

MASTER
ESA

Université d'Orléans

SEGMENTATION DE LGD

BARRAUD Lorenzo
ENIONA Raphaël
MIRZA Simon



CONTEXTE

La Loss Given Default (LGD), est un indicateur crucial dans la gestion des risques financiers, particulièrement lorsqu'on opère sur une base caution.

Son rôle clé dans la rentabilité des prêteurs et le calcul des réserves réglementaires en fait un élément essentiel pour les institutions cherchant à minimiser les pertes et stabiliser le système financier.

OBJECTIF

Réaliser une segmentation pour former des groupes de contrats qui partagent des caractéristiques communes et des niveaux de LGD homogènes.

SOMMAIRE

- ◆ Analyse des données
- ◆ Modèle de segmentation
- ◆ Conclusion



1. ANALYSE DES DONNÉES

1.1 STATISTIQUES DESCRIPTIVES

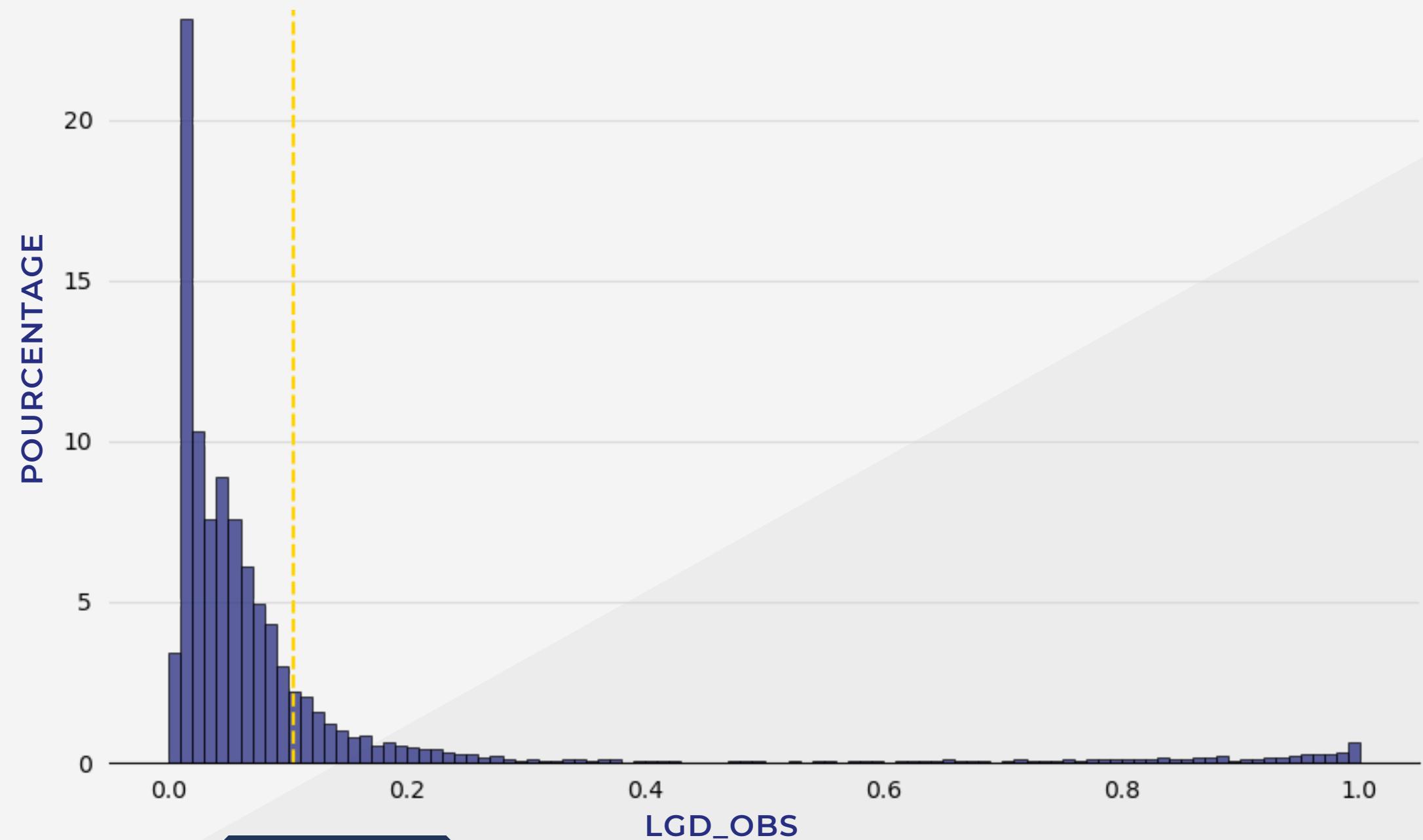
1.2 MODIFICATIONS EFFECTUÉES

1.3 SÉLECTION DES VARIABLES



1.1 Statistiques descriptives

1.1 Statistiques descriptives



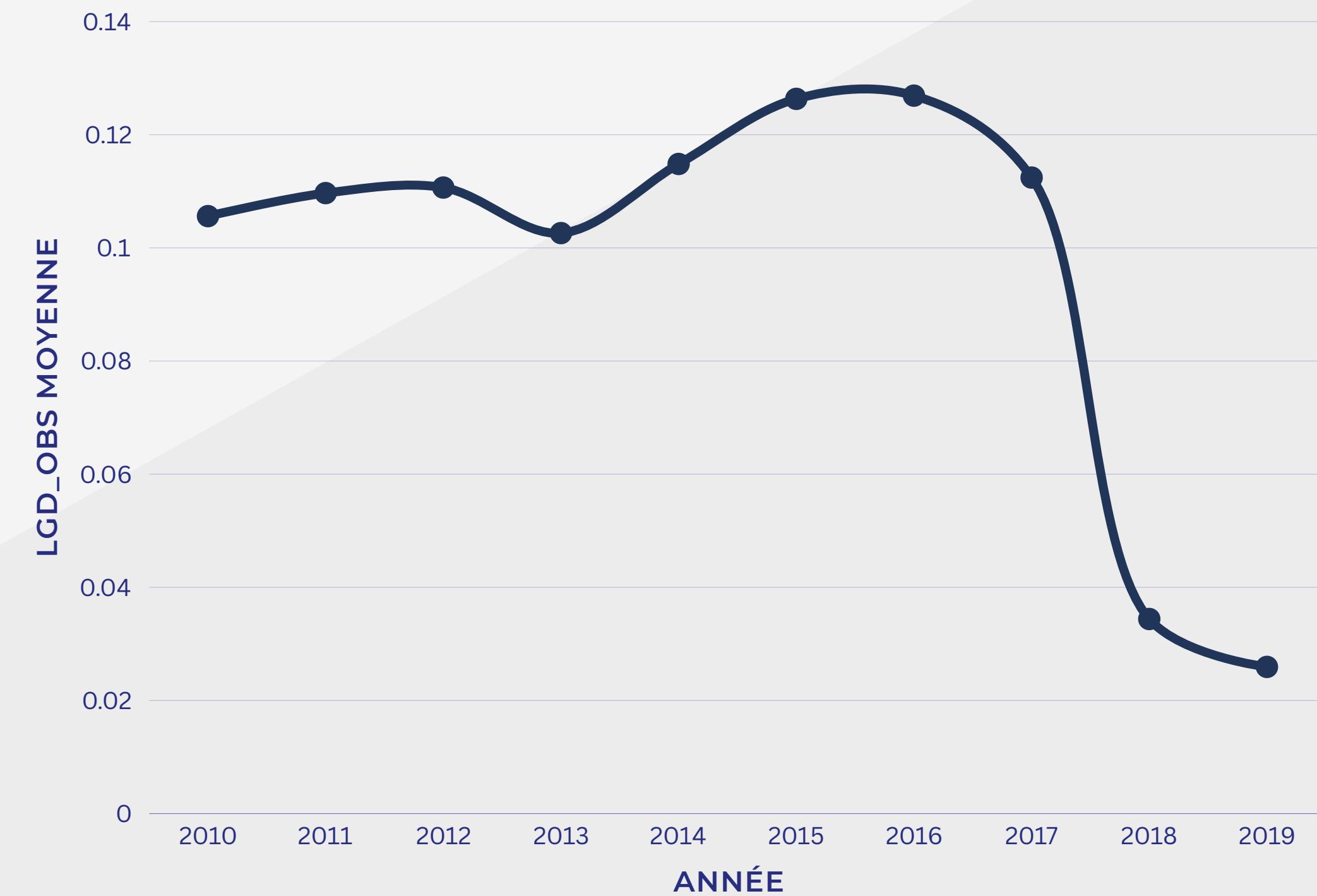
DISTRIBUTION DE
LA LGD (%)

mean	0.105
std	0.193
25%	0.0193
50%	0.047
75%	0.087

1.1 Statistiques descriptives

LGD MOYENNE
AU COURS DU
TEMPS

Diminution de la LGD moyenne depuis 2018



1.2 Modification effectuées

1.2 Modification effectuées

Les valeurs manquantes

Variable	annee_construction	EPARGNE	MONTANT_APPORT	COUT_ACQUISITION	LTV_ACTI	LTV_OCTR_OI	TAUX_APPORT	ANCIENNE_TE_PROFESSIONELLE
Traitement	Variable supprimée	Variable supprimée	Variable supprimée	Variable supprimée	Remplacement par la valeur moyenne	Remplacement par la valeur moyenne	Remplacement par la valeur moyenne	Remplacement par l'âge de l'individu moins dix-huit ans

Valeurs manquantes trop nombreuses ← → **Variables corrigées**

1.2 Modification effectuées

Les valeurs aberrantes et création de variables

Variable	AGE_EMP_PRI	NOTE_OCTROI et NOTE_COMPORTE MTALE_OCTROI	tx_experience	montant_bien	restant_du
Traitement	Remplacement des valeurs supérieures à 90 par ANCIENNETE_PRO FESSIONNELLE plus dix-huit ans*	Remplacement des valeurs mal renseignées par le mode	tx_experience est égale à AGE_EMP_PRI divisée ANCIENNETE_PRO FESSIONNELLE	montant_bien est égale à MNT_PRET divisée par LTV_OCTROI	restant_du est égale à LTV_ACTI multipliée par montant_bien

*Quelques observations avaient une ancienneté professionnelle plus grande que leur âge, ces valeurs ont été remplacées à l'aide d'un KNN

1.2 Modification effectuées

Regroupement des modalités

Variable	NATURE_BIEN	ETAT_DU_BIEN	USAGE_BIEN	CONTRAT_TRAVAIL	EPARGNE
Traitement	Regroupement en 'Autres' : Terrain, Garage, Annexes, Local mixte, Local professionnel, Autres	Regroupement en 'Neuf' : Contrat Construction MI, VEFA, Neuf	Regroupement en 'Autres' : Résidence secondaire, Logement locatif secondaire, Résidence de retraite, Autres	Regroupement en 'Autres' : Retraite, Intérimaire, Chomeur, Autres	Variable dummy : 1 si EPARGNE > 0 0 sinon

1.3 Sélection des variables

1.3 Sélection des variables

A l'aide du test de corrélation de Spearman et du V de Cramer, nous avons examiné les corrélations entre les variables et avec la variable cible. Après nos traitements les variables significatives retenues sont :

- TAUX_INTERET
- MNT_PRET
- PROJET_IMMOBILIER
- USAGE_BIEN
- AGE_EMP_PRI
- TAUX_APPORT
- LTV_OCTROI
- LTV_ACTI
- TAUX_ENDETTEMENT
- NOTE_COMPORTEMENTALE_OCTROI
- INDICATEUR_SCI
- NBR_ENF_CHARGE
- REVENUS_EMPRUNTEUR
- ANCIENNUTE_PROFESSIONNELLE
- EPARGNE_DUMMY
- CONTRAT_TRAVAIL_ENCODED
- MONTANT_BIEN
- RESTANT_DU
- TX_EXPERIENCE
- MARCHE_ENCODED
- ETAT_DU BIEN_ENCODED
- NATURE_BIEN_ENCODED
- FEUILLE
- ANNEE_DEFAUT
- NOTE_OCTROI

*Le suffixe 'encoded' correspond aux variables qualitatives encodées. Ce traitement était nécessaire pour l'arbre de classification.

2. MODÈLE DE SEGMENTATION

2.1 ARBRE DE DÉCISION

2.2 XGBOOST

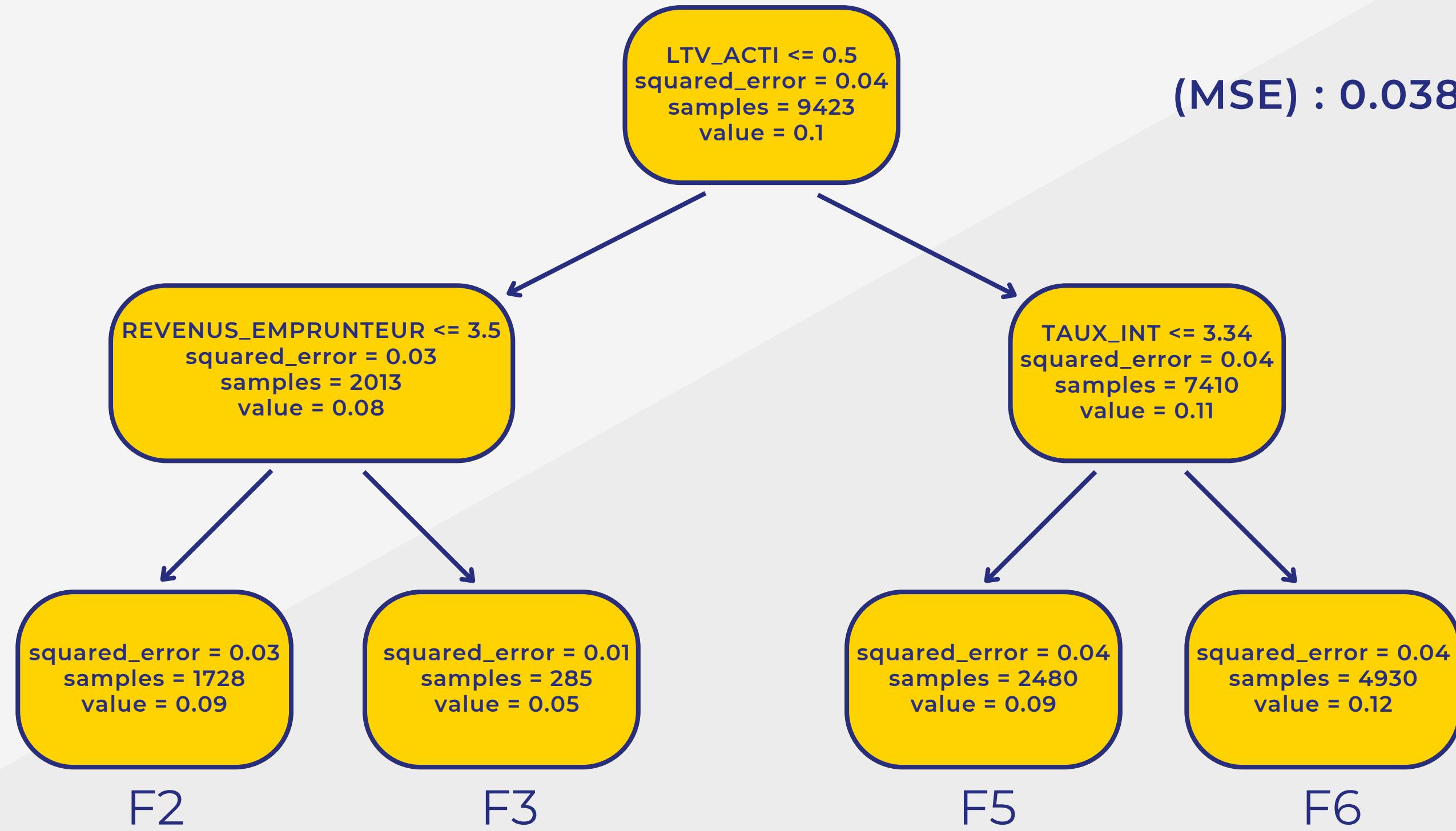
2.3 CLUSTERING

2.4 INTÉGRATION DU DPE

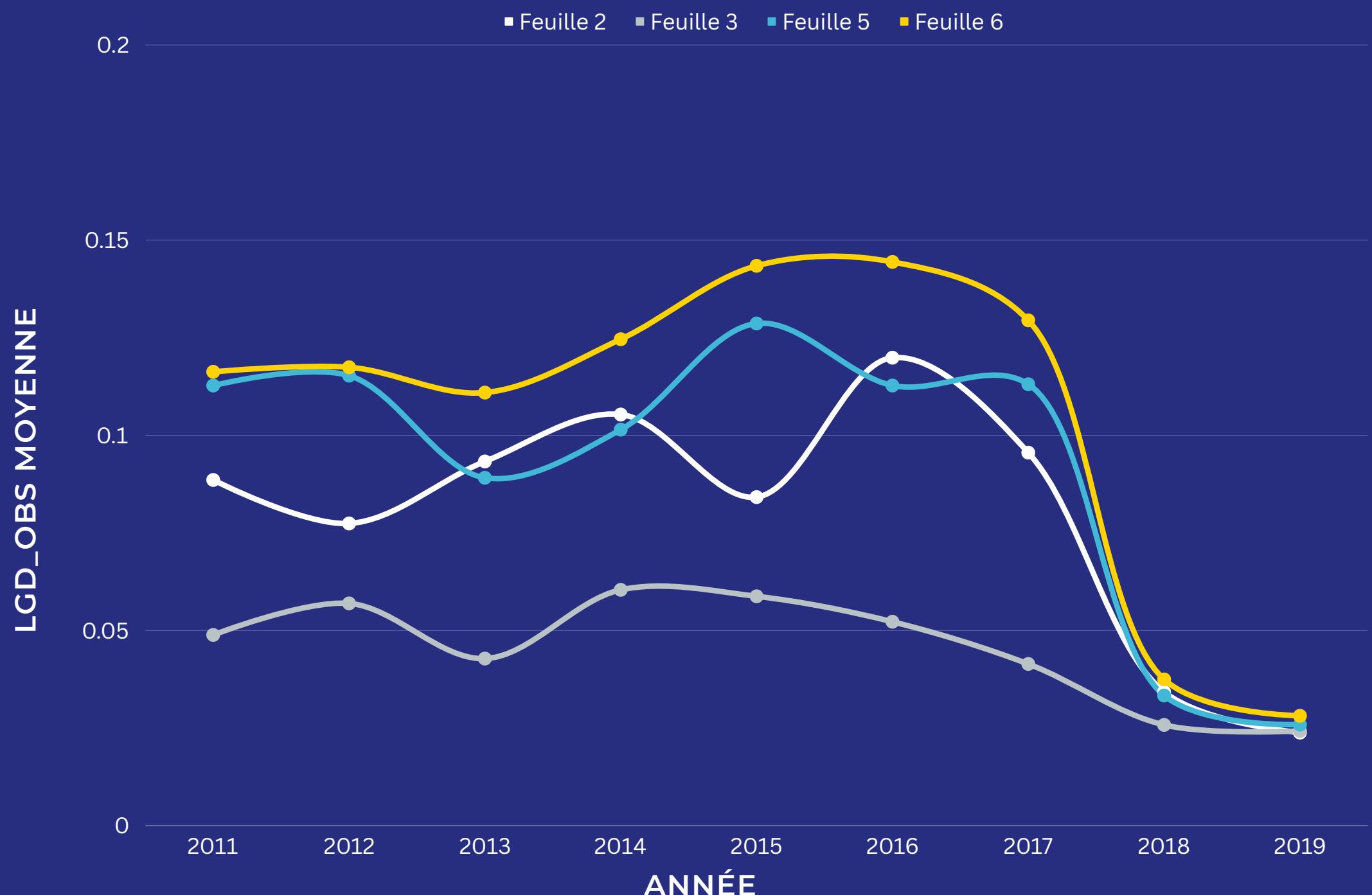


2.1 Arbre de décision

2.1 Arbre de décision



2.1 Arbre de décision



STABILITÉ EN RISQUE

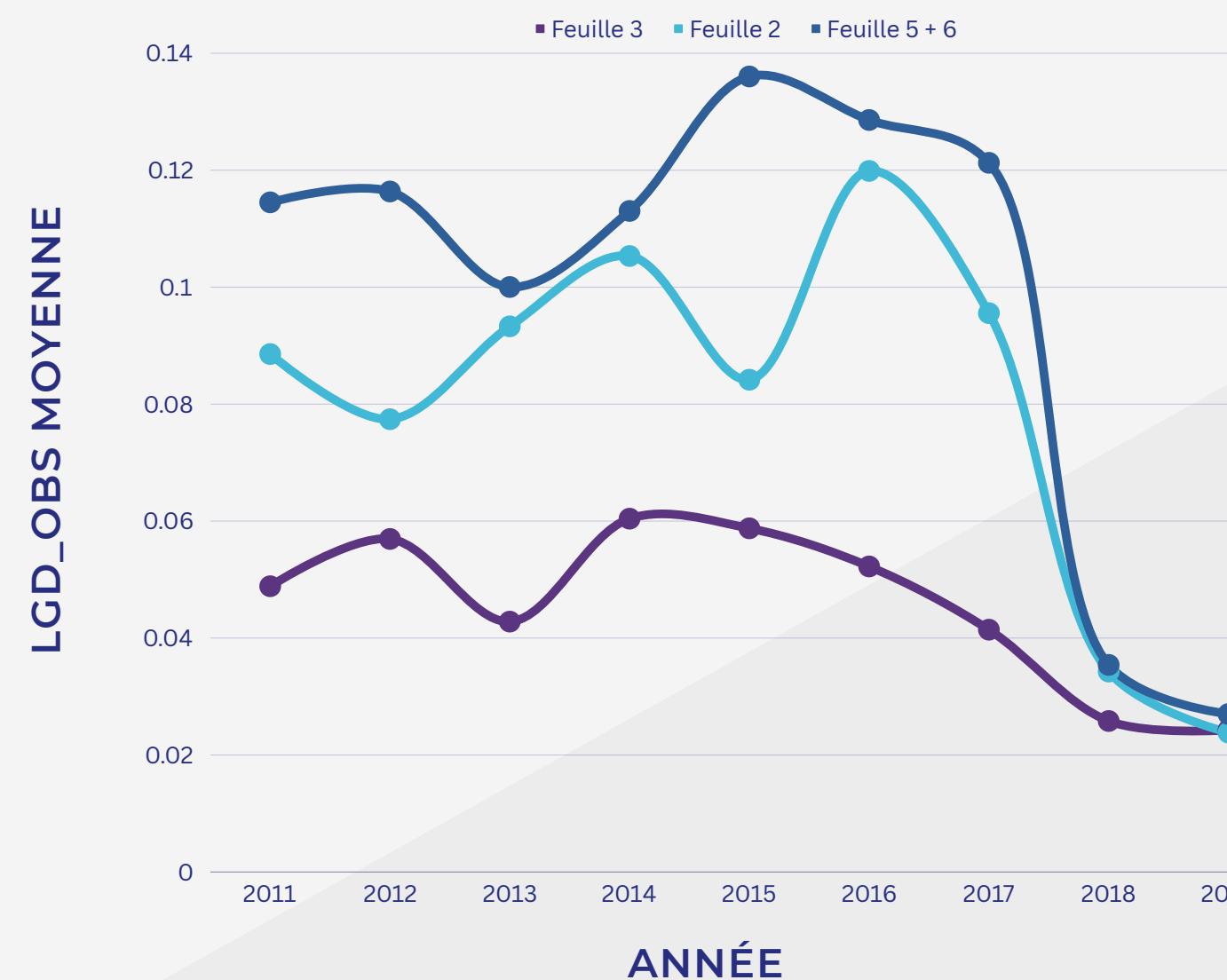
Nous observons des croisements entre les feuilles 5 et 6, la stabilité en risque n'est pas assurée.

A noter : 2010 et 2019 ne sont pas pris en compte car n'avons pas assez d'individus pour 2010 et 2019

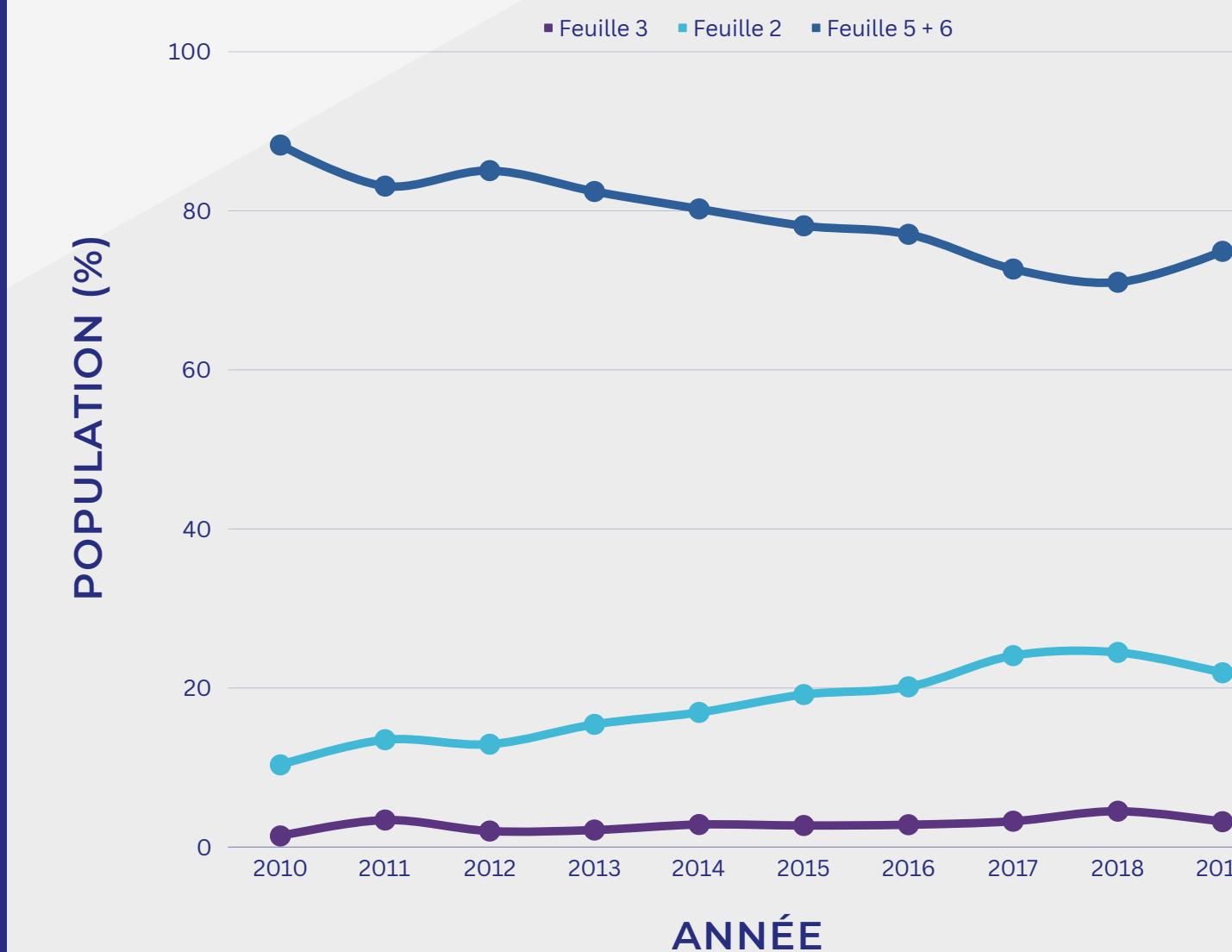
2.1 Arbre de décision

Regroupons les **feuilles 5 et 6** afin d'assurer la stabilité en risque et en volume

STABILITÉ EN RISQUE



STABILITÉ EN VOLUME



2.1 Arbre de décision

RÉSULTATS

F2	0.05
F3	0.09
F5 + F6	0.11

ÉCARTS RELATIFS

F2 vs F3	0.80
F2 vs F5 + F6	0.22
F3 vs F5 + F6	1.20

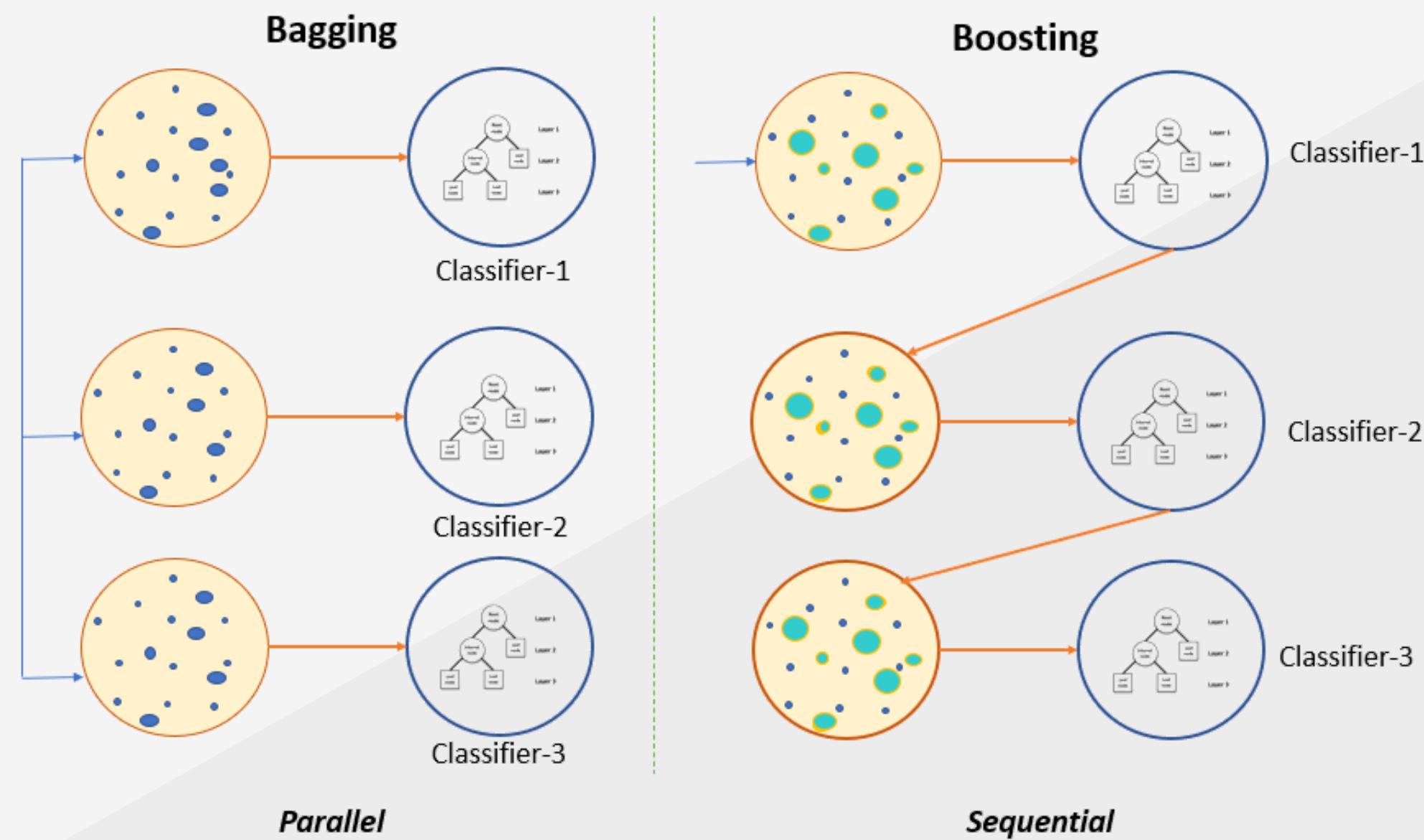
En élaguant notre arbre de décision, nous pouvons obtenir à la fois une stabilité en risque et en volume. Notre segmentation est robuste dans le temps.

Nous avons un des écarts relatifs qui n'est pas respecté entre la feuille 2 et la combinaison de la feuille 5 et 6.

2.2 XGBOOST

2.2 XGBOOST

Nous avons décidé d'utiliser l'algorithme **XGBoost** afin d'estimer les LGD de nos clients à l'aide des variables explicatives.



2.2 XGBOOST

RÉSULTATS

- Utilisation de la **validation croisée** afin de limiter le surapprentissage
- Les LGD estimées vont être utilisées pour former des clusters par la suite.

Hyperparamètres

Learning rate	0.1
Profondeur	4
Nombre d'arbres	40

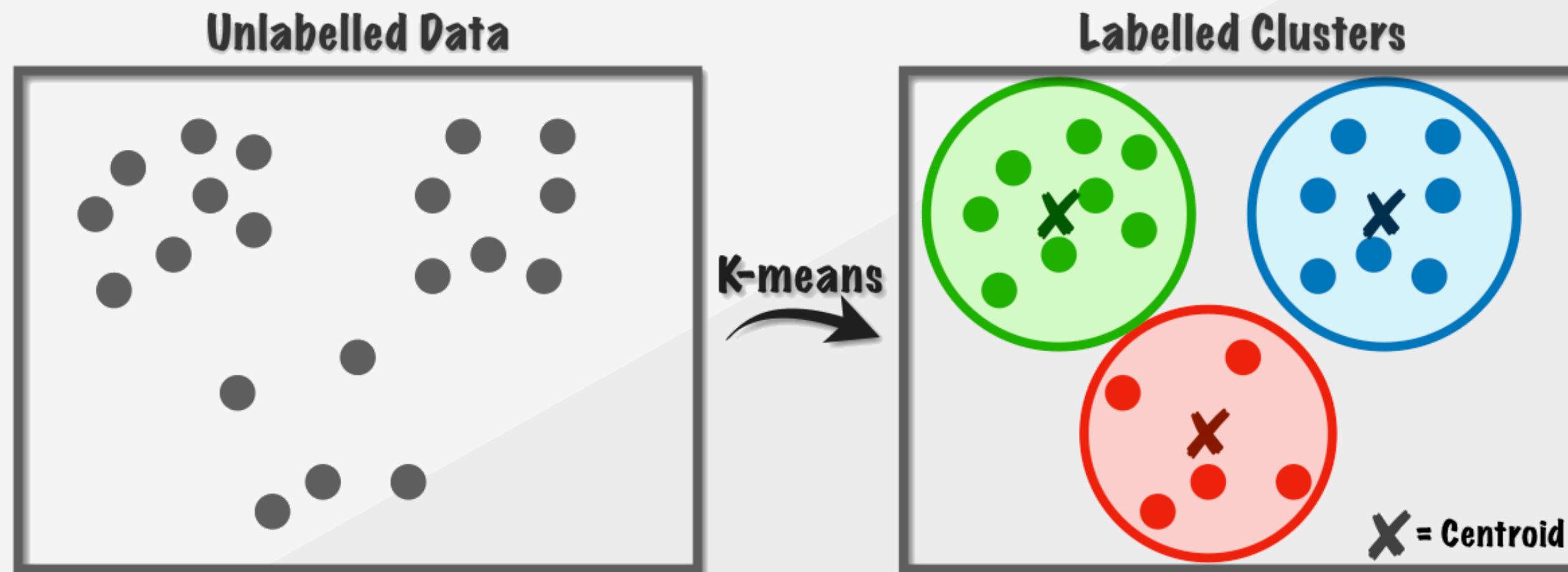
Résultats

MSE	0.04
MAE	0.10

2.3 Clustering

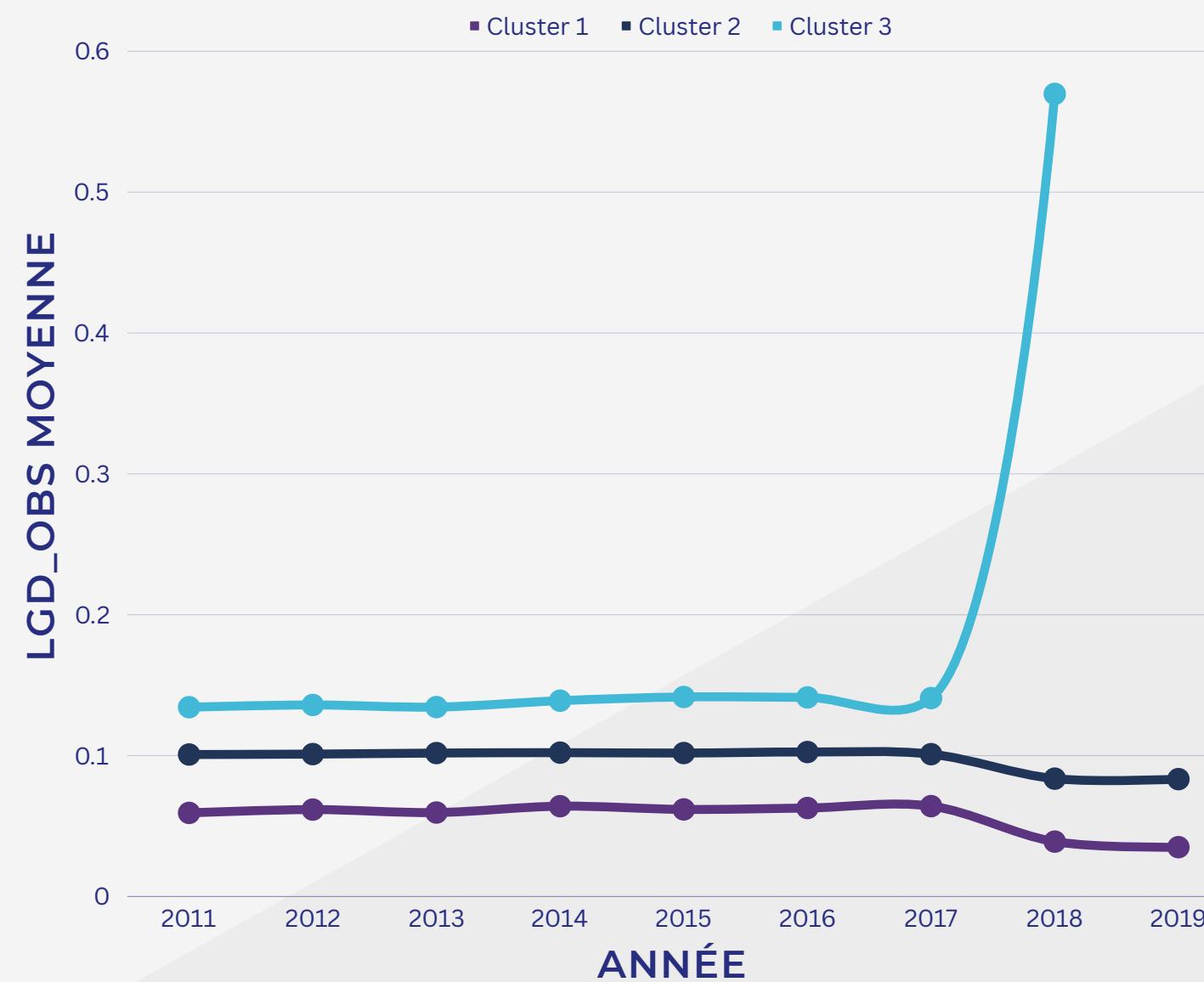
2.3 Clustering

Nous avons décidé d'utiliser l'algorithme **K-Means** afin de former des classes de LGD grâce à nos prédictions du XGBoost.

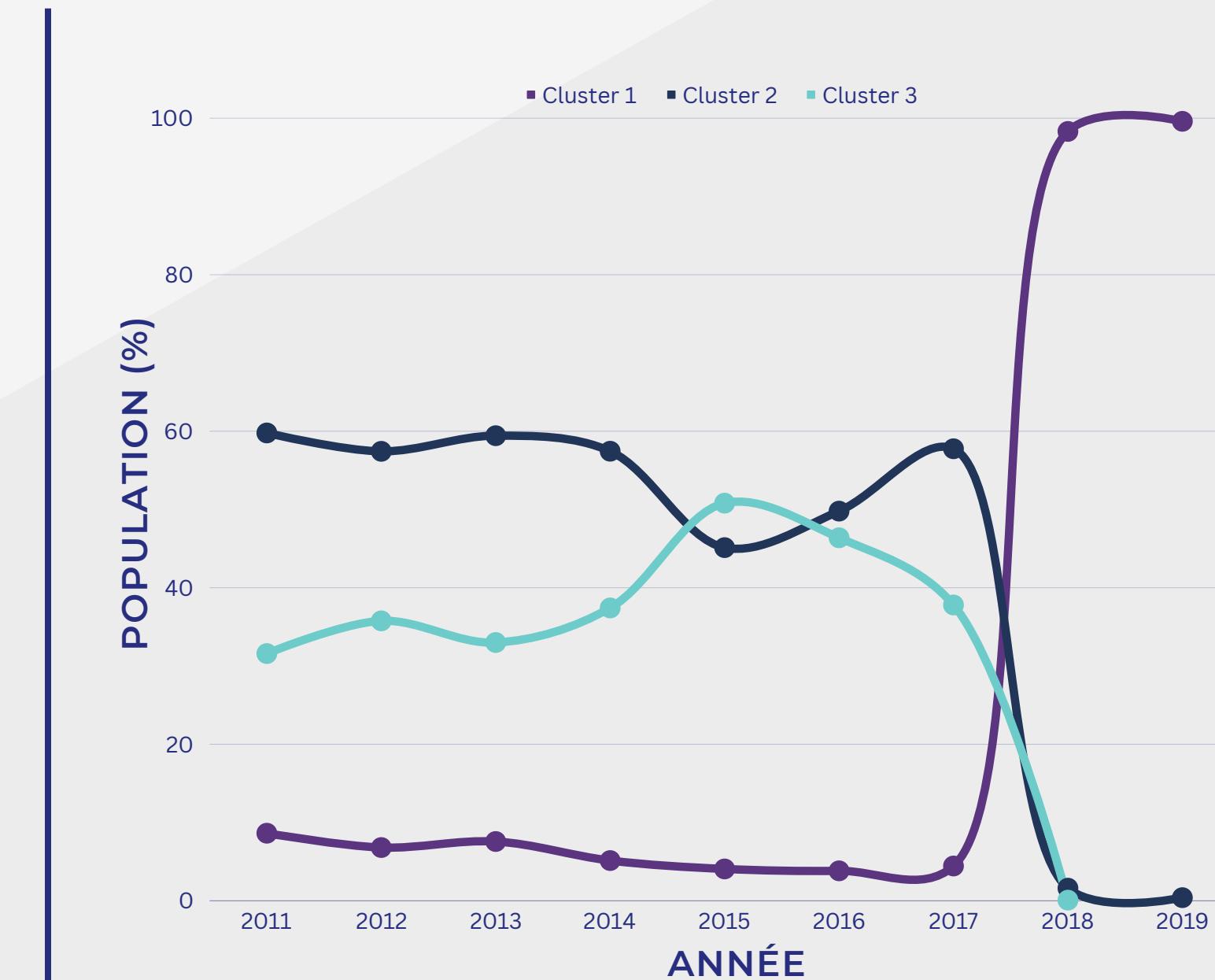


2.3 Clustering

STABILITÉ EN RISQUE



STABILITÉ EN VOLUME



2.3 Clustering

RÉSULTATS

C1	0.04
C2	0.10
C3	0.14

ECARTS RELATIFS

C1 vs C2	1.31
C1 vs C3	2.16
C2 vs C3	0.37

La stabilité en risque est assurée. A contrario la stabilité en volume n'est pas respectée. Cela peut s'expliquer par le fait que la LGD moyenne des clients a **fortement baissé** durant les dernières années

Les écarts relatifs sont tous supérieurs à 30 %

2.3 Clustering

COMPARAISON

	Arbre	K-Means
LGD Faible	0.05	0.04
LGD Moyenne	0.09	0.10
LGD Haute	0.11	0.14

- **Similarité** entre nos classes créées par notre arbre de décision et celles créées par le K-Means

2.4 Intégration du DPE

2.4 Intégration du DPE

Etiquette_DPE et Etiquette_GES :

- V-Cramer : 0.63

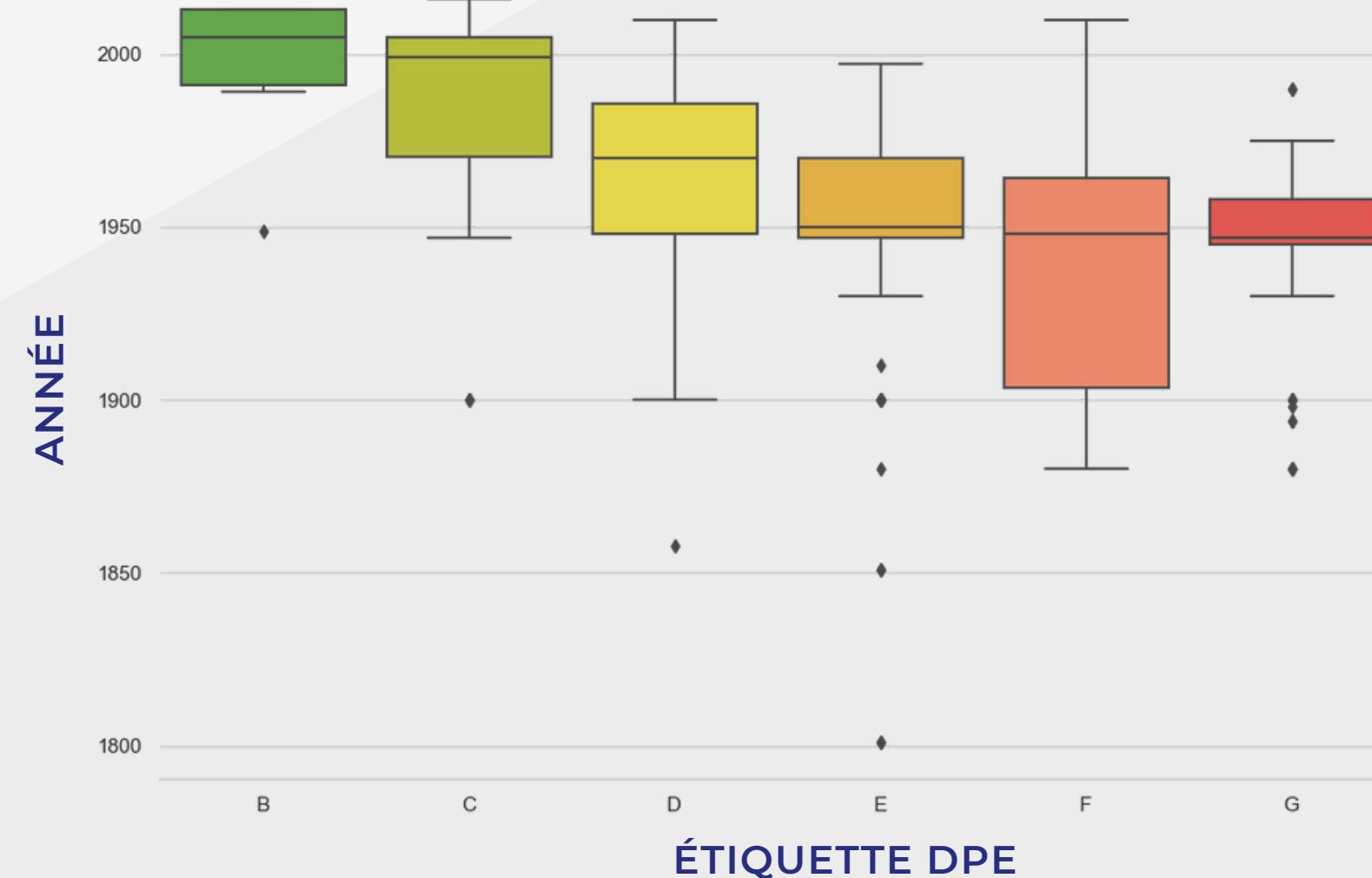
Indicateur de l'association entre les deux variables

Etiquette_DPE et Etiquette_GES

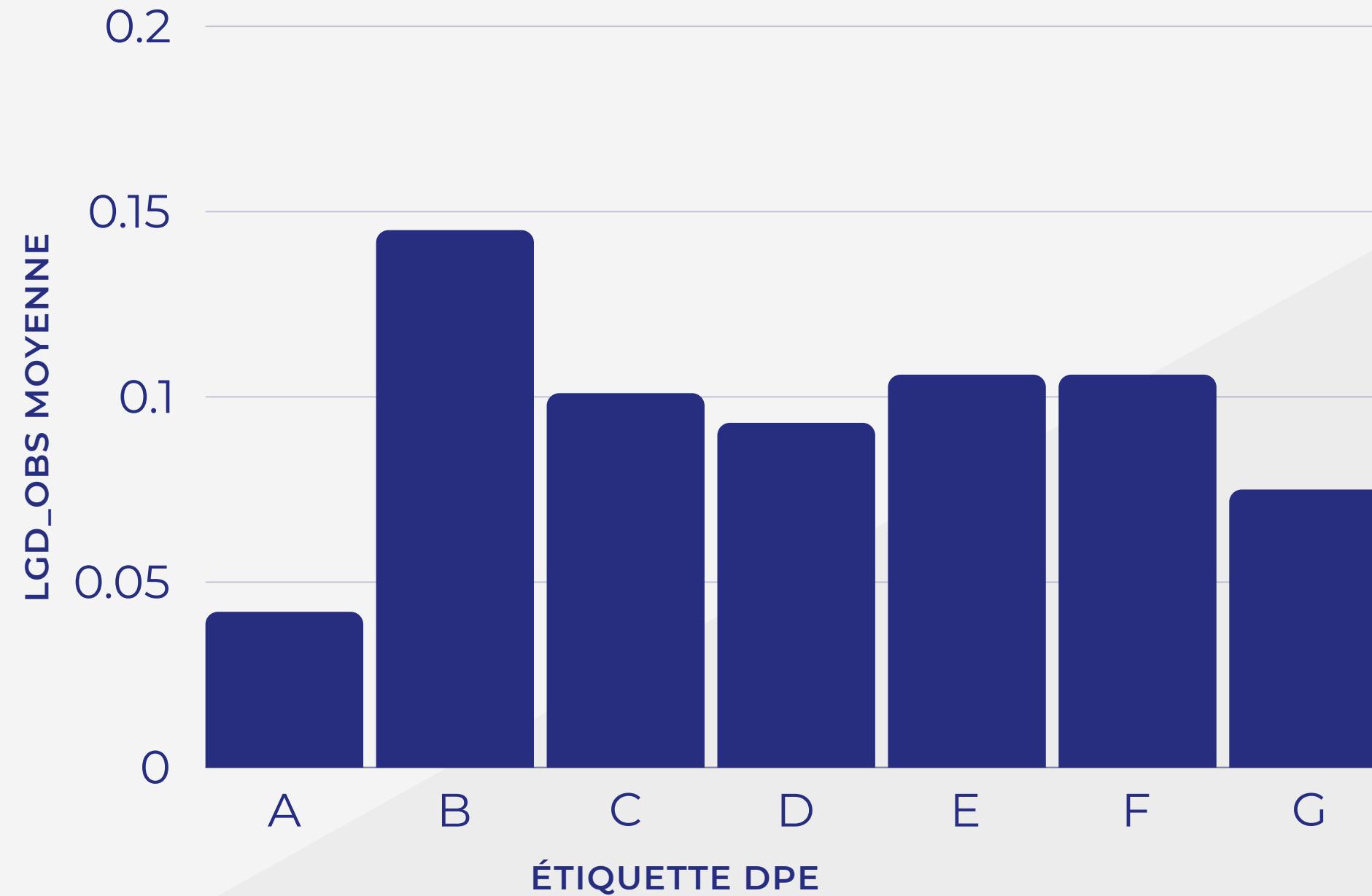
- Chi2 : 3578.72
- P-value : 0.0
- Degrees of Freedom : 36

Rejet de H0 donc on peut dire qu'il y a une association entre les deux variables

Les boxplots révèlent des différences significatives dans les niveaux de Diagnostic de Performance Énergétique (DPE) selon la classe d'année de construction



2.4 Intégration du DPE

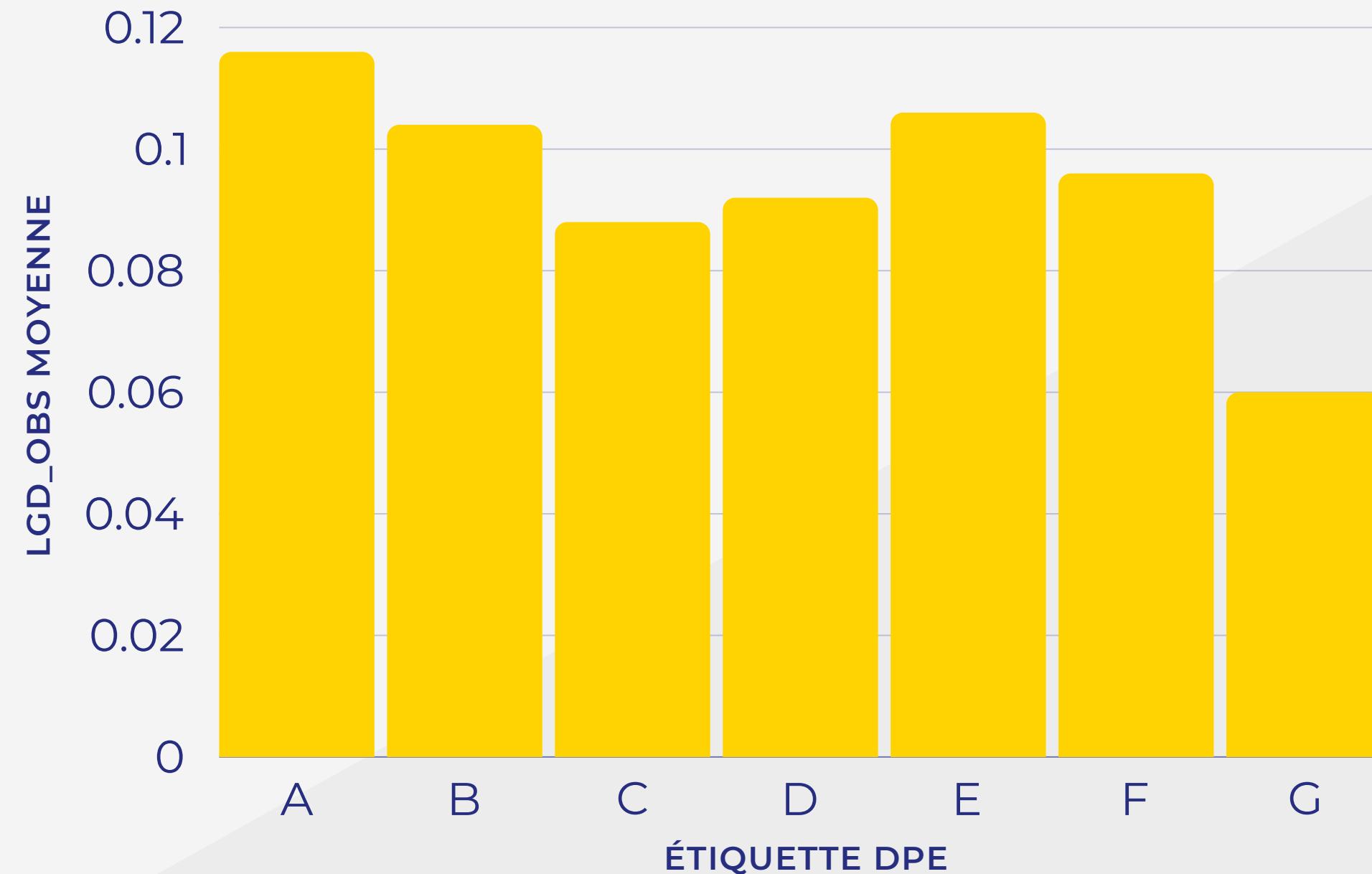


Globalement, on observe un lien peu marqué entre le niveau de LGD et le Diagnostic de Performance Énergétique (DPE).

Statistique du test de Kruskal-Wallis : 6.0
P-valeur : 0.423

La médiane diffère significativement entre au moins deux groupes.

2.4 Intégration du DPE



En général, on constate un lien peu significatif entre le niveau de LGD et les émissions de gaz à effet de serre (GES).

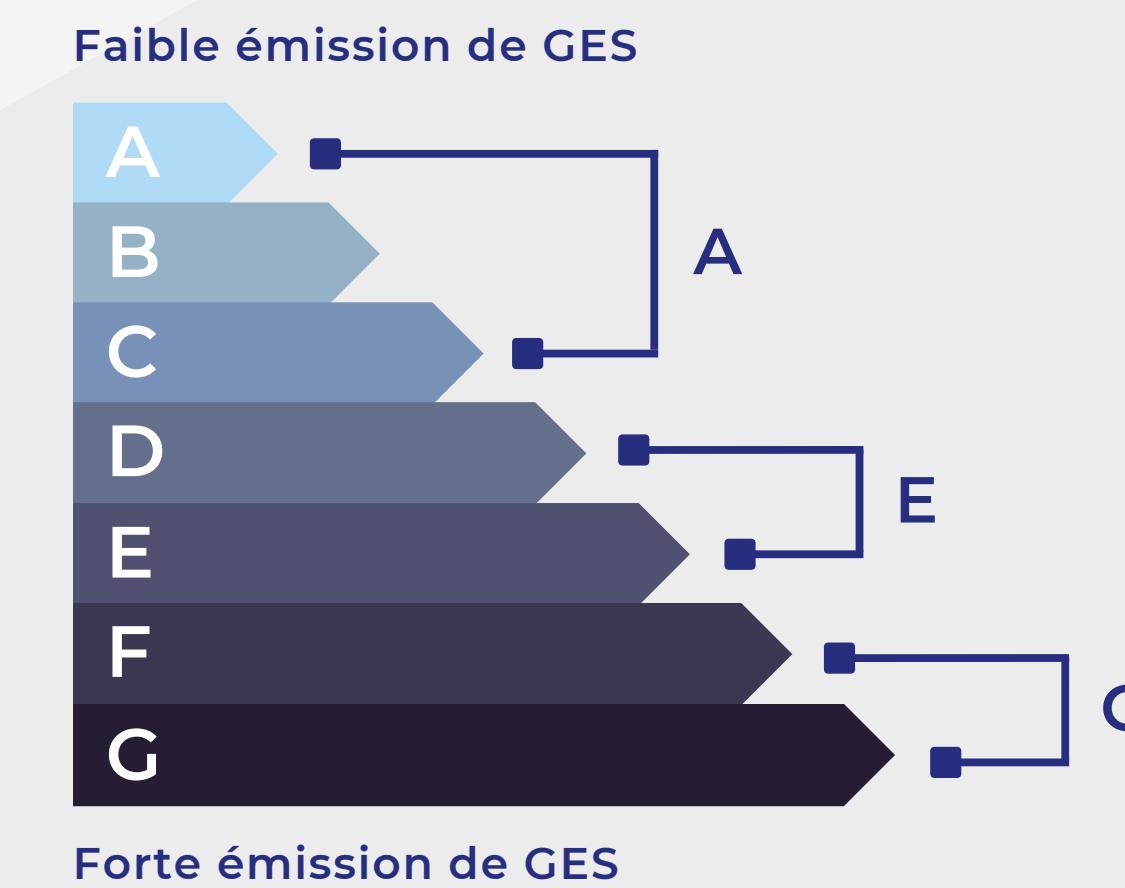
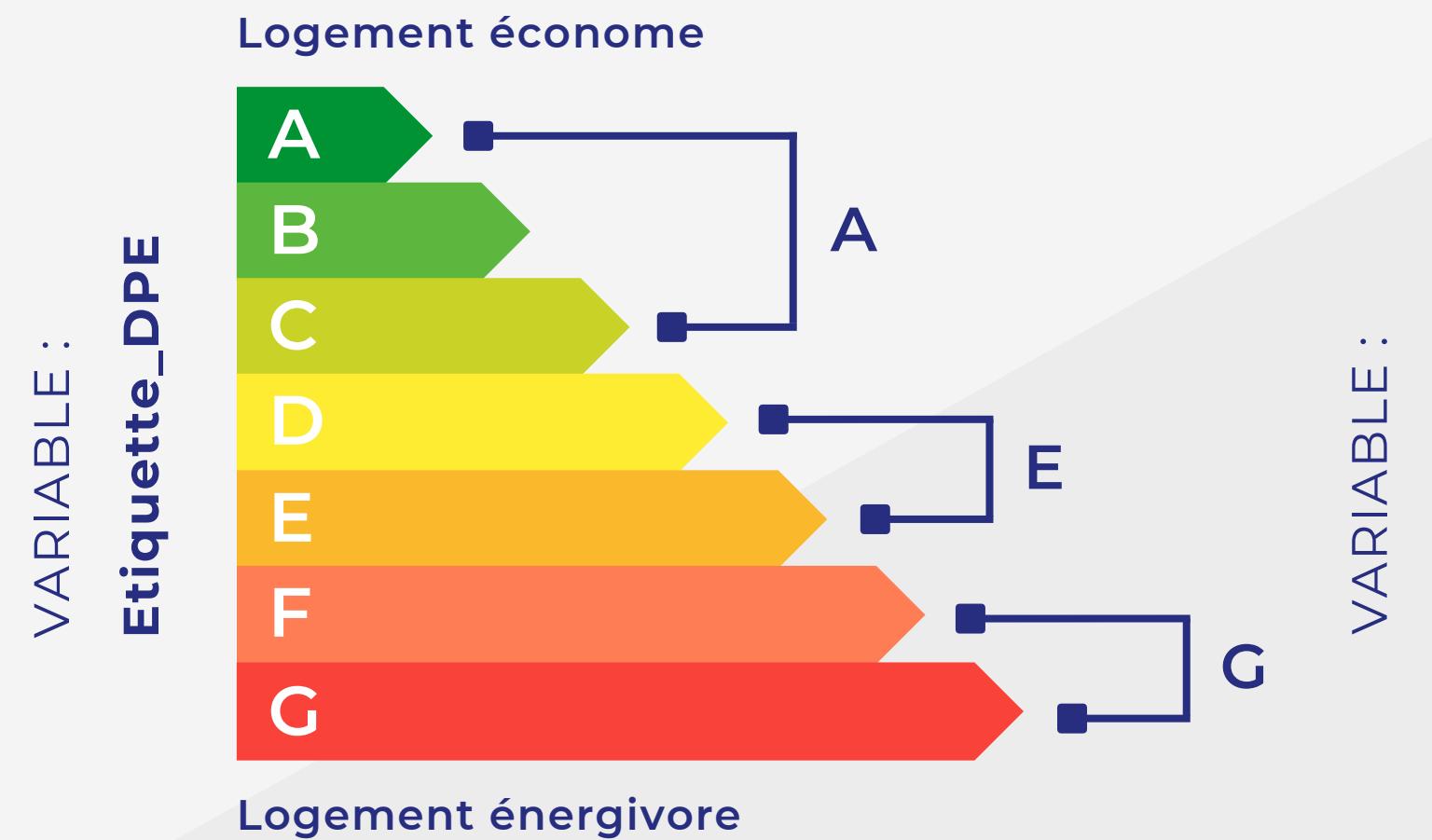
Statistique de Kruskal-Wallis : 17.381
P-value de Kruskal-Wallis : 0.0151

La médiane diffère significativement entre au moins deux groupes.

2.4 Intégration du DPE

Dans cette partie, nous ne gardons que les observations **non manquantes** des étiquettes énergétiques, soit 2056 observations.

Nous avons également **regroupé** les modalités des classes de la manière suivante :



2.4 Intégration du DPE



Regroupement des Etiquettes énergétiques en 3 Classes :

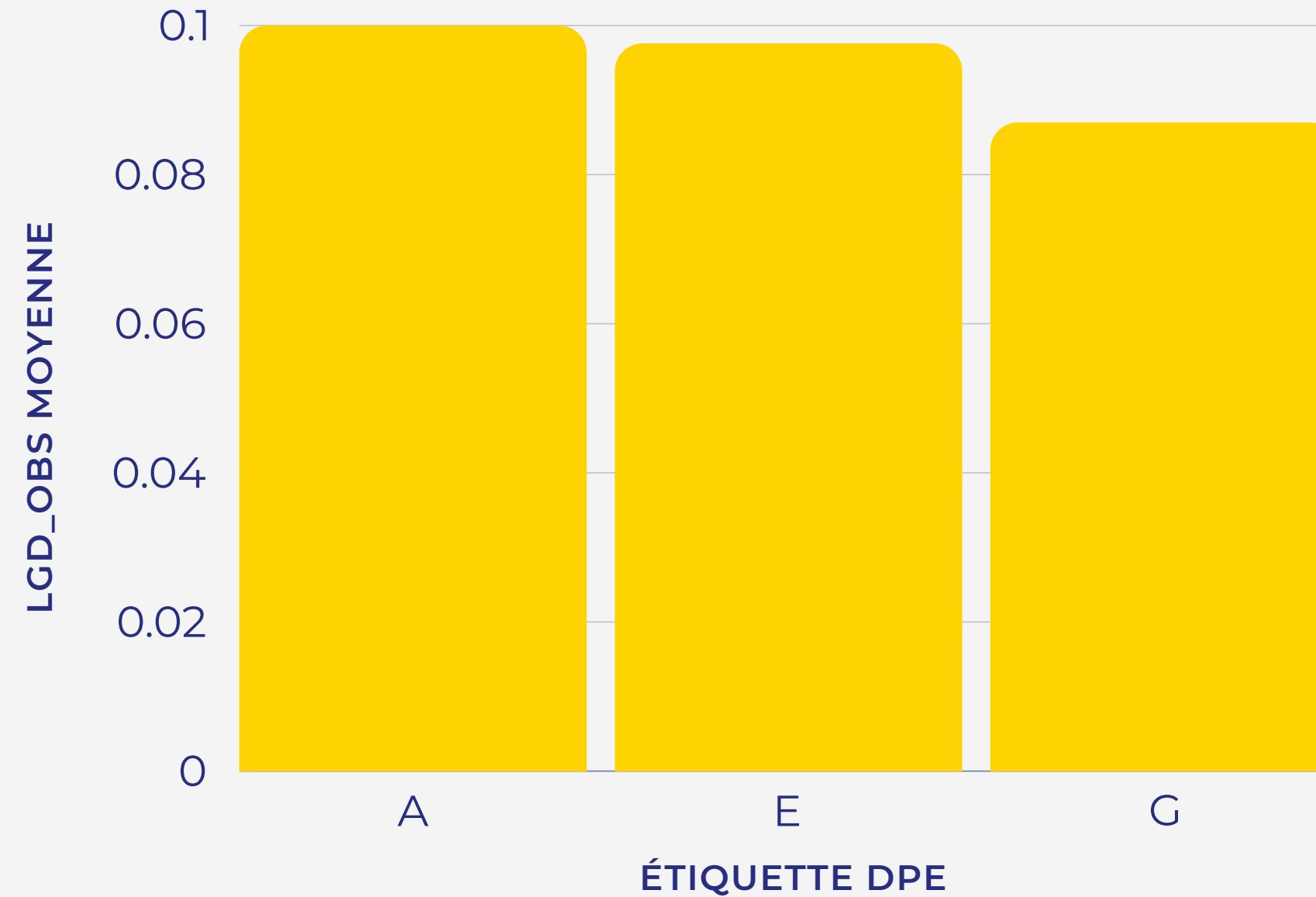
- A : A, B et C
- E : D et E
- G : F et G

Statistique de Kruskal-Wallis : 1.549

P-value de Kruskal-Wallis : 0.461

Il n'y a pas de différence significative dans les médianes entre les groupes.

2.4 Intégration du DPE



Regroupement des Etiquettes énergétiques en 3 Classes :

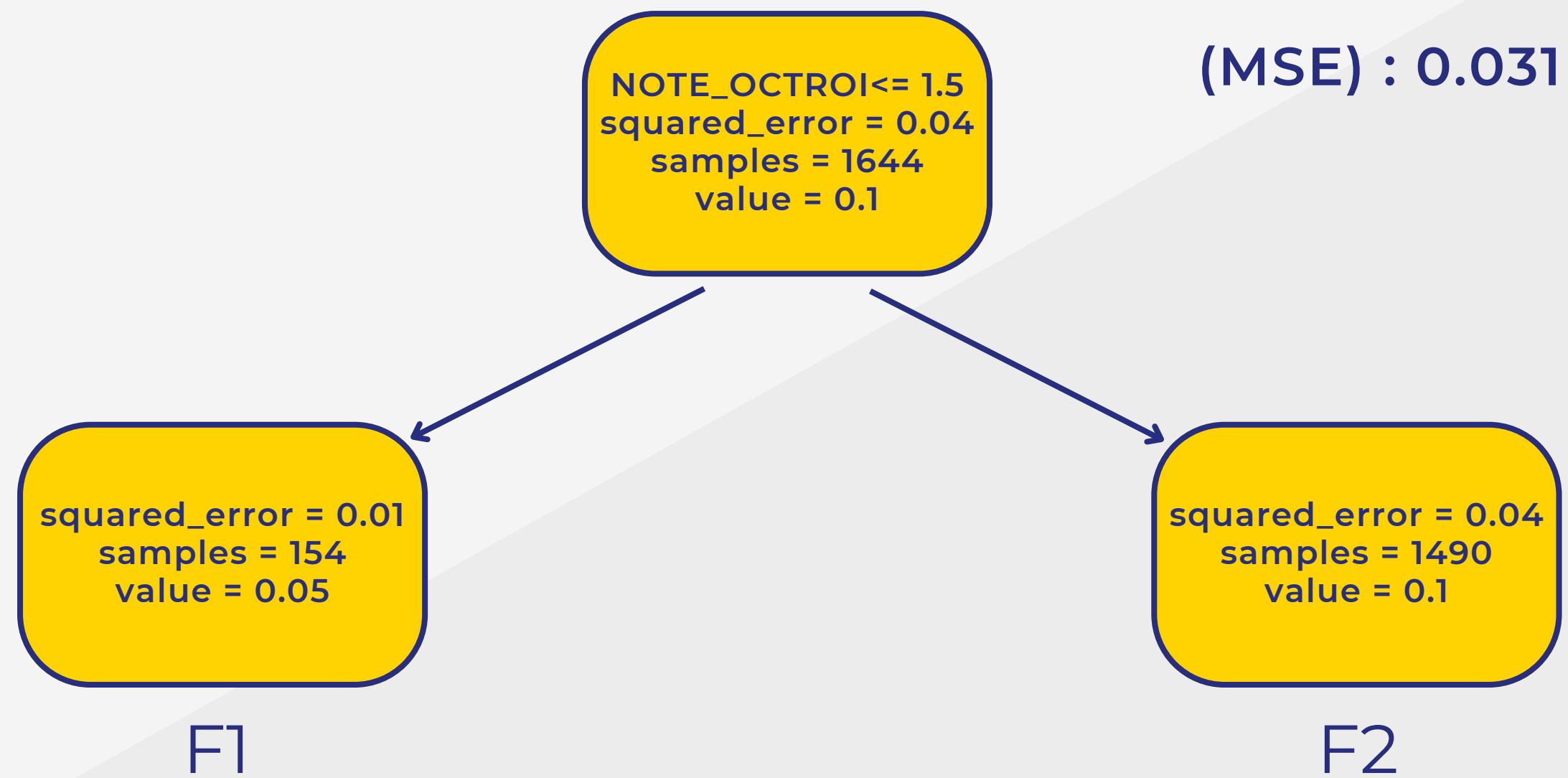
- A : A, B et C
- E : D et E
- G : F et G

Statistique de Kruskal-Wallis : 3.225

P-value de Kruskal-Wallis : 0.199

Il n'y a pas de différence significative dans les médianes entre les groupes.

2.4 Intégration du DPE



2.4 Intégration du DPE



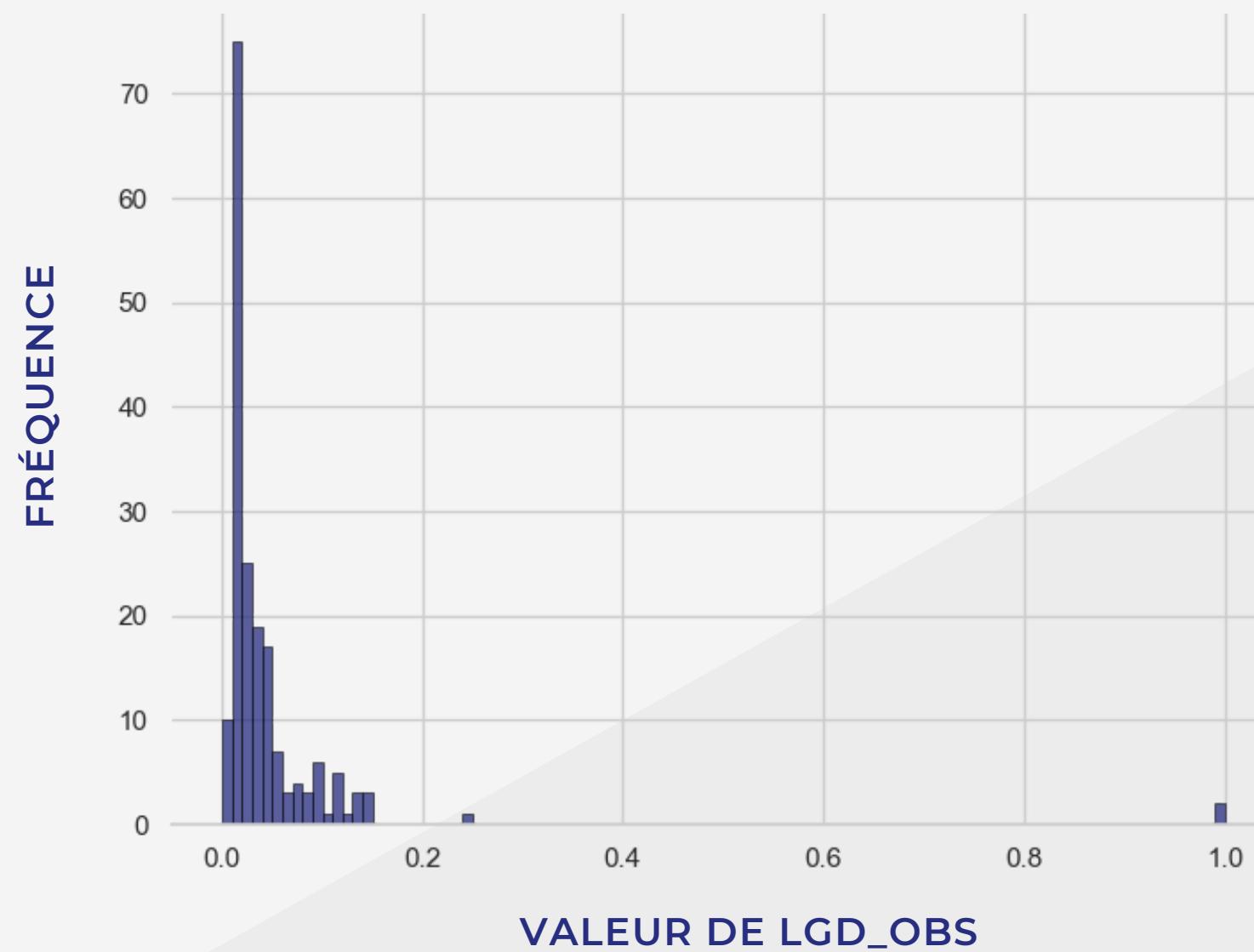
**LGD EN FONCTION DES
ANNÉES POUR CHAQUE
FEUILLE**
STABILITÉ EN RISQUE

L'absence de croisement garantit la stabilité des risques pour nos deux segments de LGD.

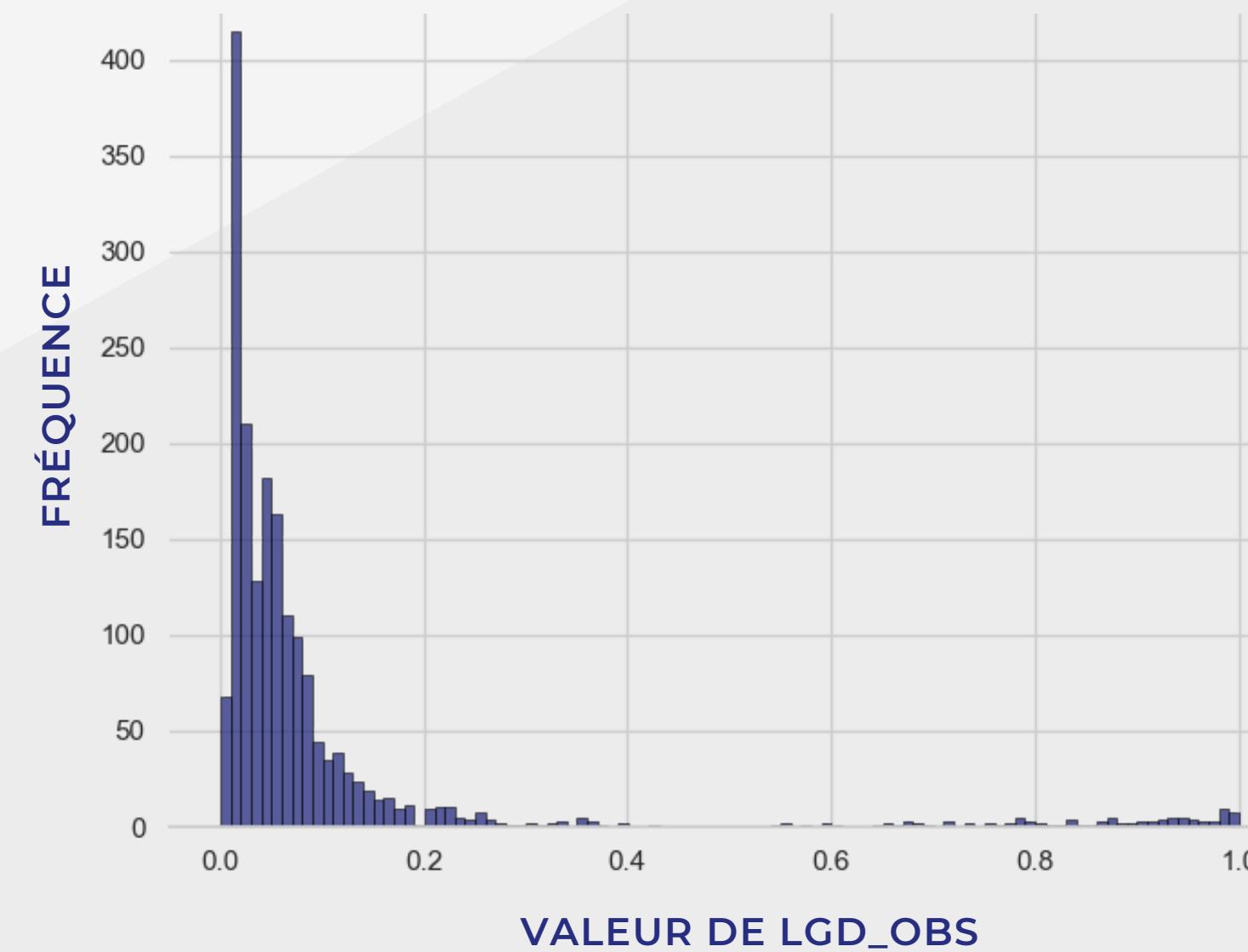
À noter : les courbes sont relativement proches en 2010, 2014 et 2018.

2.4 Intégration du DPE

Distribution de LGD_OBS pour les individus
de la feuille 1

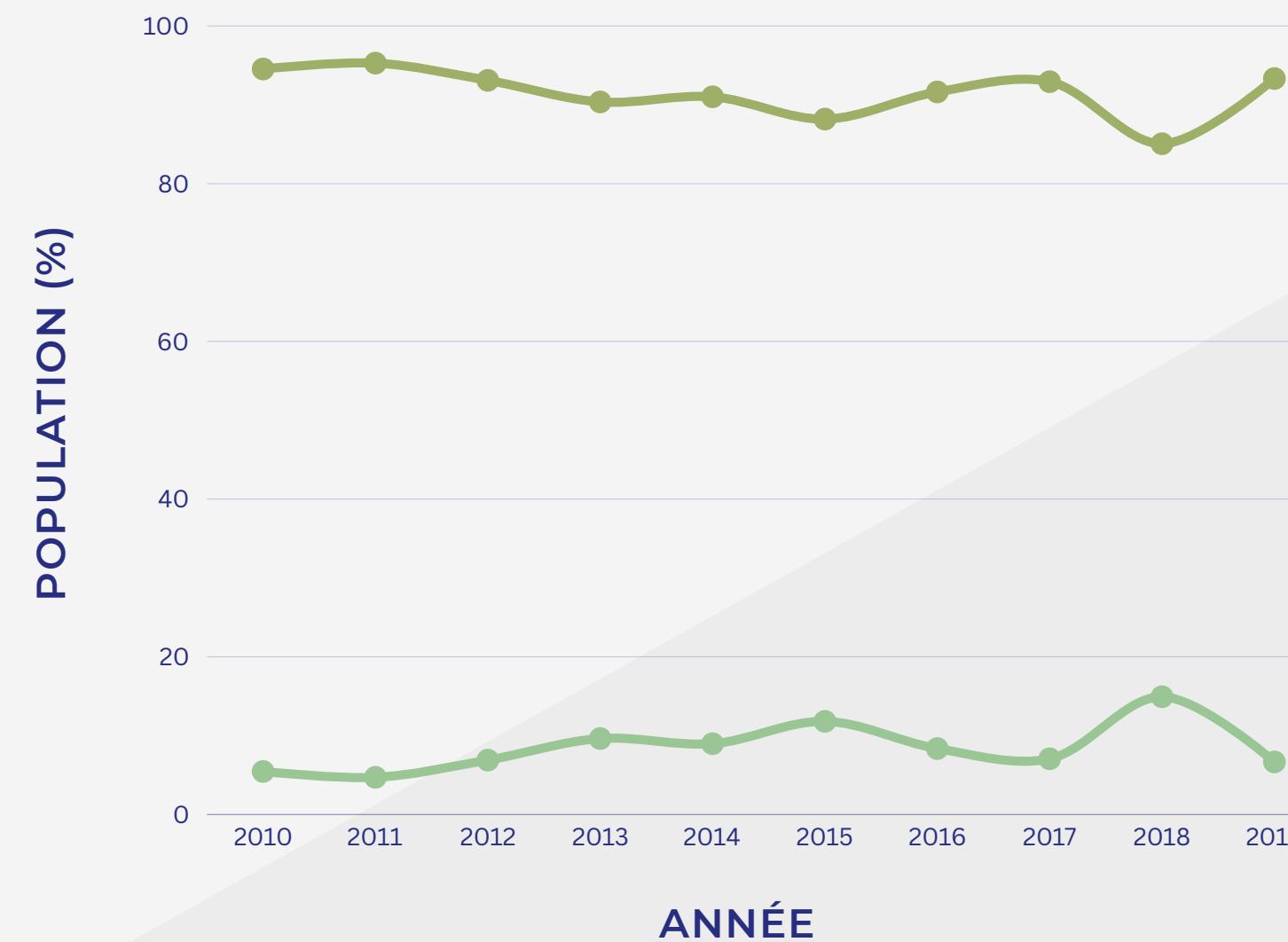


Distribution de LGD_OBS pour les individus
de la feuille 2



2.4 Intégration du DPE

POURCENTAGE DE PERSONNE PAR FEUILLE POUR
CHAQUE ANNÉE

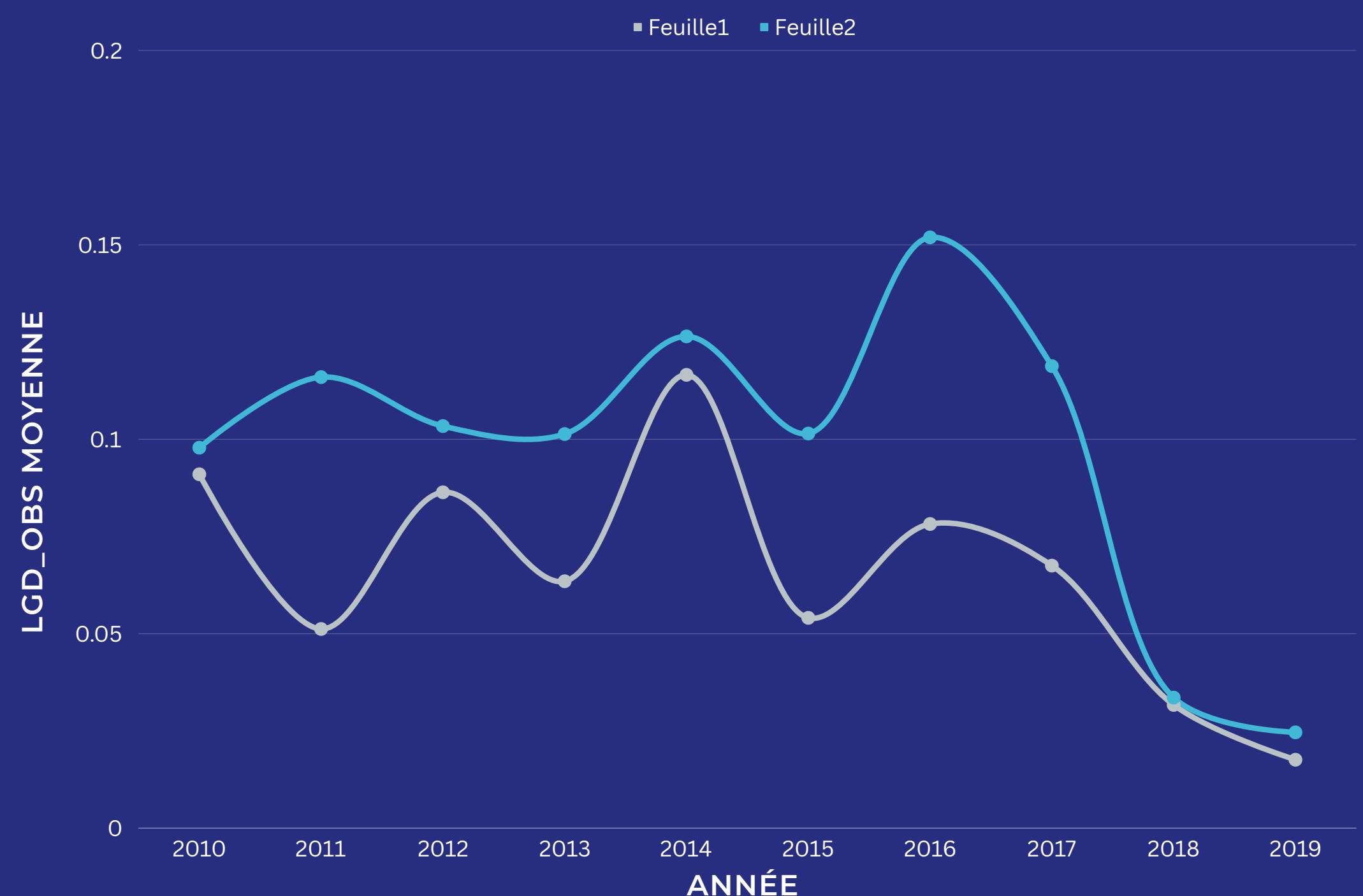


STABILITÉ EN VOLUME

L'absence de croisement assure la stabilité en termes de volume pour nos deux segments de LGD.

À noter : Un déséquilibre significatif entre les deux groupes, avec plus de 90% des individus dans le segment 1.

2.4 Intégration du DPE



STABILITÉ EN RISQUE

L'absence de croisement garantit la stabilité des risques pour nos deux segments de LGD.

À noter : les courbes sont relativement proches en 2010, 2014 et 2018.

3. RÉSULTATS

3.1 LA NOTE D'OCTROI EST UNE VARIABLE DISCRIMINANTE

3.2 STABILITÉ EN VOLUME ET EN RISQUE ASSURÉES

3.3 PAS DE RELATION ENTRE LGD ET LES DPE

3.4 MANQUE DE DONNÉES POUR LES DPE



OUVERTURE

Immobilier : le DPE, un critère observé avec attention par les banques avant d'octroyer un prêt



Romain Rouillard / Crédit photo : GARO / PHANIE / PHANIE VIA AFP • 07h00, le 19 septembre
2023

SOURCE : EUROPE 1

CONTACT



BARRAUD LORENZO

barraudlorenzopro@gmail.com



ENIONA RAPHAËL

eniona.raphael@gmail.com



MIRZA SIMON

simon.mrza@gmail.com