

# Analyse de l'écart salarial entre les genres au sein d'une entreprise fictive

## X

Approches économétriques et implications organisationnelles

Woodnalie S. JOSEPH

Décembre 2025

**Nature du projet :** Ce document constitue un projet d'apprentissage réalisé à des fins pédagogiques. Il s'appuie sur des *données fictives* issues de Kaggle pour explorer et approfondir des méthodes économétriques appliquées à l'analyse des inégalités salariales. L'objectif principal est de consolider mes compétences en analyse quantitative, décomposition statistique et interprétation de résultats empiriques.

Certains patterns observés (rendements éducatifs supérieurs pour les femmes, effet d'âge négatif) diffèrent de la littérature empirique sur données réelles et sont signe de limites inhérentes aux données simulées.

*Ce travail demeure un **exercice** en cours de perfectionnement et est ouvert à toute suggestion, correction ou commentaire constructif visant à améliorer sa rigueur méthodologique et sa clarté analytique.*

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Revue de littérature</b>	<b>1</b>
<b>3</b>	<b>Données et variables</b>	<b>2</b>
3.1	Valeurs manquantes . . . . .	3
3.2	Exploration des données . . . . .	3
3.2.1	Résumé des variables . . . . .	3
3.2.2	Distribution des niveaux d'éducation . . . . .	4
3.2.3	Distributions des variables clés . . . . .	4
3.2.4	Salaire par genre et éducation . . . . .	4
3.2.5	Corrélations entre les variables continues . . . . .	5
<b>4</b>	<b>Méthodologie</b>	<b>6</b>
4.1	Moindres carrés ordinaires (MCO) . . . . .	6
4.2	Décomposition d'Oaxaca-Blinder . . . . .	6
4.3	Régression quantile . . . . .	7
4.4	Appariement par score de propension (PMS) . . . . .	7
4.5	Synthèse . . . . .	8
<b>5</b>	<b>Résultats</b>	<b>8</b>
5.1	Moindres Carrés Ordinaires (MCO) . . . . .	9
5.1.1	Interprétation des coefficients . . . . .	9
5.2	Décomposition d'Oaxaca-Blinder . . . . .	10
5.2.1	Écart salarial brut de genre . . . . .	10
5.2.2	Décomposition globale . . . . .	10
5.2.3	Contribution des variables . . . . .	11
5.2.4	Différence de rendements par genre . . . . .	12
5.3	Régression Quantile . . . . .	12
5.4	Appariement par Score de Propension (PSM) . . . . .	13
5.5	Résumé . . . . .	13
<b>6</b>	<b>Discussion</b>	<b>14</b>
6.1	Une corrélation critique : le désavantage salarial lié à l'âge . . . . .	14
6.1.1	Limites d'interprétation et constats robustes . . . . .	14
6.2	Le paradoxe des rendements supérieurs . . . . .	14
6.3	Hétérogénéité distributionnelle . . . . .	15
6.3.1	Le plancher collant domine le plafond de verre . . . . .	15
6.4	Dimension organisationnelle et cadres théoriques . . . . .	15
6.4.1	Validation empirique des cadres théoriques . . . . .	15
6.4.2	Mécanismes organisationnels multicouches . . . . .	15
6.4.3	Le rôle central des effets d'interaction . . . . .	16
6.4.4	Valorisation organisationnelle différenciée selon l'âge . . . . .	16
<b>7</b>	<b>Implications pour l'entreprise X étudiée</b>	<b>16</b>
<b>8</b>	<b>Limites</b>	<b>16</b>
<b>9</b>	<b>Conclusion</b>	<b>17</b>
9.1	Quatre constats majeurs . . . . .	17
9.2	Contributions principales . . . . .	17
9.3	Implications pour l'action . . . . .	18
<b>10</b>	<b>Références</b>	<b>19</b>

<b>11 Annexes</b>	<b>21</b>
11.1 Statistiques descriptives par genre . . . . .	21
11.2 Équations complètes . . . . .	21
11.2.1 Décomposition d'Oaxaca-Blinder . . . . .	21
11.2.2 Régression quantile . . . . .	22
11.2.3 Appariement par score de propension . . . . .	22
11.3 Tests de robustesse et diagnostics . . . . .	22
11.3.1 Tests de diagnostic MCO . . . . .	22
11.3.2 Tests de spécification . . . . .	22
11.4 Décomposition d'Oaxaca-Blinder . . . . .	23
11.4.1 Note sur les structures de référence (Group_Weight) . . . . .	23

# 1 Introduction

En 2025, malgré des décennies de lutte pour l'équité, les femmes gagnent encore moins que les hommes. Rien qu'en France, cet écart atteignait 22.2% en faveur des hommes en 2023 (INSEE, 2023). Pourquoi cet écart persiste-t-il ? S'agit-il de différences de qualifications, ou de mécanismes structurels invisibles ?

Les disparités salariales fondées sur le genre demeurent parmi les formes d'inégalité les plus persistantes sur les marchés du travail à travers le monde. Les femmes continuent de gagner moins que les hommes en moyenne, **même lorsque des caractéristiques observables** telles que l'éducation, l'expérience et la profession **sont prises en compte** (Blau & Kahn, 2017; Forum économique mondial, 2023).

La littérature économique décompose l'écart salarial en deux composantes distinctes :

- les différences en matière de capital humain (composante "expliquée") et
- les différences résiduelles inexpliquées, souvent interprétées comme de la discrimination ou des barrières structurelles (Oaxaca, 1973; Blinder, 1973).

Les méthodes économétriques permettent de séparer ces effets et de fournir des preuves pour les interventions politiques (Fortin et al., 2011).

Cette analyse porte sur 6 700 employés d'une entreprise fictive (données publiées sur Kaggle par Ahmad Raza Kashif<sup>1</sup>). Dans un cadre organisationnel unique, les dynamiques internes se manifestent au-delà des analyses macroéconomiques. Nous combinons quatre méthodes complémentaires :

- Régression MCO : écart salarial moyen après contrôles
- Décomposition d'Oaxaca-Blinder : sépare "dotations" vs "rendements inégaux"
- Régression quantile : écart du bas au sommet de la distribution
- Appariement par score de propension : comparaison entre profils équivalents

La convergence des résultats à travers ces quatre approches renforce la fiabilité des constats et permet d'identifier les mécanismes qui façonnent les inégalités salariales au sein de cette organisation. Bien que la portée de l'étude se limite à un cas unique, les mécanismes identifiés s'inscrivent dans une littérature robuste et offrent des pistes concrètes pour la réflexion sur les pratiques organisationnelles.

## 2 Revue de littérature

Les disparités salariales entre les genres ont été analysées à travers trois cadres théoriques principaux.

La **théorie du capital humain** attribue les différences salariales aux variations dans les investissements en éducation, formation et expérience professionnelle (Becker, 1964 ; Mincer & Polachek, 1974). Les travaux récents nuancent cette approche : Goldin (2014) démontre que les écarts persistent même après contrôle pour l'éducation et l'expérience, et identifie la flexibilité temporelle comme facteur déterminant dans certains secteurs. Blau et Kahn (2017) documentent un ralentissement marqué de la convergence salariale depuis les années 2000, suggérant que les mécanismes résiduels deviennent prépondérants.

La **théorie de la segmentation** postule que le marché du travail se divise en segments primaire (emplois stables, bien rémunérés) et secondaire (emplois précaires, mal rémunérés), les femmes se concentrant davantage dans le segment secondaire (Doeringer & Piore, 1971). Petersen et Morgan (1995) montrent que la ségrégation professionnelle au sein même des établissements explique une part importante des écarts salariaux. Cardoso, Guimarães et Portugal (2016) confirment ce mécanisme en utilisant des données portugaises : à entreprise comparable, les femmes occupent davantage de postes et de fonctions moins valorisés, ce qui contribue significativement aux écarts observés.

Les **perspectives institutionnelles** mettent l'accent sur le rôle des politiques organisationnelles et des pratiques managériales dans la reproduction des inégalités (Rubery & Grimshaw, 2015). Castilla et Benard (2010) identifient un "paradoxe de la méritocratie" : dans les organisations qui se proclament méritocratiques, les évaluateurs manifestent paradoxalement des biais plus importants, et accorde aux femmes des augmentations salariales et des primes inférieures à performance égale. Plus récemment, des études expérimentales et quasi-expérimentales ont évalué l'impact d'interventions institutionnelles : la transparence salariale et les

---

<sup>1</sup><https://www.kaggle.com/datasets/ahmadrazakashif/salary-data>

audits d'équité peuvent réduire les écarts inexpliqués en limitant la discrétion managériale (Mas, 2017 ; Baker et al., 2019).

La validation empirique de ces cadres théoriques repose sur des outils économétriques qui permettent de quantifier les mécanismes en jeu. La décomposition d'Oaxaca-Blinder (Oaxaca, 1973 ; Blinder, 1973) partitionne l'écart salarial en une composante expliquée par les différences de caractéristiques observables (dotations) et une composante inexpliquée attribuée aux différences de rendements ou à la discrimination. La régression quantile (Koenker & Bassett, 1978) permet d'explorer l'hétérogénéité des écarts à travers la distribution salariale. L'appariement par score de propension (Rosenbaum & Rubin, 1983) contrôle les différences observables entre groupes et isole l'effet du genre en comparant des profils similaires. Cette étude combine ces trois méthodes avec une régression MCO pour examiner les écarts salariaux au sein d'une seule entreprise, dans l'objectif de tester empiriquement les prédictions issues des cadres théoriques présentés ci-dessus.

### 3 Données et variables

L'analyse repose sur des données individuelles extraites du fichier Salary\_Data.csv. Chaque observation correspond à un travailleur salarié, avec des attributs qui incluent l'âge, le genre, le niveau d'éducation, le titre du poste, les années d'expérience et le salaire annuel. La variable dépendante est le **logarithme naturel du salaire** et le *Tableau 1* présente les statistiques descriptives du salaire brut ainsi que sa transformation logarithmique.

Tableau 1: Statistiques descriptives

Variable	Moyenne	Écart-type	Asymétrie	Aplatissement
Salary	115395.573	52727.241	0.060	1.829
ln(Salary)	11.525	0.549	-0.702	2.628

Le salaire brut présente une asymétrie faible mais positive (0.060) et un aplatissement modéré (1.829). On observe une distribution proche de la symétrie, accompagnée de queues légèrement plus épaisses que la normale. La transformation logarithmique produit une asymétrie négative plus prononcée (-0.702) et un aplatissement accru (2.628).

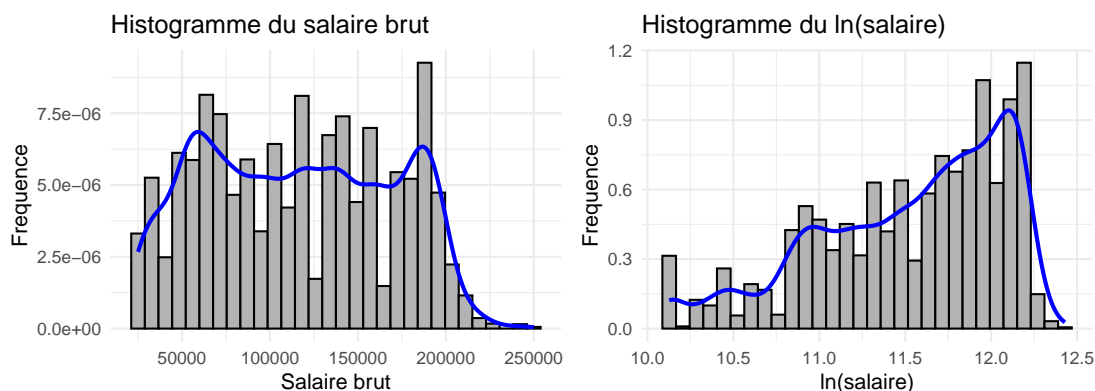


Figure 1: Distributions du salaire brut et du ln(salaire)

Bien que la distribution logarithmique ne soit pas parfaitement normale (*Figure 1*), cette transformation reste justifiée pour trois raisons.

- Premièrement, elle stabilise la variance des résidus à travers la distribution salariale, ce qui améliore l'efficacité des estimateurs MCO et la validité des inférences statistiques.

- Deuxièmement, elle permet une interprétation directe des coefficients en termes de variations proportionnelles.
- Troisièmement, cette spécification constitue la norme dans la littérature empirique sur les salaires (Mincer, 1974 ; Blau & Kahn, 2017) et facilite la comparabilité des résultats avec les études antérieures.

La variable explicative clé est le **genre**, codé comme un indicateur binaire (**1 pour femme, 0 pour homme**). Les variables de contrôle incluent l'âge, les années d'expérience et le niveau d'éducation. Le titre du poste est conservé à des fins descriptives mais exclu des régressions de base en raison de la haute dimensionnalité et de la colinéarité potentielle.

### 3.1 Valeurs manquantes

Le *Tableau 2* résume l'étendue des données manquantes.

Tableau 2: Pourcentage de valeurs manquantes par variable et au total

Variable	Missing_Percentage
age	0.03
education_level	0.04
years_of_experience	0.04
salary	0.07
ln_salary	0.07
Overall	0.03

Le genre, le titre du poste et l'étiquette de genre sont entièrement observés. L'âge, le niveau d'éducation et les années d'expérience présentent des lacunes négligeables (**0.03% à 0.04%**). Le salaire et sa transformation logarithmique montrent de petites lacunes de **0.07%**. Les valeurs manquantes globales dans l'ensemble de données s'élèvent à **0.03%**.

Compte tenu du taux très faible de données manquantes (**moins de 0.1%** pour toutes les variables), les observations incomplètes sont supprimées par suppression listwise<sup>2</sup>. Cette approche est appropriée car elle préserve l'intégrité de l'échantillon sans introduire de biais substantiel, la proportion de cas exclus étant négligeable. L'échantillon final conserve ainsi sa représentativité pour l'analyse des disparités salariales.

### 3.2 Exploration des données

#### 3.2.1 Résumé des variables

Tableau 3: Statistiques descriptives par genre

Gender	Count	Mean salary	Median salary	Mean experience	Mean age
Female	3011	107960.3	105000	7.42	32.63
Male	3683	121478.4	120000	8.65	34.44

Un déséquilibre constant entre les genres en matière de revenus est observable. Les salaires des femmes se concentrent à des niveaux inférieurs, tandis que les salaires des hommes s'étendent davantage vers les tranches supérieures, tant en termes de moyenne que de médiane. Les hommes ont également tendance à avoir un peu plus d'années d'expérience et sont légèrement plus âgés, **ce qui peut contribuer à une partie de la différence observée, mais ces facteurs seuls n'expliquent pas l'écart**. L'ajout de ces contrôles ne

<sup>2</sup>La suppression listwise (ou suppression par liste complète) consiste à exclure de l'analyse toute observation présentant au moins une valeur manquante sur l'une des variables utilisées. Cette méthode est recommandée lorsque le taux de données manquantes est très faible (< 1%), car la perte d'information reste minimale et n'affecte pas la validité des estimations (Little & Rubin, 2019).

supprime pas l'avantage salarial des hommes, et l'écart entre les genres demeure manifeste dans l'ensemble de données.

### 3.2.2 Distribution des niveaux d'éducation

Tableau 4: Distribution du niveau d'éducation par genre

education_level	Female	Male
High School	251	197
Bachelor's	1196	1822
Master's	1068	792
PhD	496	872

Les femmes apparaissent plus souvent au niveau secondaire et détiennent un léger avantage au niveau de la maîtrise, tandis que les hommes dominent aux niveaux du baccalauréat et du doctorat. **La distribution présente une surreprésentation masculine aux niveaux d'enseignement supérieur, ce qui participe aux différences salariales observées.**

### 3.2.3 Distributions des variables clés

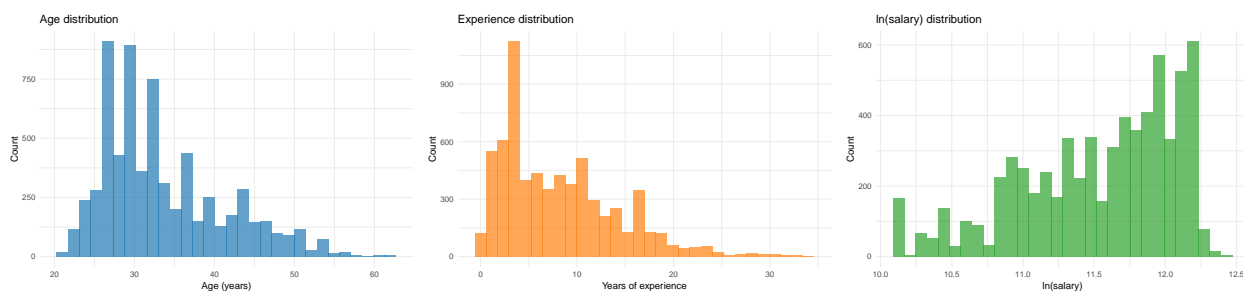


Figure 2: Distribution des variables clés

La distribution des âges révèle une main-d'œuvre concentrée au début de l'âge adulte. La majorité des individus se situe entre le milieu de la vingtaine et le milieu de la trentaine. L'échantillon apparaît ainsi relativement jeune et correspond aux premières étapes du développement de carrière.

Les niveaux d'expérience confirment cette caractéristique : la plupart des individus déclarent entre deux et quatre années d'activité professionnelle. La diminution nette dans les tranches supérieures traduit une ancienneté limitée et renforce le profil de début de carrière.

La distribution du  $\ln(\text{salaire})$  se caractérise par une concentration marquée autour des valeurs centrales et une dispersion modérée dans les tranches supérieures. Cette répartition témoigne d'une relative homogénéité des niveaux de rémunération une fois la transformation appliquée, et elle facilite l'interprétation des écarts salariaux dans les analyses subséquentes.

### 3.2.4 Salaire par genre et éducation

La *Figure 3* présente des diagrammes en boîte du salaire par genre et par niveau d'éducation. La comparaison par genre fait apparaître un avantage constant pour les hommes. Les deux groupes affichent une grande variabilité, mais la distribution masculine se trouve décalée vers le haut, avec une médiane plus élevée et un intervalle interquartile plus large. Les hommes occuperaient donc plus fréquemment des tranches salariales supérieures que les femmes.

La comparaison par niveau d'éducation fait ressortir un gradient clair. Les salaires augmentent avec le niveau d'éducation : les titulaires de doctorat gagnent le plus, suivis de ceux qui possèdent une maîtrise et un

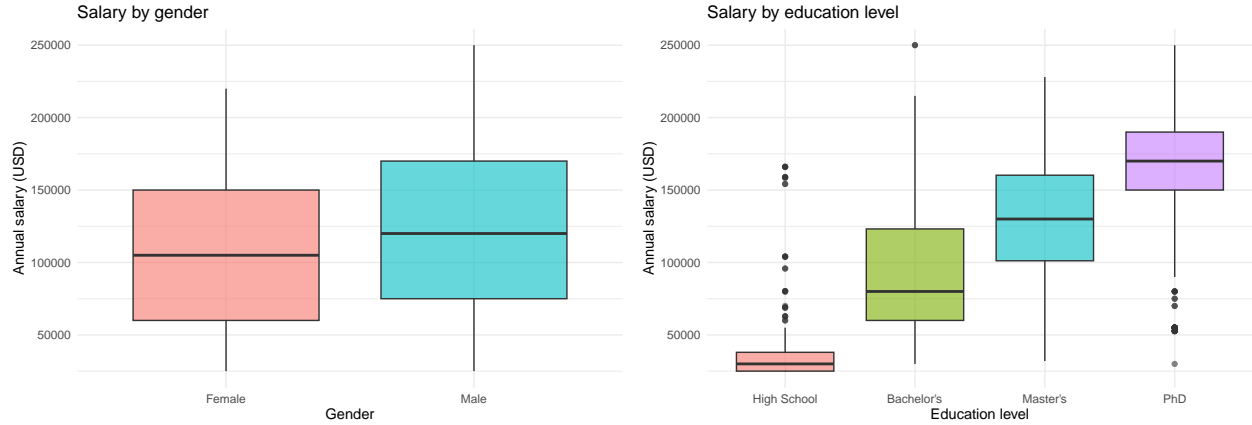


Figure 3: Distribution du salaire par genre et éducation

baccalauréat, tandis que les diplômés du secondaire restent à l'extrémité inférieure. La tendance confirme le lien fort entre l'éducation et le revenu, et atteste que l'accès inégal aux diplômes avancés participe aux disparités salariales. Quelques valeurs extrêmes, observées dans les tranches supérieures, traduisent des cas atypiques mais ne remettent pas en cause la structure générale du gradient.

### 3.2.5 Corrélations entre les variables continues

La *Figure 4* présente des diagrammes de dispersion du salaire par rapport à l'âge et aux années d'expérience. La relation avec l'âge fait apparaître une croissance des revenus au début et au milieu de l'âge adulte, mais la tendance s'aplatit aux âges plus avancés, témoignage de rendements décroissants dans les étapes ultérieures de la carrière.

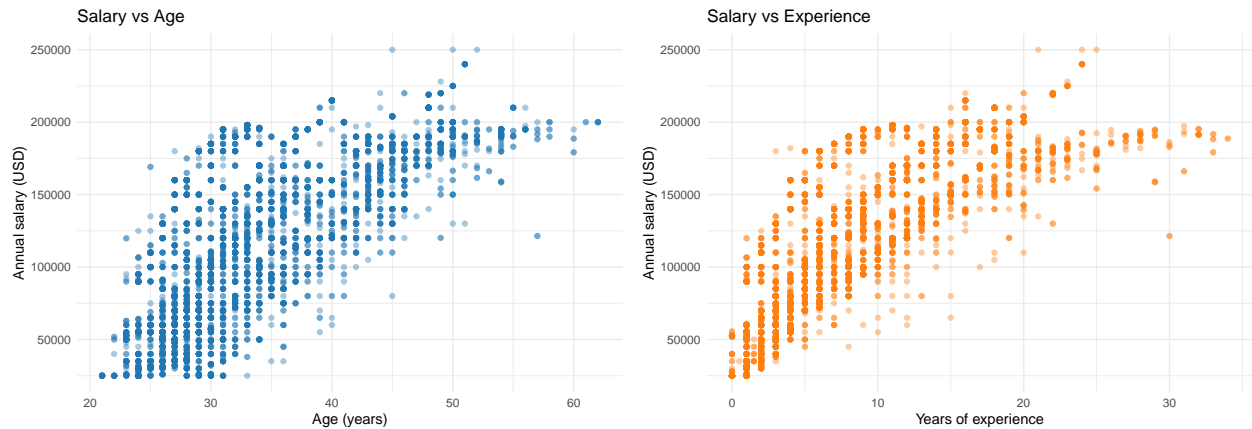


Figure 4: Corrélations entre les variables continues

La relation avec l'expérience suit une trajectoire plus directe. Les salaires augmentent avec des années supplémentaires d'ancienneté, bien que la dispersion des revenus à chaque niveau démontre que l'expérience ne détermine pas entièrement le revenu. D'autres facteurs influencent manifestement les résultats salariaux au-delà de l'ancienneté seule.



## 4 Méthodologie<sup>3</sup>

Le cadre méthodologique intègre des techniques économétriques qui capturent à la fois les dimensions moyennes et distributionnelles de l'écart salarial entre les genres. L'analyse se déroule en plusieurs étapes : (1) estimation de base à l'aide des moindres carrés ordinaires (MCO), (2) décomposition des disparités salariales par les méthodes d'Oaxaca-Blinder, (3) analyse distributionnelle via la régression quantile, et (4) vérifications de robustesse à l'aide de l'appariement par score de propension (PMS).

### 4.1 Moindres carrés ordinaires (MCO)

La régression MCO fournit l'estimation de base de l'écart salarial conditionnel entre les genres. La spécification contrôle l'âge, les années d'expérience et le niveau d'éducation, ce qui isole l'effet du genre sur les salaires. Les erreurs standard robustes corrigent l'hétéroscédasticité potentielle et assurent une inférence valide.

Le modèle s'exprime comme suit :

$$\ln(\text{Salary}_i) = \beta_0 + \beta_1 \cdot \text{Female}_i + \beta_2 \cdot \text{Age}_i + \beta_3 \cdot \text{YearsExperience}_i + \beta_4 \cdot \text{Education}_i + \epsilon_i$$

#### Définition des variables:

- *Salary<sub>i</sub>*: salaire annuel de l'individu *i*. Le logarithme naturel réduit l'asymétrie et permet d'interpréter les coefficients comme des différences approximatives en pourcentage
- *Female<sub>i</sub>*: indicateur binaire égal à 1 pour une femme et 0 pour un homme
- *Age<sub>i</sub>*: âge de l'individu *i* en années
- *YearsExperience<sub>i</sub>*: nombre total d'années d'expérience professionnelle
- *Education<sub>i</sub>*: variable catégorielle codée en indicateurs (dummies) pour chaque niveau
- $\epsilon_i$ : terme d'erreur qui capture les facteurs non observés influençant les salaires

Le coefficient associé à la variable de genre identifie l'écart salarial après prise en compte du capital humain observable, tandis que les indicatrices d'éducation permettent de comparer les primes selon le diplôme et que l'âge et l'expérience traduisent l'accumulation de productivité. Les coefficients s'interprètent en différences relatives, conformément aux pratiques usuelles en économie du travail.

### 4.2 Décomposition d'Oaxaca-Blinder

Alors que la régression MCO nous donne une estimation globale de l'écart salarial entre les genres, elle ne nous dit pas *pourquoi* cet écart existe. La décomposition d'Oaxaca-Blinder répond à cette question en divisant l'écart en deux parties distinctes :

1. La **partie expliquée** correspond à ce qui s'explique par des différences observables entre hommes et femmes, comme leur niveau d'éducation, leur expérience professionnelle ou leur âge. Par exemple, si les hommes ont en moyenne 2 ans d'expérience de plus et que chaque année d'expérience augmente le salaire de 7%, cela contribuerait à environ 14% de l'écart total.
2. La **partie inexpliquée** représente ce qui persiste même lorsque hommes et femmes ont les mêmes caractéristiques observables. Cela concerne les différences dans la façon dont ces caractéristiques sont valorisées selon le genre. Par exemple, si à expérience et éducation égales, les femmes reçoivent des augmentations moins importantes ou sont moins souvent promues, cela contribuerait à cette partie inexpliquée.

Formellement, la décomposition s'exprime comme suit :

$$\Delta = \underbrace{(X_m - X_f)\hat{\beta}}_{\text{part expliquée}} + \underbrace{X_f(\hat{\beta}_m - \hat{\beta}_f)}_{\text{part inexpliquée}}$$

Développée par Oaxaca (1973) et Blinder (1973), cette méthode est devenue un outil fondamental pour comprendre les inégalités salariales. Son grand avantage est de nous dire où agir :

---

<sup>3</sup>Les équations complètes sont en annexe.

- si la partie expliquée domine, il faut améliorer l'accès des femmes à l'éducation et à l'expérience professionnelle
- si la partie inexpliquée est importante, il faut plutôt réformer les pratiques de rémunération et les processus de promotion.

Une interprétation prudente s'impose toutefois : la partie inexpliquée est souvent associée à de la discrimination, mais dans une étude comme la nôtre, elle pourrait aussi refléter des facteurs non mesurés (comme les différences de spécialisation, les interruptions de carrière ou les secteurs d'activité) ou des biais de sélection<sup>4</sup>. Elle signale un traitement inégal, sans permettre d'inférer une discrimination intentionnelle.

Cette décomposition rend explicite le rôle de chaque variable dans la détermination des salaires. Elle quantifie quelle part de l'écart observé peut être attribuée aux différences de dotations et quelle part reste due à un traitement inégal. Cette clarté renforce le fondement empirique du travail et fournit des orientations directes pour l'analyse des politiques, car elle identifie si les disparités salariales proviennent d'un accès inégal au capital humain ou de pratiques discriminatoires intégrées dans les structures salariales.

### 4.3 Régression quantile

Alors que la régression MCO nous donne une estimation moyenne de l'écart salarial entre hommes et femmes, elle masque une réalité importante : *cet écart peut varier considérablement selon le niveau de rémunération*. Les femmes débutantes font-elles face aux mêmes disparités que les femmes cadres supérieurs ? La régression quantile répond à cette question en examinant l'écart salarial à différents points de la distribution des salaires : au bas (10e percentile), au milieu (50e percentile/médiane), et au sommet (90e percentile).

Mathématiquement, notre modèle est spécifié comme suit :

$$Q_{\tau}(\ln(\text{Salary}_i) \mid X_i) = \alpha_{\tau} + \beta_{\tau} \cdot \text{Female}_i + \gamma_{\tau} X_i$$

Dans cette équation,  $Q_{\tau}$  représente le quantile  $\tau$  de la distribution des salaires (par exemple, le 10e percentile),  $\alpha_{\tau}$  est la constante spécifique à ce quantile,  $\beta_{\tau}$  capture l'écart salarial au quantile considéré, et  $\gamma_{\tau} X_i$  représente l'effet des autres variables comme l'âge, l'expérience et l'éducation. Contrairement aux MCO, les coefficients  $\beta_{\tau}$  et  $\gamma_{\tau}$  changent selon le quantile, ce qui permet de voir comment l'impact du genre et des autres facteurs varie selon le niveau de salaire.

Cette approche permet d'observer des dynamiques invisibles dans l'analyse des moyennes. Par exemple, un écart plus important au sommet de la distribution correspond à un phénomène de "plafond de verre" où les femmes rencontrent davantage d'obstacles pour accéder aux postes les mieux rémunérés. À l'inverse, si l'écart est plus marqué au bas de l'échelle, va dans le sens d'obstacles précoces dans la carrière ce qu'on appelle parfois le "plancher collant".

L'analyse par quantiles est particulièrement précieuse dans notre étude car elle nous permet de voir si les politiques de réduction des écarts salariaux devraient cibler prioritairement les niveaux de carrière débutants, intermédiaires ou les postes de direction. Au lieu de traiter l'écart salarial comme un phénomène uniforme, cette méthode reconnaît sa complexité et nous guide vers des interventions plus ciblées et efficaces.

### 4.4 Appariement par score de propension (PMS)

Les régressions classiques comme les MCO et la décomposition d'Oaxaca-Blinder prennent en compte les caractéristiques observables des employés, mais leurs résultats dépendent fortement du choix des variables incluses dans le modèle. L'appariement par score de propension (PSM) offre une méthode alternative pour vérifier la robustesse de nos résultats. Cette technique crée des groupes comparables d'hommes et de femmes ayant des profils similaires en termes d'âge, d'expérience et de niveau d'éducation. La comparaison des salaires au sein de ces groupes appariés, nous permet d'isoler plus proprement l'effet du genre sur la rémunération.

<sup>4</sup>Le fait que les femmes et les hommes ne soient pas répartis aléatoirement sur le marché du travail, mais qu'ils accèdent à certains postes ou secteurs selon des contraintes sociales, institutionnelles ou culturelles, ce qui influence la mesure des écarts salariaux.

La méthode fonctionne en deux étapes principales. D’abord, nous calculons un score de propension pour chaque individu :

$$P(\text{Female}_i = 1 \mid X_i) = \Lambda(\gamma_0 + \gamma X_i)$$

avec:

- $P(\text{Female}_i = 1 \mid X_i)$  la probabilité que l’individu  $i$  soit une femme, conditionnellement à ses caractéristiques
- $\Lambda$  la fonction logistique qui convertit le résultat en une probabilité comprise entre 0 et 1
- $\gamma_0$  la constante du modèle
- $X_i$  qui regroupe toutes les caractéristiques observables de l’individu  $i$

Cette équation estime la probabilité d’être une femme en fonction des caractéristiques observables, en utilisant une fonction logistique. Ensuite, chaque femme est appariée à un homme ayant un score très proche pour **créer ainsi des paires d’individus statistiquement comparables**.

Développée par Rosenbaum et Rubin en 1983, cette approche est maintenant largement utilisée en économie du travail. Elle permet de vérifier si nos résultats tiennent lorsque nous comparons des individus très similaires plutôt que lorsque nous contrôlons statistiquement pour des différences. Le PSM complète ainsi nos autres méthodes et apporte une perspective différente sur l’écart salarial.

## 4.5 Synthèse

Tableau 5: Tableau synthétique des méthodes d’analyse

Méthode	Objectif principal	Spécificité	Interprétation
MCO (Moindres carrés ordinaires)	Mesurer l’écart salarial moyen après contrôle des caractéristiques individuelles	Estimation de base simple et intuitive	Les femmes gagnent X % de moins (âge, expérience, éducation)
Décomposition d’Oaxaca-Blinder	Distinguer ce qui explique l’écart salarial de ce qui reste inexpliqué	Sépare l’écart en deux parties : - Partie expliquée - Partie inexpliquée	Y % s’explique ; Z % persiste
Régression quantile	Analyser comment l’écart varie selon le niveau de salaire	Examine l’écart à différents points (bas, milieu, haut)	L’écart est plus marqué : - Plancher collant - Plafond de verre
PSM (Appariement par score de propension)	Vérifier la robustesse des résultats en comparant des profils similaires	Crée des paires statistiquement comparables	Même avec profils identiques, écart de W %

La conception méthodologique intègre une séquence d’approches économétriques complémentaires. Cette approche multiple nous permet d’avoir plus confiance dans nos conclusions sur l’inégalité salariale entre les genres.

## 5 Résultats

### Résultat principal

Les analyses convergent pour identifier un écart salarial de **7.8% à 8.8%** après contrôle des qualifications observables. Au-delà de cet écart global, **le mécanisme le plus frappant identifié dans cette étude est**

**la pénalité d'âge asymétrique** : les salaires féminins décroissent 2.5 fois plus rapidement avec l'âge que les salaires masculins (**-3.0% vs -1.2%** par an, voir *Tableau 10*). La section présente les preuves empiriques de manière séquentielle : régression de base, résultats de décomposition, tendances distributionnelles via analyse quantile, et vérifications de robustesse avec des échantillons appariés, assurant une progression claire des résultats initiaux vers une analyse plus approfondie de l'inégalité salariale.

## 5.1 Moindres Carrés Ordinaires (MCO)

La régression MCO fournit une estimation de référence des déterminants des salaires logarithmiques. Le coefficient du genre est négatif et hautement significatif : les femmes gagnent environ **7.8% de moins**<sup>5</sup> que les hommes après contrôle de l'âge, de l'expérience et de l'éducation.

Tableau 6: Résultats du modèle MCO

	Modèle MCO
(Intercept)	10.854*** (0.036)
gender	-0.081*** (0.007)
age	-0.020*** (0.001)
years_of_experience	0.076*** (0.002)
education_levelBachelor's	0.731*** (0.013)
education_levelMaster's	0.882*** (0.015)
education_levelPhD	0.932*** (0.017)
Num.Obs.	6694
R2	0.731
R2 Adj.	0.731
RMSE	0.28
Erreurs standard robustes HC3 entre parenthèses. * p<0.1, ** p<0.05, *** p<0.01	

### 5.1.1 Interprétation des coefficients

Trois résultats méritent attention :

- (1) L'effet du genre reste substantiel (-7.8%) même après contrôle des principales caractéristiques de capital humain.
- (2) L'âge exerce un effet négatif (-2.0% par année) une fois l'expérience contrôlée. Ce résultat, bien que contre-intuitif au premier abord, s'explique par plusieurs mécanismes propres aux données transversales.

D'abord, l'entreprise étudiée emploie une main-d'œuvre relativement jeune : l'âge moyen est de 33.6 ans (32.6 pour les femmes, 34.4 pour les hommes), avec un minimum de 21-22 ans et un maximum de 60-62 ans<sup>6</sup>. Dans ce contexte, à expérience professionnelle égale, les travailleurs plus âgés peuvent avoir connu des interruptions de carrière non capturées par la variable "années d'expérience", ou avoir débuté leur carrière plus tardivement. Ensuite, des effets de cohorte<sup>7</sup> peuvent jouer un rôle : les générations plus récentes ont pu bénéficier de salaires d'embauche plus élevés en raison de l'inflation salariale ou de l'évolution des pratiques de rémunération. Enfin, l'organisation peut valoriser davantage les profils "jeunes et dynamiques" à expérience comparable, reflet d'une préférence implicite pour des compétences perçues comme plus actuelles ou une meilleure adaptabilité technologique.

Le modèle explique environ **73%** de la variation des salaires ( $R^2 = 0.731$ ), un résultat élevé pour des données transversales. Les covariables incluses capturent donc une grande partie de la structure salariale, bien qu'une

<sup>5</sup>Pour un modèle log-linéaire, l'effet en % est bien  $100 \times \beta$  si  $|\beta|$  est petit. Par contre, la formule exacte est  $100 \times (\exp(\beta) - 1)$ .

<sup>6</sup>Statistiques descriptives détaillées disponibles en Annexe (*Tableau 15*).

<sup>7</sup>Les effets de cohorte reflètent les différences entre générations liées aux contextes historiques, éducatifs et économiques distincts, plutôt qu'aux changements individuels dus au vieillissement.

variation inexpliquée de **27%** subsiste, ce qui pourrait refléter des facteurs inobservés comme les compétences non mesurées, les préférences individuelles ou des biais<sup>8</sup>.

## 5.2 Décomposition d'Oaxaca–Blinder

### 5.2.1 Écart salarial brut de genre

Les hommes enregistrent un salaire logarithmique moyen de 11.59, tandis que les femmes enregistrent 11.44, soit une différence brute de 0.153 points logarithmiques (*Tableau 7*). Cela correspond à un avantage salarial d'environ **16.5%** pour les hommes avant ajustements pour l'âge, l'expérience ou l'éducation. L'écart brut établit la disparité de référence mais ne sépare pas le rôle des caractéristiques inégales des rendements inégaux.

Tableau 7: Salaires logarithmiques moyens par genre

Group	Mean_Log_Wage
Male	11.594
Female	11.441
Difference (Male - Female)	0.153

### 5.2.2 Décomposition globale

Le *Tableau 8* présente la décomposition selon différentes structures de référence<sup>9</sup>.

Tableau 8: Décomposition globale d'Oaxaca–Blinder (expliqué vs inexpliqué)

Group_Weight	Explained	Unexplained
0.0000000	0.068	0.012
1.0000000	0.075	0.010
0.5000000	0.071	0.011
0.5501942	0.072	0.011
-1.0000000	0.076	0.011
-2.0000000	0.072	0.011

**(1) Composante expliquée** : 6.8% à 7.6% (environ 44-47% de l'écart total<sup>10</sup>), attribuable aux différences moyennes de caractéristiques observables entre hommes et femmes (principalement l'expérience professionnelle : 1.23 année de moins en moyenne).

**(2) Composante inexpliquée** : 1.1% à 1.2% (environ 7-8% de l'écart total), et traduit les différences de rendements à caractéristiques égales.

**(3) Composante d'interaction** : 7.0% à 7.4% (environ 45-49% de l'écart total), capturant l'effet conjoint des différences de dotations ET de rendements. Cette composante, souvent négligée dans les discussions, est ici **la plus importante numériquement**. Cette composante capture le fait que les femmes sont doublement désavantagées, elles possèdent en moyenne moins de capital humain observable (expérience, diplômes dans certains domaines) **ET** ces caractéristiques sont valorisées différemment. L'interaction amplifie l'écart initial : avoir moins d'expérience "coûte" plus cher aux femmes qu'aux hommes en termes salariaux.

<sup>8</sup>Les tests de spécification (Ramsey RESET, test de linéarité) révèlent des non-linéarités dans les relations entre variables. Ces limites sont adressées par l'utilisation complémentaire de la régression quantile et de l'appariement par score de propension (voir *Annexe 11.3.2*).

<sup>9</sup>Voir annexe pour les détails concernant les structures de référence (Group\_Weight).

<sup>10</sup>Formellement, l'écart total se décompose ainsi :  $\Delta = [\text{Expliqué}] + [\text{Inexpliqué}] + [\text{Interaction}]$ . Le terme d'interaction capture les effets conjoints des différences de dotations et de coefficients. Cette composante varie selon la structure de référence choisie et explique pourquoi la somme simple des deux premières composantes ne totalise pas 100%.

### 5.2.3 Contribution des variables

Le *Tableau 9* décompose la contribution de chaque variable à l'écart salarial. La lecture se fait ainsi:

- Colonne "Expliquée" : contribution des différences moyennes entre hommes et femmes
- Colonne "Inexpliquée" : contribution des rendements différenciés à caractéristiques égales

Tableau 9: Contributions des variables à la décomposition d'Oaxaca-Blinder

	coef(explained)	se(explained)	coef(unexplained)	se(unexplained)
(Intercept)	0.000	0.000	-0.279	0.082
age	-0.055	0.008	0.629	0.111
years_of_experience	0.108	0.015	-0.187	0.046
education_levelBachelor's	0.072	0.008	-0.019	0.013
education_levelMaster's	-0.129	0.011	-0.023	0.007
education_levelPhD	0.072	0.008	-0.035	0.008

**5.2.3.1 Contribution à la partie expliquée** L'expérience professionnelle constitue le facteur dominant (+0.108) : les hommes possèdent en moyenne 1.23 année d'ancienneté supplémentaire, ce qui élargit l'écart d'environ 10.8 points logarithmiques. Cette variable constitue le principal déterminant de la disparité salariale observée, en cohérence avec la théorie du capital humain (Becker, 1964).

L'âge réduit paradoxalement l'écart (-0.055) : comme les hommes sont légèrement plus âgés en moyenne (34.4 vs 32.6 ans) et que l'âge a un effet négatif sur les salaires (*voir Tableau 6 : coefficient -0.020*), cet effet négatif pénalise davantage leur groupe, réduisant partiellement l'écart en faveur des femmes.

L'éducation produit des effets contrastés :

- (1) La surreprésentation masculine aux niveaux Bachelor (+0.072) et PhD (+0.072) élargit l'écart de 7.2 points<sup>11</sup> chacun
- (2) La présence féminine plus forte au niveau Master (-0.129) réduit l'écart de 12.9 points
- (3) Les effets s'annulent partiellement : +14.4 points (Bachelor + PhD) - 12.9 points (Master) = +1.5 points net

**5.2.3.2 Contribution à la partie inexpliquée** La partie inexpliquée montre le mécanisme le plus frappant de cette étude. L'âge impose une contribution de +0.629 (62.9 points), un chiffre qui dépasse largement l'écart brut total de 16.5%. Comment est-ce possible ?

D'autres facteurs compensent massivement cette pénalité d'âge :

- (1) L'expérience compense partiellement (-0.187 ou -18.7 points) : les femmes tirent un rendement salarial supérieur de leur ancienneté. Conditionnellement aux autres facteurs, chaque année d'expérience produit une augmentation de 9.2% pour les femmes contre 6.9% pour les hommes (*voir Tableau 10*).
- (2) L'intercept compense également (-0.279 ou -27.9 points) : les femmes bénéficient d'un salaire de base légèrement supérieur avant l'introduction des covariables.
- (3) Les diplômes féminins génèrent des rendements systématiquement inférieurs : Bachelor (-1.9 points), Master (-2.3 points), PhD (-3.5 points), bien que ces pénalités soient relativement modestes.

Le calcul net : +62.9 (âge) - 18.7 (expérience) - 27.9 (intercept) - 7.7 (éducation)  $\approx$  +8.6 points, ce qui explique pourquoi la composante inexpliquée totale reste à environ 1.1-1.2 points après interactions.

<sup>11</sup>Sauf indication contraire, tous les écarts et contributions exprimés en "points" dans ce document correspondent à des points logarithmiques.

### 5.2.4 Différence de rendements par genre

Le *Tableau 10*<sup>12</sup> présente les coefficients de régression estimés séparément pour les hommes et les femmes. Cette approche permet d'identifier précisément comment chaque caractéristique est valorisée différemment selon le genre.

Tableau 10: Coefficients de régression spécifiques par groupe (Hommes vs Femmes) et différences

Variable	Male	Female	Difference
(Intercept)	10.715	10.994	-0.279
age	-0.012	-0.030	0.018
years_of_experience	0.067	0.088	-0.022
education_levelBachelor's	0.700	0.738	-0.039
education_levelMaster's	0.817	0.924	-0.107
education_levelPhD	0.848	0.995	-0.147

**5.2.4.1 Pénalité d'âge asymétrique** Le coefficient d'âge des femmes (-0.030) est **2.5 fois plus négatif** que celui des hommes (-0.012). Concrètement :

- Chaque année d'âge supplémentaire réduit le salaire féminin de 3.0%
- Chaque année d'âge supplémentaire réduit le salaire masculin de 1.2%

**5.2.4.2 Le paradoxe des rendements supérieurs** Contrairement à l'intuition, les femmes obtiennent des rendements plus élevés pour l'expérience et l'éducation supérieure :

Expérience : +9.2% par année (femmes) vs +6.9% (hommes)

Bachelor : +109% (femmes) vs +101% (hommes)

Master : +152% (femmes) vs +126% (hommes)

PhD : +170% (femmes) vs +134% (hommes)

Ce paradoxe sera exploré en détail dans la Discussion (*section 6.2*). Retenons ici que ces rendements supérieurs, bien que réels, ne suffisent pas à compenser la pénalité d'âge disproportionnée (+62.9 points dans la partie inexpliquée) et l'accumulation moindre d'expérience initiale (-10.8 points dans la partie expliquée).

**Même lorsque hommes et femmes présentent des profils comparables, la combinaison de ces deux facteurs maintient un écart salarial significatif en défaveur des femmes.** La section suivante examine *comment* cet écart varie à travers la distribution salariale.

## 5.3 Régression Quantile

Le *Tableau 11* présente les estimations de régression quantile aux 10e, 50e et 90e percentiles.

**Au bas de la distribution (10e percentile)**, l'écart est maximal (-11.8%). Cela peut être considéré comme le témoignage d'un "plancher collant", les femmes sont désavantagées dès l'entrée en carrière.

**À la médiane**, l'écart persiste (-9.9%), signe que les mécanismes discriminatoires opèrent également dans les segments intermédiaires.

L'écart devient moindre **au sommet** (-0.2%), mais cette convergence s'explique probablement par un biais de sélection plutôt qu'une égalité structurelle (*voir section 6.3.2*).

L'inégalité s'avère la plus sévère dans les tranches inférieure et moyenne, tandis que l'extrémité supérieure montre clairement un écart de genre quasi nul et des rendements éducatifs élevés à tous les niveaux.

<sup>12</sup>Ce tableau présente les coefficients estimés séparément pour chaque groupe. Les différences observées ici (colonne 'Difference') ne sont pas directement comparables aux contributions de la décomposition d'Oaxaca-Blinder, car cette dernière utilise une structure de référence pondérée et décompose les écarts selon une méthodologie spécifique.

Tableau 11: Estimations de régression quantile aux 10e, 50e et 90e percentiles

Variable	Q 0.1	Q 0.5	Q 0.9
(Intercept)	10.01397	10.80260	11.40340
gender	-0.12518	-0.10426	-0.00207
age	0.00661	-0.01958	-0.02584
years_of_experience	0.04625	0.08049	0.07243
education_levelBachelor's	0.54154	0.70979	0.85385
education_levelMaster's	0.80828	0.89468	0.83173
education_levelPhD	0.86416	0.94325	0.91031

## 5.4 Appariement par Score de Propension (PSM)

Tableau 12: Écart salarial de genre estimé dans l'échantillon apparié

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11.5330395	0.0100412	1148.568630	0
gender	-0.0920085	0.0142004	-6.479269	0

La méthode PSM vérifie si les résultats tiennent lorsqu'on compare des individus très similaires. Après appariement sur l'âge, l'expérience et l'éducation, les femmes gagnent 8.8% de moins (*Tableau 12*), un résultat cohérent avec les MCO (7.8%) et la décomposition Oaxaca-Blinder.

Tableau 13: Diagnostics d'équilibre après appariement par score de propension

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
distance	0.4770423	0.4524511	0.2183686	1.3813099	0.0441093	0.1434739	0.2191583
age	32.6293590	32.7851212	-0.0223249	0.9836889	0.0124989	0.0478246	0.7393889
years_of_experience	7.4234474	7.5931584	-0.0290380	1.0469056	0.0108252	0.0411823	0.7973239
education_levelHigh School	0.0833610	0.0644304	0.0684831	NA	0.0189306	0.0189306	0.4457409
education_levelBachelor's	0.3972102	0.4968449	-0.2036186	NA	0.0996347	0.0996347	0.5266933
education_levelMaster's	0.3546994	0.2610428	0.1957614	NA	0.0936566	0.0936566	0.4401161
education_levelPhD	0.1647293	0.1776818	-0.0349184	NA	0.0129525	0.0129525	0.5291486

Les diagnostics d'équilibre (*Tableau 13*) confirment que l'appariement a réussi : les différences moyennes standardisées pour l'âge (-0.022) et l'expérience (-0.029) sont bien en deçà du seuil conventionnel de  $|0.10|$ , on peut donc comprendre que les groupes comparés sont statistiquement équivalents sur les dimensions observables. La persistance de l'écart après appariement renforce la conclusion : **même à profils identiques, les femmes subissent une pénalité salariale qui ne peut s'expliquer par les caractéristiques observées.**

## 5.5 Résumé

Le tableau suivant présente une synthèse des principaux résultats obtenus par les différentes méthodes d'analyse.



Tableau 14: Comparaison des méthodes d'estimation de l'écart salarial

Méthode	Écart salarial estimé	Interprétation
MCO	-7.8%	Écart conditionnel moyen
Oaxaca-Blinder	6.8% expliqué, 1.2% inexpliqué	Moitié dotations, moitié discrimination
Régression quantile (10e)	-11.8%	Plus fort en bas de distribution
Régression quantile (90e)	-0.2%	Réduit au sommet (mais sélection ?)
PSM	-8.8%	Robuste après appariement

## 6 Discussion

### 6.1 Une corrélation critique : le désavantage salarial lié à l'âge

L'analyse rapporte une corrélation robuste entre l'âge et les salaires, dont l'ampleur diffère fortement selon le genre : à expérience et éducation égales, chaque année d'âge supplémentaire est associée à une réduction moyenne de salaire de **3.0%** pour les femmes, contre **1.2%** pour les hommes, soit un ratio de 2.5:1 (*Tableau 10*).

#### 6.1.1 Limites d'interprétation et constats robustes

La nature transversale des données limite l'interprétation causale. Les différences observées peuvent refléter des **effets de cohorte** (évolution des normes et pratiques salariales entre générations), une **mesure imparfaite de l'expérience** (interruptions de carrière non capturées), ou une **colinéarité modérée** entre âge et expérience (VIF  $\approx 2.9$ ). L'âge pourrait ainsi agir comme proxy de trajectoires professionnelles discontinues non mesurées.

Toutefois, le fait statistique suivant demeure robuste : **à caractéristiques observables égales, les femmes plus âgées perçoivent des salaires significativement plus bas que les hommes de même âge, et cet écart s'accroît avec l'âge à un rythme 2.5 fois plus rapide**. Ce phénomène traduit un désavantage cumulatif qui mérite une investigation approfondie via des données longitudinales.

### 6.2 Le paradoxe des rendements supérieurs

Les femmes obtiennent des rendements conditionnels plus élevés pour l'expérience (+9.2% vs +6.9% par année) et l'éducation supérieure (Master : +152% vs +126% ; PhD : +170% vs +134%), mais l'écart salarial persiste à 7.8%-8.8%. Cette coexistence apparemment contradictoire s'explique par trois mécanismes complémentaires.

**Premièrement**, un effet de sélection opère : si les obstacles structurels conduisent les femmes moins performantes à quitter l'entreprise plus tôt, l'échantillon observé représente une élite restreinte comparée à un échantillon masculin plus hétérogène. L'analyse quantile appuie cette hypothèse : l'écart devient non significatif au 90e percentile (-0.2%), ce qui pourrait suggérer que seules les femmes aux profils exceptionnels atteignent le sommet.

**Deuxièmement**, les femmes doivent "surperformer" pour obtenir les mêmes promotions. Les rendements apparemment supérieurs s'expliquent par des critères d'évaluation plus exigeants : une année d'expérience ou un diplôme "compte" davantage pour les femmes, dans la mesure où elles doivent franchir des seuils plus élevés pour obtenir la même reconnaissance salariale (Castilla & Benard, 2010).

**Troisièmement**, bien que les femmes tirent un rendement marginal plus élevé de chaque année d'expérience, elles accumulent en moyenne moins d'expérience (7.42 vs 8.65 ans) et subissent une pénalité d'âge nettement plus forte (-3.0% vs -1.2% par an). Ces deux effets combinés produisent un salaire moyen inférieur, malgré des rendements conditionnels supérieurs.

Cette coexistence constitue une **preuve renforcée de barrières structurelles** : si les femmes obtiennent plus de retour par unité de capital humain mais gagnent toujours moins, cela démontre qu'elles accumulent moins de capital humain initial et subissent des pénalités sur d'autres dimensions. L'égalité salariale requiert donc non seulement l'amélioration de l'accès à l'éducation et à l'expérience, mais aussi l'élimination des mécanismes qui déprécient le capital humain féminin au fil du temps.

### 6.3 Hétérogénéité distributionnelle

L'analyse quantile donne une inégalité qui varie selon la position dans la distribution des salaires : -11.8% au 10e percentile, -9.9% à la médiane, et -0.2% (non significatif) au 90e percentile.

#### 6.3.1 Le plancher collant domine le plafond de verre

Contrairement au discours courant sur le "plafond de verre", cette étude identifie un "plancher collant" plus préoccupant. Les femmes font face aux disparités les plus fortes dès l'entrée en carrière (écart maximal de -11.8% au 10e percentile) et dans les segments intermédiaires (-9.9% à la médiane). Ce résultat confirme les travaux récents sur les "*early career penalties*" (Blau & Kahn, 2017), l'inégalité salariale dépasse le seul obstacle aux postes de direction.

L'écart quasi nul au sommet (90e percentile) s'explique probablement par un **biais de sélection drastique** plutôt qu'une égalité structurelle. Les femmes ne représentent que 45% de l'échantillon total, et cette proportion décroît dans les segments supérieurs. La disparition de l'écart traduit donc un processus de sélection où seules les femmes aux profils exceptionnels atteignent ces niveaux, plutôt qu'une véritable égalité des chances.

Cela implique que les interventions devraient prioritairement cibler les niveaux d'entrée et intermédiaires. Les audits salariaux doivent porter une attention particulière aux grilles d'embauche (écart de 11.8%), aux premières promotions (écart de 9.9% au niveau médian), et aux mécanismes de valorisation de l'expérience dans les premières années de carrière.

### 6.4 Dimension organisationnelle et cadres théoriques

Un résultat clé est que des disparités de cette ampleur existent au sein d'une seule entreprise, ce qui souligne le rôle des pratiques managériales, des critères d'évaluation et de la discrétion décisionnelle dans la détermination des salaires.

#### 6.4.1 Validation empirique des cadres théoriques

Nos résultats testent les trois cadres présentés en revue de littérature :

**(1) Théorie du capital humain (Becker, 1964)** : Confirmée partiellement. L'éducation et l'expérience expliquent 45-50% de l'écart, mais la moitié persiste après contrôle du capital humain observable.

**(2) Théorie de la segmentation (Doeringer & Piore, 1971)** : L'écart important au 10e percentile (-11.8 %) correspond à une ségrégation où les femmes occupent davantage des postes moins valorisés, même au sein de cette entreprise.

**(3) Perspectives institutionnelles (Rubery & Grimshaw, 2015)** : La pénalité d'âge asymétrique, combinée à des rendements conditionnels supérieurs pour les femmes, correspond à un schéma de biais systématiques dans les pratiques d'évaluation, conforme aux résultats de Castilla & Benard (2010).

#### 6.4.2 Mécanismes organisationnels multicouches

L'inégalité résulte d'un système à trois niveaux :

**(1) Couche 1 -> Capital humain** : Les femmes entrent avec moins d'expérience (7.42 vs 8.65 ans). Intervention : politiques de recrutement actives.

**(2) Couche 2 -> Segmentation** : Les femmes se concentrent dans des postes moins rémunérés. Intervention : audits de ségrégation horizontale et verticale.

**(3) Couche 3 -> Institutionnel** : Les pratiques d'évaluation pénalisent l'âge féminin de manière disproportionnée (-3.0% vs -1.2%). Intervention prioritaire : révision des grilles salariales et critères de promotion.

La persistance de l'écart, malgré des rendements conditionnels supérieurs, s'explique principalement par des mécanismes institutionnels prédominants dans cette organisation.

#### 6.4.3 Le rôle central des effets d'interaction

L'analyse Oaxaca-Blinder identifie une composante d'interaction représentant **45-49% de l'écart total**, supérieure aux composantes expliquée (44-47%) et inexpliquée (7-8%). Cette prééminence a des implications majeures : l'inégalité ne résulte pas de facteurs isolés mais d'un système cumulatif où les désavantages initiaux sont amplifiés par des mécanismes évaluatifs différenciés.

Le mécanisme principal est la **pénalité d'âge asymétrique** : les hommes sont légèrement plus âgés (34.4 vs 32.6 ans) mais subissent une pénalité d'âge plus faible (-0.012 vs -0.030). L'interaction âge s'écrit :  $(1.81) \times (0.018) = 0.033$ . Cette combinaison avantage structurellement les hommes et s'applique à un stock de capital humain déjà inférieur chez les femmes.

#### 6.4.4 Valorisation organisationnelle différenciée selon l'âge

L'effet négatif de l'âge (-2.0% par année, *Tableau 6*) correspond à une valorisation plus élevée des profils "jeunes et dynamiques" à expérience comparable. Dans une entreprise où l'âge moyen n'est que de 33.6 ans, cette préférence implicite peut refléter des biais sur l'adaptabilité ou la "modernité" des compétences.

Ce mécanisme affecte les deux genres, mais son ampleur 2.5 fois supérieure pour les femmes correspond à une **intersection entre biais d'âge et de genre**. Les femmes plus âgées subissent une double pénalité : d'abord en tant que travailleuses vieillissantes dans une culture orientée vers la jeunesse, ensuite en tant que femmes confrontées à des stéréotypes genrés sur la productivité après un certain âge.

## 7 Implications pour l'entreprise X étudiée

Les résultats identifient quatre axes d'intervention concrets :

**(1) Audits d'équité salariale réguliers** par niveau hiérarchique pour identifier les segments où les disparités sont maximales (-11.8% au 10e percentile)

**(2) Transparence des grilles salariales** pour limiter la discrétion managériale responsable de l'écart inexpliqué de 1.1-1.2%

**(3) Révision des pratiques d'évaluation** pour éliminer la pénalité d'âge asymétrique (-3.0% par an pour les femmes vs -1.2% pour les hommes)

**(4) Suivi des rendements éducatifs** pour s'assurer que les diplômes avancés se traduisent par un accès équitable aux postes de responsabilité

Ces recommandations s'appliquent d'abord à l'entreprise étudiée et, par extension, à des organisations qui présente des caractéristiques structurelles comparables.

## 8 Limites

L'étude repose sur des données d'une seule entreprise employant environ 6 700 personnes. Cette portée fournit un cadre institutionnel cohérent mais réduit la possibilité de généraliser les résultats. Les structures de rémunération, les pratiques de promotion et les critères d'évaluation varient selon les entreprises, les secteurs et les contextes nationaux. Les conclusions s'appliquent principalement aux organisations présentant des caractéristiques similaires.

Les modèles incluent l'âge, l'expérience et l'éducation, mais d'autres dimensions essentielles restent absentes : profession détaillée, évaluations de performance, affectation départementale, secteur d'activité, interruptions

de carrière et marge de décision des responsables. Ces éléments non observés peuvent contribuer aux écarts salariaux et réduire la capacité explicative des résultats. La composante inexpliquée de 1.1 à 1.2%, souvent interprétée comme reflétant une discrimination structurelle, peut également capturer l’effet de ces variables omises.

La nature transversale des données empêche l’exploration des trajectoires salariales dans le temps. Un suivi longitudinal permettrait d’analyser la progression de carrière, les schémas de promotion et l’accumulation des désavantages au fil des trajectoires professionnelles. Les résultats décrivent un instantané de l’inégalité salariale à un moment donné, sans capturer les processus dynamiques qui façonnent les écarts sur l’ensemble de la vie professionnelle, notamment les effets de cohorte ou l’interaction entre interruptions de carrière et pénalité d’âge.

L’effet négatif de l’âge conditionnellement à l’expérience, bien qu’expliqué par des effets de cohorte et des caractéristiques organisationnelles (voir *sections 5.1.1 et 6.1.1*), mériterait une validation sur données longitudinales pour distinguer clairement les effets d’âge des effets de période et de cohorte.

Bien que la convergence des résultats à travers quatre méthodes renforce la robustesse des constats, l’inférence causale reste limitée. Les différences observées peuvent refléter des mécanismes de sélection non aléatoire : les femmes et les hommes n’entrent pas aléatoirement dans l’entreprise ni n’accèdent aux différents postes selon les mêmes contraintes institutionnelles et sociales. Ces biais de sélection affectent particulièrement l’interprétation des résultats au 90e percentile, où la réduction de l’écart peut masquer un processus de sélection drastique plutôt qu’une égalité structurelle.

## 9 Conclusion

Cette étude examine les disparités salariales de genre au sein d’une entreprise à l’aide d’un cadre économétrique combinant régression MCO, décomposition d’Oaxaca-Blinder, régression quantile et appariement par score de propension.

### 9.1 Quatre constats majeurs

#### (1) L’écart persiste même à qualifications égales

Les quatre méthodes convergent pour confirmer un écart salarial robuste de 7.8% à 8.8% après contrôle de l’âge, de l’expérience et de l’éducation.

#### (2) L’interaction domine numériquement la décomposition

La décomposition d’Oaxaca-Blinder révèle que 45-49% de l’écart provient de la composante d’interaction, c’est-à-dire de l’effet multiplicatif entre différences de dotations et différences de rendements. L’inégalité ne résulte pas de facteurs isolés mais d’un système cumulatif où les désavantages initiaux (moins d’expérience : -1.23 année) sont amplifiés par des mécanismes évaluatifs différenciés selon le genre.

#### (3) Un désavantage salarial corrélé à l’âge, 2.5 fois plus marqué pour les femmes

À expérience et éducation égales, les femmes plus âgées subissent un désavantage salarial 2.5 fois plus important que les hommes du même âge (corrélation de -3.0% vs -1.2% par année). Cette corrélation robuste ne prouve pas une causalité directe, elle pourrait refléter des interruptions de carrière non mesurées, des effets de cohorte, ou des biais d’évaluation à l’intersection âge-genre. Néanmoins, elle signale un désavantage structurel croissant pour les femmes au fil de leur carrière.

#### (4) L’inégalité touche surtout le bas et le milieu de carrière

L’analyse distributionnelle donne des écarts plus élevés dans les segments inférieurs (-11.8% au 10e percentile) et intermédiaires (-9.9% à la médiane), illustrant un “plancher collant”. Au 90e percentile, l’écart devient non significatif (-0.2%), reflétant probablement un biais de sélection drastique plutôt qu’une égalité structurelle.

### 9.2 Contributions principales

La convergence des résultats obtenus par quatre méthodes indépendantes renforce la conclusion : l’inégalité salariale de genre demeure profondément liée aux structures organisationnelles. Cette étude apporte deux contributions :

- (1) Elle démontre l'intérêt d'une approche multi-méthodes pour comprendre la nature duale de l'inégalité salariale : accumulation inégale d'expérience (composante expliquée) et pénalité d'âge disproportionnée avec valorisation différenciée des qualifications (composante inexpliquée).
- (2) Elle montre que l'inégalité salariale n'est pas uniquement un phénomène macroéconomique mais aussi un défi organisationnel interne. Des disparités de cette ampleur au sein d'une seule entreprise soulignent le rôle des pratiques de promotion, des critères d'évaluation et de la discrétion managériale.

### **9.3 Implications pour l'action**

Ces résultats fournissent une base empirique pour la conception d'interventions ciblées :

- (1) Audits d'équité salariale réguliers par département et niveau hiérarchique
- (2) Transparence des grilles salariales pour réduire la discrétion managériale
- (3) Critères de promotion explicites pour prévenir les pénalités liées à l'âge
- (4) Suivi des rendements éducatifs pour garantir une valorisation équitable des qualifications

Ces actions s'inscrivent dans le cadre de l'Objectif de Développement Durable 5 des Nations Unies (égalité des genres et autonomisation des femmes, cible 5.5 sur la participation pleine et l'égalité des chances en leadership).

L'écart salarial de genre résulte d'une interaction complexe entre dotations inégales et valorisation différenciée. Comprendre cette dualité constitue la première étape vers des politiques organisationnelles plus justes et plus efficaces.

## 10 Références

### Ouvrages et chapitres de livre

- Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. National Bureau of Economic Research; distributed by Columbia University Press.
- Doeringer, P. B., & Piore, M. J. (1971). Internal labor markets and manpower analysis. D.C. Heath and Company.
- Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In O. Ashenfelter & D. Card (Eds.), Handbook of labor economics (Vol. 4, pp. 1–102). Elsevier. [https://doi.org/10.1016/S0169-7218\(11\)00407-2](https://doi.org/10.1016/S0169-7218(11)00407-2)
- Greene, W. H. (2018). Econometric analysis (8th ed.). Pearson.
- Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119482260>
- Mincer, J. (1974). Schooling, experience, and earnings. National Bureau of Economic Research; Columbia University Press.

### Méthodes statistiques et économétriques

- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. The Journal of Human Resources, 8(4), 436–455. <https://doi.org/10.2307/144855>
- Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles. Econometrica, 46(1), 33–50. <https://doi.org/10.2307/1913643>
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review, 14(3), 693–709. <https://doi.org/10.2307/2525981>
- Palmer, C. (2011). Interpretation of  $\beta$  in log-linear models.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838. <https://doi.org/10.2307/1912934>

### Articles empiriques sur les écarts salariaux

- Baker, M., Halberstam, Y., Kroft, K., Mas, A., & Messacar, D. (2019). Pay transparency and the gender gap. American Economic Journal: Applied Economics 15 (2): 157–83. DOI: 10.1257/app.20210141
- Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. Journal of Economic Literature, 55(3), 789–865. <https://doi.org/10.1257/jel.20160995>
- Cardoso, A. R., Guimarães, P., & Portugal, P. (2016). What drives the gender wage gap? A look at the role of firm and job-title heterogeneity. Oxford Economic Papers, 68(2), 506–524. <https://doi.org/10.1093/oep/gpv069>
- Castilla, E. J., & Benard, S. (2010). The paradox of meritocracy in organizations. Administrative Science Quarterly, 55(4), 543–576. <https://doi.org/10.2189/asqu.2010.55.4.543>
- Goldin, C. (2014). A grand gender convergence: Its last chapter. American Economic Review, 104(4), 1091–1119. <https://doi.org/10.1257/aer.104.4.1091>
- Mas, A. (2017). Does transparency lead to pay compression? Journal of Political Economy, 125(5), 1683–1721. <https://doi.org/10.1086/693137>
- Mincer, J., & Polachek, S. (1974). Family investments in human capital: Earnings of women. Journal of Political Economy, 82(2, Part 2), S76–S108. <https://doi.org/10.1086/260293>
- Petersen, T., & Morgan, L. A. (1995). Separate and unequal: Occupation-establishment sex segregation and the gender wage gap. American Journal of Sociology, 101(2), 329–365. <https://doi.org/10.1086/230727>
- Rubery, J., & Grimshaw, D. (2015). The 40-year pursuit of equal pay: A case of constantly moving goalposts. Cambridge Journal of Economics, 39(2), 319–343. <https://doi.org/10.1093/cje/beu053>

### Rapports et sources institutionnelles

- Forum économique mondial. (2023). Global gender gap report 2023. <https://www.weforum.org/publications/global-gender-gap-report-2023/>
- Institut national de la statistique et des études économiques (INSEE). (2023). Écarts de rémunération femmes-hommes. <https://www.insee.fr/fr/statistiques>
- Nations Unies. (2015). Objectifs de développement durable. <https://sdgs.un.org/fr/goals>

**Note sur le dataset**

Kashif, A. R. (n.d.). Salary data [Dataset]. Kaggle. <https://www.kaggle.com/datasets/ahmadrazakashif/salary-data>

## 11 Annexes

### 11.1 Statistiques descriptives par genre

Tableau 15: Statistiques descriptives quantitatives par genre

Statistique	Female	Male	Différence
Sample Size	3011	3683	-672
Age (Mean)	32.63	34.44	-1.81
Age (Min)	21	22	-1
Age (Max)	60	62	-2
Experience (Mean)	7.42	8.65	-1.23
Experience (Min)	0	0	0
Experience (Max)	34	32	2
Salary (Mean)	107960.31	121478.37	-13518.06
Salary (Min)	25000	25000	0
Salary (Max)	220000	250000	-30000
Log Salary (Mean)	11.44	11.59	-0.15

Tableau 16: Répartition du niveau d'éducation par genre

Education level	Female	Male
High School	8.3%	5.3%
Bachelor's	39.7%	49.5%
Master's	35.5%	21.5%
PhD	16.5%	23.7%

### 11.2 Équations complètes

#### 11.2.1 Décomposition d'Oaxaca-Blinder

$$\begin{aligned} \overline{\ln(\text{Salary})}_M - \overline{\ln(\text{Salary})}_F &= (\overline{\text{Age}}_M - \overline{\text{Age}}_F) \hat{\beta}_{\text{Age}} + (\overline{\text{Experience}}_M - \overline{\text{Experience}}_F) \hat{\beta}_{\text{Experience}} + \\ &\quad (\overline{\text{Education}}_M - \overline{\text{Education}}_F) \hat{\beta}_{\text{Education}} + \left( \hat{\beta}_{0,M} - \hat{\beta}_{0,F} \right) + \left( \hat{\beta}_{\text{Age},M} - \hat{\beta}_{\text{Age},F} \right) \overline{\text{Age}}_F + \\ &\quad \left( \hat{\beta}_{\text{Experience},M} - \hat{\beta}_{\text{Experience},F} \right) \overline{\text{Experience}}_F + \left( \hat{\beta}_{\text{Education},M} - \hat{\beta}_{\text{Education},F} \right) \overline{\text{Education}}_F \end{aligned}$$

#### Interprétation des termes :

- Le côté gauche représente la différence de salaires logarithmiques moyens entre les hommes ( $M$ ) et les femmes ( $F$ ).
- Le premier ensemble de termes capture la **composante expliquée**, des différences moyennes d'âge, d'expérience et d'éducation entre les deux groupes.
- Le deuxième ensemble de termes capture la **composante inexpliquée**, qui découle des différences dans les rendements de ces caractéristiques, y compris l'ordonnée à l'origine.
- Les coefficients  $\hat{\beta}$  représentent les rendements estimés pour chaque caractéristique. Les indices  $M$  et  $F$  indiquent des structures salariales distinctes pour les hommes et les femmes.



### 11.2.2 Régression quantile

$$Q_{\tau}(\ln(\text{Salary}_i) \mid X_i) = \beta_{0,\tau} + \beta_{1,\tau} \cdot \text{Female}_i + \beta_{2,\tau} \cdot \text{Age}_i + \beta_{3,\tau} \cdot \text{YearsExperience}_i + \beta_{4,\tau} \cdot \text{HighSchool}_i + \beta_{5,\tau} \cdot \text{Bachelor}_i + \beta_{6,\tau} \cdot \text{Master}_i + \beta_{7,\tau} \cdot \text{PhD}_i$$

$Q_{\tau}(\cdot)$  désigne la fonction quantile conditionnelle au percentile  $\tau$  et  $\beta_{j,\tau}$  les coefficients qui varient selon les quantiles

### 11.2.3 Appariement par score de propension

$$P(\text{Female}_i = 1 \mid X_i) = \Lambda(\gamma_0 + \gamma_1 \cdot \text{Age}_i + \gamma_2 \cdot \text{YearsExperience}_i + \gamma_3 \cdot \text{EducationLevel}_i)$$

où  $\Lambda(\cdot)$  désigne la fonction logistique.

## 11.3 Tests de robustesse et diagnostics

### 11.3.1 Tests de diagnostic MCO

Tableau 17: VIF par variable (multicolinéarité)

Variable	VIF
gender	1.03
age	2.92
years_of_experience	2.94
education_level	1.09

Tableau 18: Tests de diagnostic du modèle MCO

Test	Statistique	p_value
BP Breusch-Pagan (hétéroscédasticité)	582.497	0.00e+00
Jarque-Bera (normalité des résidus)	21.180	2.52e-05

### 11.3.2 Tests de spécification

Tableau 19: Tests de spécification du modèle MCO

Test	Statistique	p_value
RESET Ramsey RESET	1177.208	0
Linéarité âge/expérience	1577.654	0

#### Interprétation des tests diagnostiques :

- **Multicolinéarité (VIF)** : Tous les VIF sont inférieurs à 5, donc absence de multicolinéarité **problématique** entre les variables explicatives.
- **Hétéroscédasticité (Breusch-Pagan)** : Le test rejette l'homoscédasticité ( $p < 0.001$ ). Des erreurs standard robustes HC3 sont utilisées dans toutes les estimations MCO pour garantir la validité de l'inférence statistique (White, 1980).
- **Normalité des résidus (Jarque-Bera)** : Le test rejette la normalité ( $p < 0.001$ ). Avec  $n = 6\,694$  observations, le Théorème Central Limite assure néanmoins la validité asymptotique des estimateurs et des

tests (Greene, 2018).

- **Spécification (RESET et linéarité)** : Les tests rejettent la spécification linéaire ( $p < 0.001$ ), il y a donc présence de non-linéarités. Cela justifie l'utilisation complémentaire de la régression quantile (*section 5.3*) et de l'appariement par score de propension (*section 5.4*), méthodes moins dépendantes de la forme fonctionnelle.

Ces limites ne remettent pas en cause les conclusions principales mais renforcent la nécessité d'une approche multi-méthodes. La convergence des estimations MCO, Oaxaca-Blinder, régression quantile et PSM confirme la persistance de l'écart salarial de genre.

Tableau 20: Synthèse des tests diagnostiques et corrections appliquées

Test	Problème détecté	Correction appliquée
Breusch-Pagan	Hétéroscédasticité	Erreurs robustes HC3
Jarque-Bera	Non-normalité résidus	Échantillon large (n=6694)
Ramsey RESET	Spécification incorrecte	Approche multi-méthodes
Linéarité	Non-linéarité âge/exp	Régression quantile + PSM

## 11.4 Décomposition d'Oaxaca-Blinder

### 11.4.1 Note sur les structures de référence (Group\_Weight)

Les différentes valeurs de Group\_Weight dans le *Tableau 8* correspondent à différentes structures de référence utilisées pour la décomposition :

Group\_Weight = 0 : utilise la structure salariale masculine comme référence

Group\_Weight = 1 : utilise la structure salariale féminine comme référence

Group\_Weight = 0.5 : utilise une moyenne simple des deux structures

Group\_Weight = 0.55 : utilise une moyenne pondérée proche de la répartition observée

Group\_Weight = -1 et -2 : pondérations alternatives pour tests de robustesse

La variation limitée des résultats selon ces différentes spécifications atteste de la robustesse de la décomposition.

## Liste des tableaux

1	Statistiques descriptives . . . . .	2
2	Pourcentage de valeurs manquantes par variable et au total . . . . .	3
3	Statistiques descriptives par genre . . . . .	3
4	Distribution du niveau d'éducation par genre . . . . .	4
5	Tableau synthétique des méthodes d'analyse . . . . .	8
6	Résultats du modèle MCO . . . . .	9
7	Salaires logarithmiques moyens par genre . . . . .	10
8	Décomposition globale d'Oaxaca–Blinder (expliqué vs inexpliqué) . . . . .	10
9	Contributions des variables à la décomposition d'Oaxaca–Blinder . . . . .	11
10	Coefficients de régression spécifiques par groupe (Hommes vs Femmes) et différences . . . . .	12
11	Estimations de régression quantile aux 10e, 50e et 90e percentiles . . . . .	13
12	Écart salarial de genre estimé dans l'échantillon apparié . . . . .	13
13	Diagnostics d'équilibre après appariement par score de propension . . . . .	13
14	Comparaison des méthodes d'estimation de l'écart salarial . . . . .	14
15	Statistiques descriptives quantitatives par genre . . . . .	21
16	Répartition du niveau d'éducation par genre . . . . .	21
17	VIF par variable (multicolinéarité) . . . . .	22
18	Tests de diagnostic du modèle MCO . . . . .	22
19	Tests de spécification du modèle MCO . . . . .	22
20	Synthèse des tests diagnostiques et corrections appliquées . . . . .	23

## Liste des figures

1	Distributions du salaire brut et du ln(salaire) . . . . .	2
2	Distribution des variables clés . . . . .	4
3	Distribution du salaire par genre et éducation . . . . .	5
4	Corrélations entre les variables continues . . . . .	5