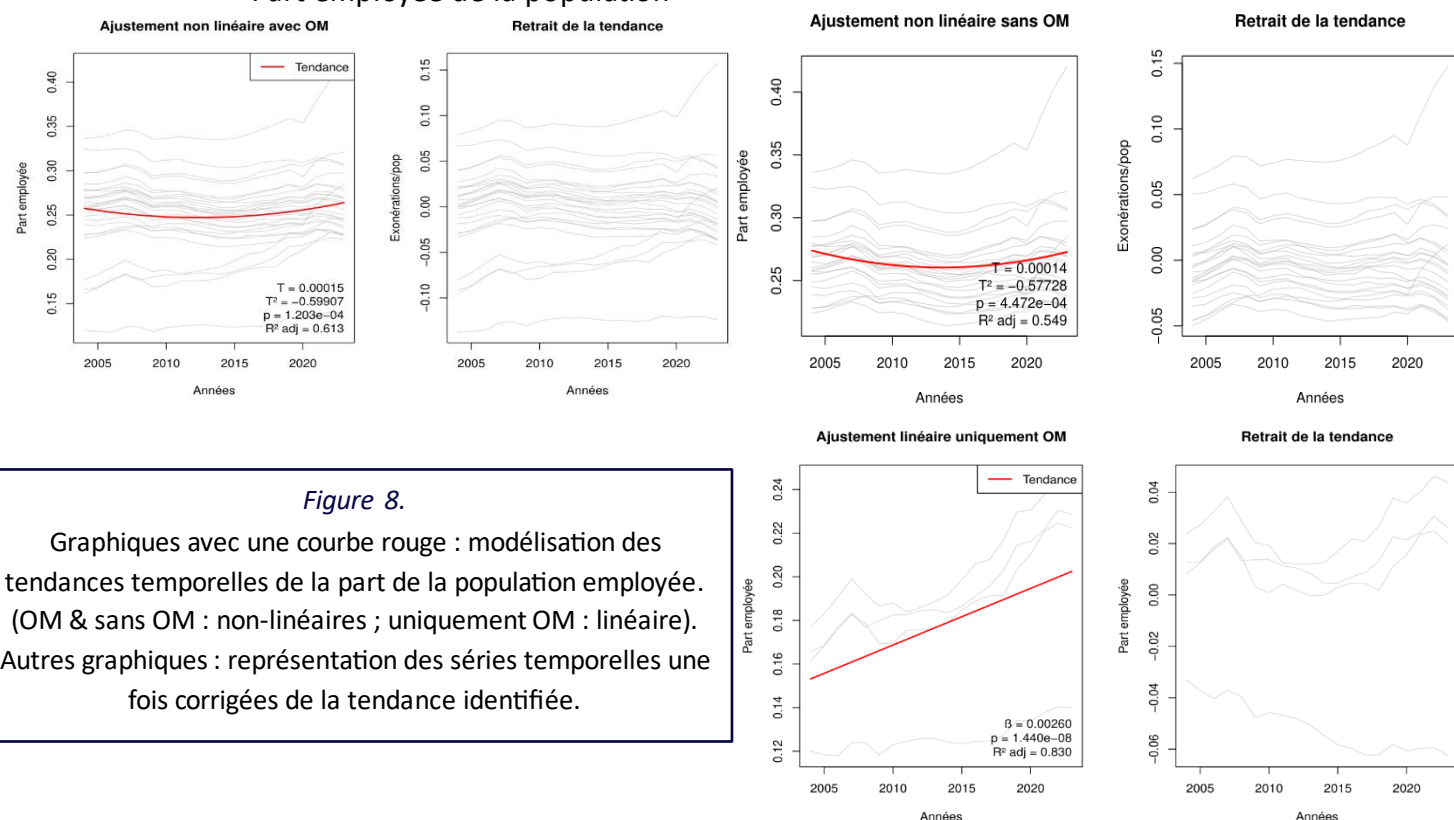
Il me fallait désormais exploiter ces observations statistiquement, en identifiant une tendance claire pour chacune des variables et en corrigeant les séries temporelles, avant d'effectuer une nouvelle série de régressions linéaires. A cette fin, j'ai tout naturellement choisi de créer une prédiction de tendances statistiques respectivement linéaire pour les exonérations par habitant et quadratique pour la part de la population employée. Par ailleurs, bien que les territoires d'outre-mer ne soient pas les uniques exceptions à la conjecture d'une distribution non-linéaire de la part employée, j'ai décidé de faire de leur caractéristique un critère dans la réalisation de mes modèles, et ce jusqu'à la fin de mon travail de modélisation.

Ce critère m'a conduit à distinguer trois séries de modèles pour chaque réalisation statistique : une prenant en compte l'ensemble des régions, une seconde excluant les régions d'outre-mer, et une dernière n'incluant que les outre-mer (avec un ajustement linéaire y compris pour la part employée) ; en ré-évaluant les totaux selon les données prises en compte. Voici, pour chacune de ces séries, les représentations tendances identifiées ainsi que des distributions de données corrigées de leur tendance (voir aussi annexe III.).

- Exonérations par habitant :



- Part employée de la population



L'identification des tendances à travers les séries, et en prenant en compte la caractéristique métropole/outre-mer m'a conduit aux conclusions suivantes. D'une part, la tendance non-linéaire que j'avais cru identifier lors de ma première analyse de surface semblait valable, puisque l'ajustement effectué sur les données de part employée de la population expliquait plus de 55% des variations dans les deux cas de figure. Néanmoins, je ne m'appliquerai pas à expliquer les raisons de cette évolution conjoncturelle.

Ensuite, il semblerait que le critère d'appartenance aux territoires d'outre-mer soit pertinent dans l'analyse des tendances des régions d'outre-mer elles-mêmes, mais que leur écartement fasse souffrir l'analyse des régions restantes d'une moins grande significativité statistique (R^2_a passant de 61 à 55% entre les deux modèles).

Enfin, malgré la persistance de variations communes importantes (comme le pic vers le bas des exonérations en 2012 ou la remontée de celles-ci après 2020), les distributions de données des deux indicateurs semblent effectivement se stabiliser autour de zéro après correction des tendances. J'ai donc jugé pertinent de prendre en compte ces tendances dans une réédition des modèles de régression décrits dans ma deuxième partie.

IV. Application des tendances temporelles.

Ayant choisi de fonder l'ajustement des tendances à un niveau moyen sur l'ensemble des régions prises en compte par chaque série (avec OM, sans OM et uniquement OM), il m'a suffi, après une réévaluation des totaux pour chaque série, de soustraire les valeurs prédites aux valeurs empiriques traitées précédemment. Voici les graphiques correspondant aux modèles les plus pertinents dans l'illustration de mon analyse conclusive (je ne présenterai ici que les séries basées sur une distinction des outre-mer, voir la série *avec OM* dans l'annexe IV.).

- Série sans les territoires d'outre-mer :

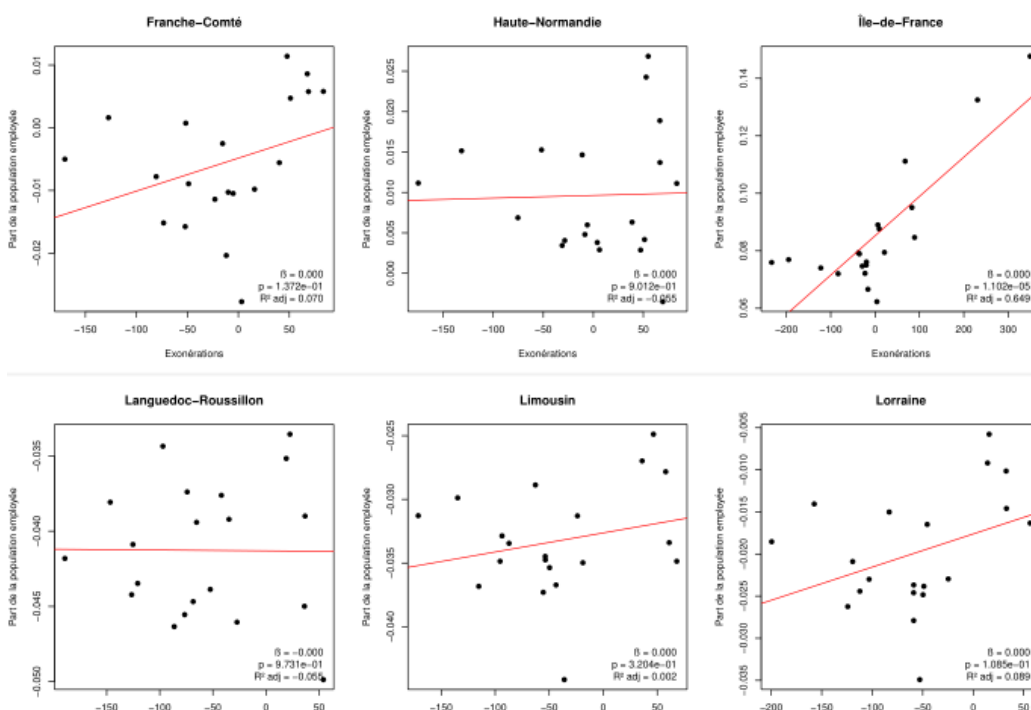


Figure 9.

Régressions linéaires simples entre exonérations par habitant corrigées par ajustement quadratique et part de la population employée corrigée par ajustement linéaire – séries temporelles.

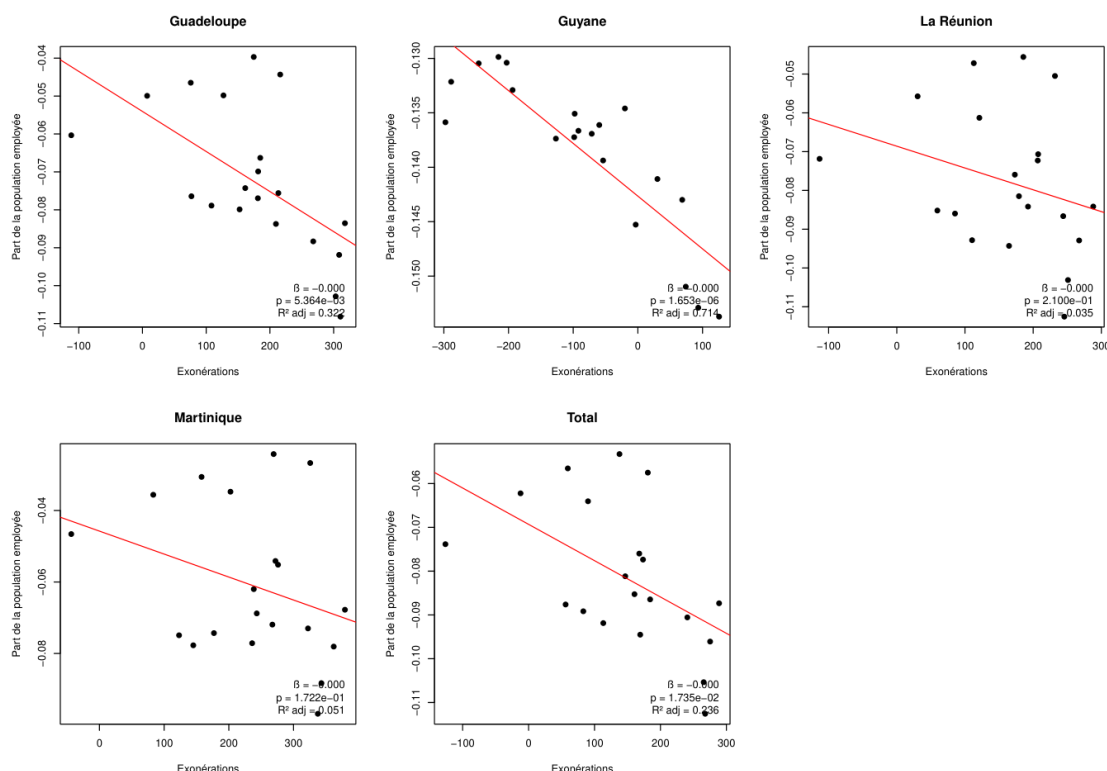
Obs. : l'Île-De-France est la seule région de la série pour laquelle le modèle explique une part significative des variations ($R^2_a \approx 65\%$).

- Série outre-mer uniquement :

Figure 10.

Régressions linéaires simples entre exonérations par habitant et part de la population employée corrigées par ajustements linéaires – séries temporelles.

Obs. : la Guadeloupe et la Guyane ont fait l'objet d'ajustements linéaires à coefficients négatifs, hautement explicatifs (valeurs de R^2_a à 32 et 71%).



Les graphiques présentés dans la figure 9 permettent d'illustrer un fait valable pour l'ensemble des régressions effectuées sur les régions en métropole : la correction des tendances temporelles des deux indicateurs mène à une disparition de toute corrélation significative entre eux. Les coefficients de déterminations des différentes régressions sont tous très proches de zéro et la p-value des coefficients de régressions très proche de 1, avec un cas critique en Aquitaine (p-v. = 99,7%) (cf. annexe IV.). Il semble donc à ce stade difficile d'inférer un quelconque lien de causalité entre exonérations et amélioration de l'emploi dans ces régions.

Une seule exception à ce constat demeure : le cas de l'Île-de-France. Les données corrigées permettent d'établir une relation linéaire très significative entre part de la population employée et exonérations par habitant, et le modèle explique une part exceptionnellement importante de la variance totale, puisque le coefficient de détermination ajusté prend une valeur supérieure à 60%. J'interprèterais cette particularité parisienne comme relevant de la singularité (comparément à la province) de la structure du marché du travail dans la région. Il est vraisemblable que la forte concentration démographique de l'endroit favorise de plus grandes fluidité et fréquence d'interaction entre les agents, rendant plus à même les patrons d'entreprises locales de répercuter leurs avantages fiscaux sur leur processus d'embauche. Les données annuelles prises en considération dans cette étude sont néanmoins trop peu nombreuses pour en attester formellement.

En ce qui s'agit du modèle appliqué aux territoires d'outre-mer uniquement, une première observation notable est celle de la similitude apparente des distributions des données à travers la Réunion et les Antilles (se répercutant donc sur le total). Cette conjoncture commune à

travers des zones géographiques pourtant très éloignées semble indiquer qu'une centralisation des politiques menées dans ces territoires a bien un effet homogénéisant des variables économiques. Mais malgré ce constat (qui demeure superficiel et invérifié statistiquement), les indicateurs statistiques liés à la régression démontrent une variabilité importante dans la nature de l'effet des politiques d'exonérations entre ces régions. En effet, alors que les ajustements appliqués à la Martinique et à la Réunion semblent confirmer l'inexistence d'un lien entre les deux indicateurs, la Guyane et la Guadeloupe connaissent des sorts différents, se matérialisant par des résultats statistiques très inattendus. Une relation négative très significative entre exonérations fiscales et part de la population employée apparaît, avec un coefficient de détermination ajusté allant de 32% à 71% pour la Réunion.

Même au travers de mon prisme critique à l'égard des pratiques économiques concernées, cela demeure tout à fait contre-intuitif pour l'instant. Un aspect de ma démarche aurait-il pu conférer davantage de biais aux distributions plutôt que de les en épurer ? Se pourrait-il qu'un mécanisme économique intermédiaire contraignant m'ait échappé ? Mon travail de traitement des données relatives aux régions d'outre-mer était-il inabouti ? Je laisse le loisir au lecteur de ce mémoire de consulter les programmes informatiques contenus dans l'annexe ainsi que sa propre culture économique pour tenter de répondre à ces interrogations lourdes de sens dans le contexte de ma problématique.

NB : en marge de ce travail, j'ai réalisé une dernière série de régressions en implémentant un décalage d'un an entre les exonérations et leurs supposés effets. Les résultats sont dans l'annexe V.