Stratégies d’assurance paramétrique pour la résilience des centrales hydrauliques

Julien ANNETTE

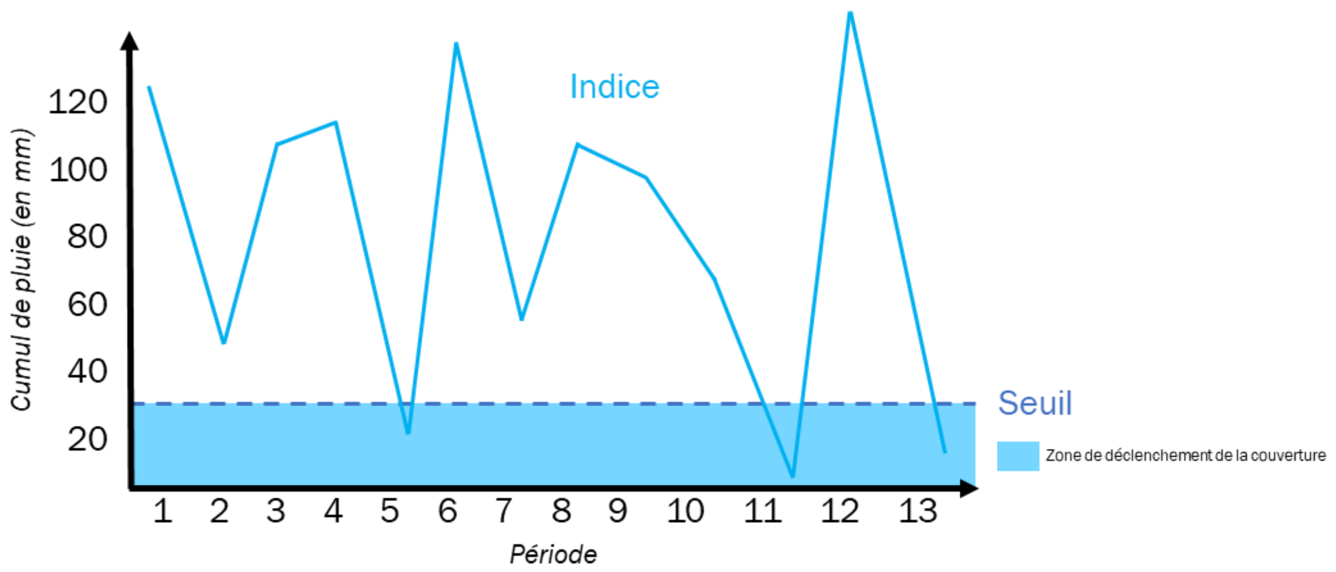
Comment protéger efficacement un producteur hydroélectrique face à l'aléa météorologique ?

SOMMAIRE

1.1 • L'assurance paramétrique : un type d'assurance innovant

- L'assurance paramétrique "couvre un assuré contre la survenance d'un événement spécifique en payant un montant fixe reposant sur l'ampleur de l'événement, plutôt que sur l'ampleur des pertes." *(Source : NAIC)*

- Exemple : protéger ses revenus contre le risque de perte de rendement des récoltes induit par une faible pluviométrie



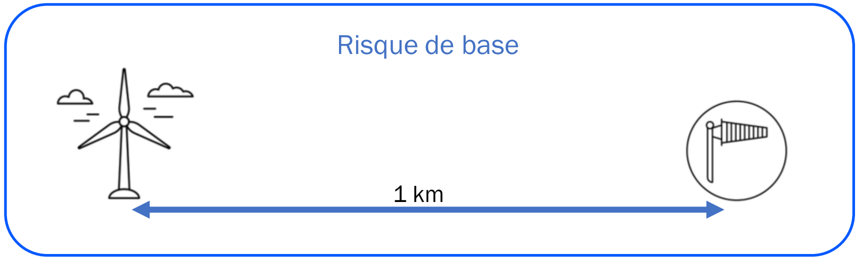
1.2 • L'assurance paramétrique : des enjeux nouveaux

Des avantages

- Des contraintes en matière d'aléa moral et d'antisélection fortement réduites
- Des opportunités de fraude beaucoup moins nombreuses
- Des délais d'indemnisation largement raccourcis

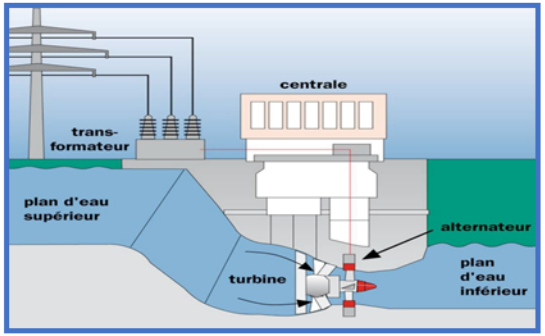
Et des inconvénients

- Un indice parfois difficile à comprendre
- Apparition du risque de base



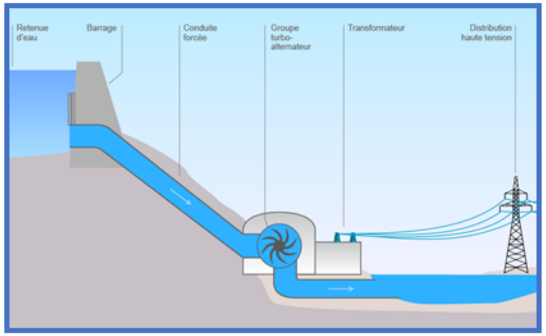
2.1 • L'hydroélectricité : une source d'énergie précieuse et atypique

L'hydroélectricité est la première source d'énergie renouvelable dans le monde. *(Source : Agence Internationale de l'Energie)*



Centrale au fil de l'eau

- Production d'électricité continue
- Pas de possibilité de stockage
- Plus nombreuses mais moins puissantes
- Prise en compte de contraintes environnementales



Centrale de lacs

- Production d'électricité modulable
- Stockage d'énergie constitué par la retenue d'eau
- Investissements importants et contrôles réguliers
- Prise en compte de contraintes environnementales

2.2 • L'hydroélectricité : une source de revenus incertains

En France, en 2022, la production d'énergie hydraulique n'avait jamais connu un niveau aussi faible depuis 1976.

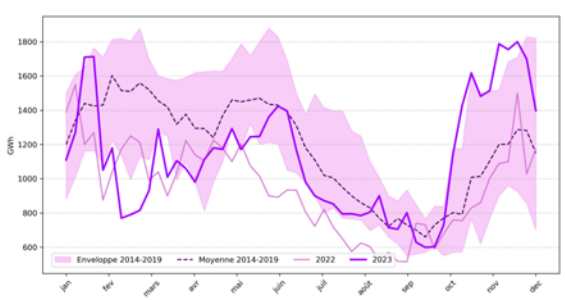
Aléa hydrologique

- Année la plus sèche en France depuis 1959
- Déficit pluviométrique de 25% sur l'ensemble de l'année

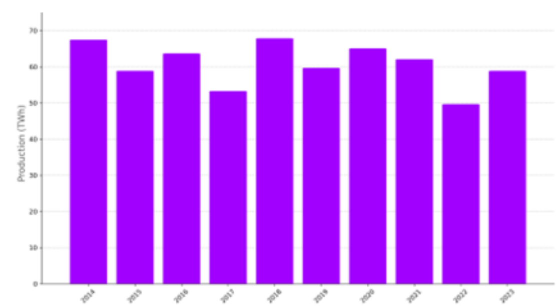
Chute de la production de 20%

Modulabilité de la production

- Gestion responsable des stocks hydrauliques



Aléa de production intra-année



Aléa de production inter-années

2.3 • Les spécificités de la donnée dans l'assurance paramétrique

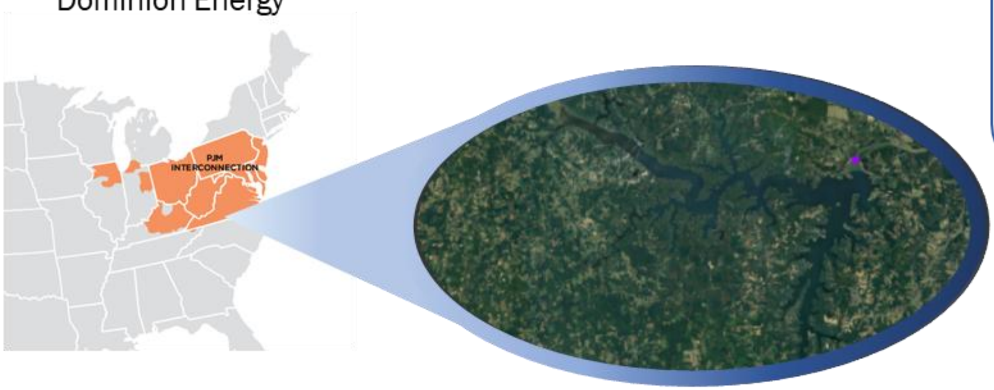
- Données accessibles par les deux parties
- Données mises à jour régulièrement ou disponibles en temps réel
- Données provenant d'un tiers indépendant

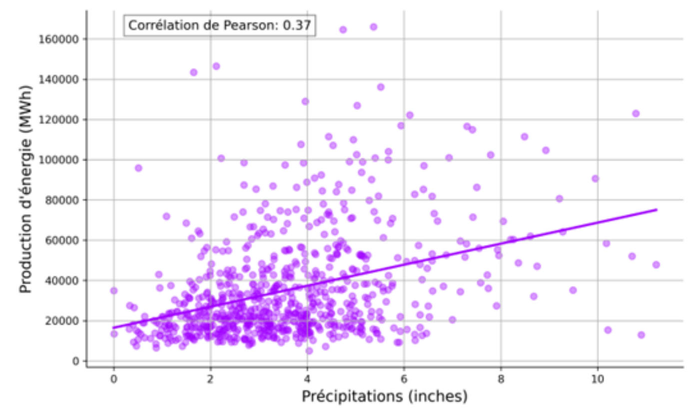
→ Conditions nécessaires à la mise en place d'une assurance paramétrique

- Données fiables, précises et pertinentes → Réduire le risque de base

Les données de nombreuses centrales états-uniennes sont en Open Data, contrairement aux centrales françaises.

Ex : la centrale de John H. Kerr, en Virginie, exploitée par Dominion Energy





Faible corrélation = risque de base accru

Institut des ACTUAIRES

2.4 • Données à disposition

Hydrologie
Afflux d'eau journalier
Reflux d'eau journalier
Niveau journalier du réservoir
Historique de 70 ans

Afflux et reflux d'eau horaires sur un historique de trois jours mis à jour toutes les heures
Données en ft^3/s

Production
Production d'électricité horaire en MWh
Historique de 3 jours

Données pour John H. Kerr

Prix
Prix horaire du MWh
Historique de 18 ans

Demande
Intensité horaire de la demande en électricité en MW
Historique de 18 ans

3.1 • L'afflux d'eau, fortement lié à la production

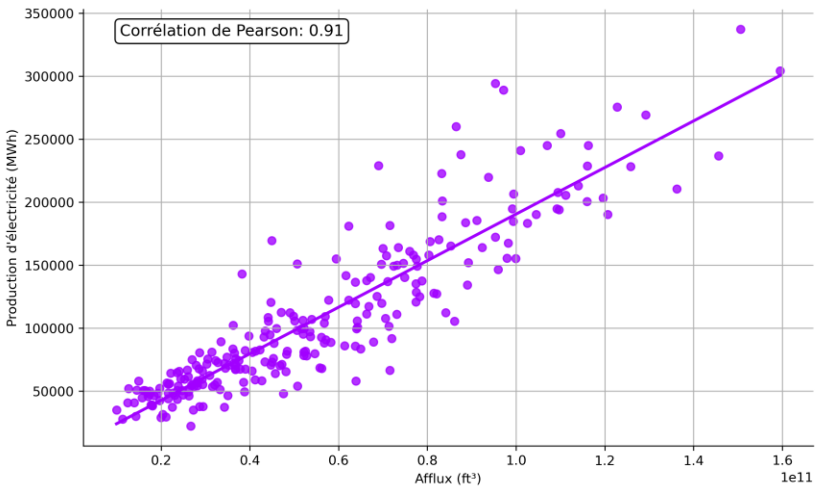
Quelle maille choisir ?

- mensuelle ? ✗
- saisonnière ? ✓
- annuelle ? ✗

Et pourquoi ?

Corrélation forte et données suffisamment nombreuses pour calibrer un modèle

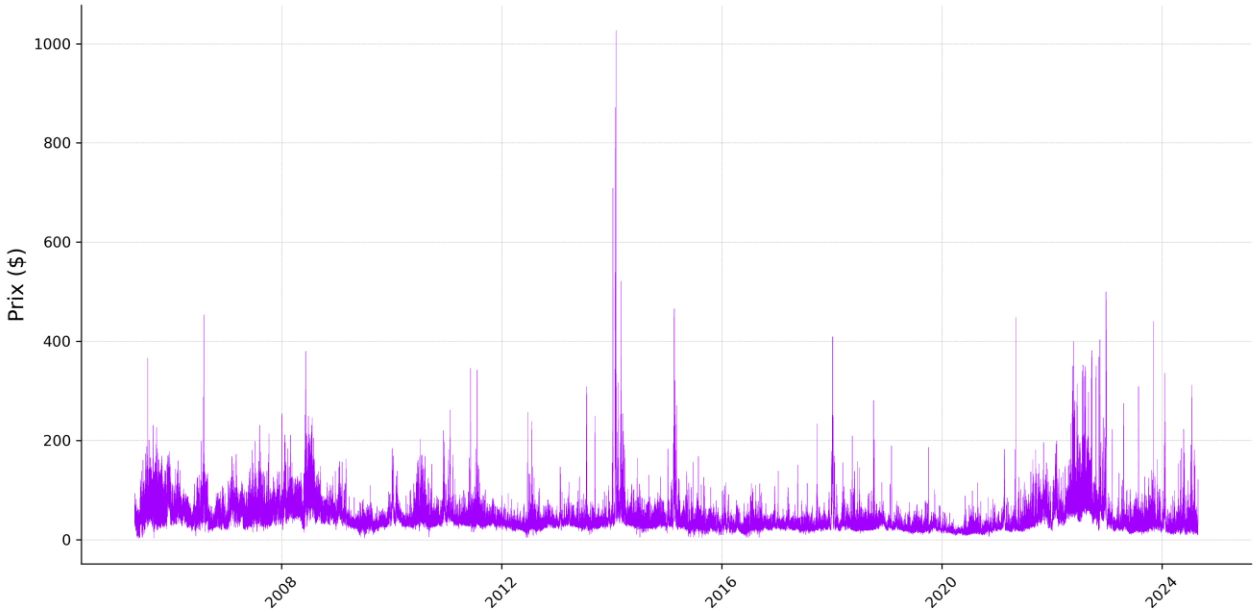
La centrale possède un levier d'action sur la production, mais pas sur l'afflux d'eau



Corrélation entre l'afflux saisonnier et la production saisonnière

3.2 • PJM Interconnection : un marché dérégulé

Le prix horaire du MWh est déterminé par un système d’offre et de demande.



Volatilité horaire du prix du MWh sur l’historique

3.3 • Avant de créer l’indice, calculer les revenus de la centrale

1

Obtention du coefficient énergétique k

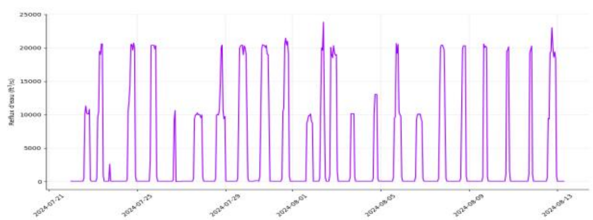
$$G_h = k \cdot r_h$$

G_h la production pour l’heure h

r_h le reflux d’eau pour l’heure h

2

Estimation du reflux horaire



3

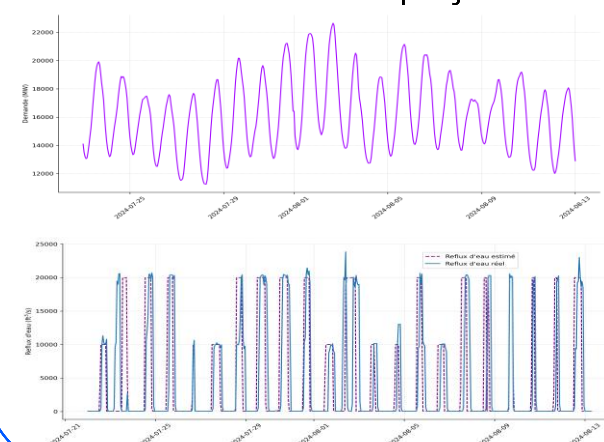
Calcul du nombre d’heures de turbinage

$$N_h(t) = \frac{r \cdot 24}{r_t}$$

r le reflux horaire estimé dans l’étape 2 r_t le reflux d’eau du jour t

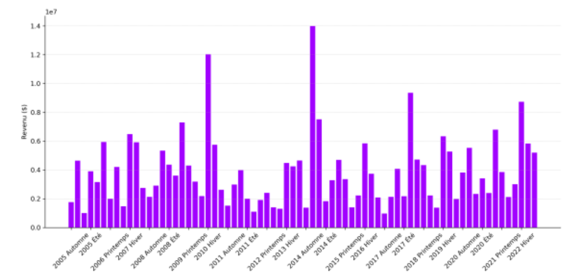
4

Modélisation du reflux d’eau horaire chaque jour



5

Calcul des revenus saisonniers



3.4 • Quel modèle pour obtenir l’indice ?

Formulation du modèle :

$$\ln(Y) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + \varepsilon$$

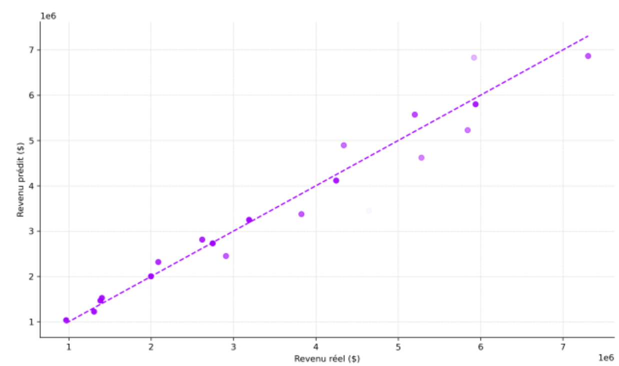
avec :
Y le revenu saisonnier
X_1 l’afflux d’eau saisonnier
X_2 le prix de l’électricité cumulé

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{y_i} \times 100$$

$$\text{SMAPE} = \frac{100}{n} \times \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}$$

Modèle	R^2	MAE	RMSE	sMAPE	MPE	Moyenne	Pearson
log-log	0,954	254 273 $	373 464 $	3,71%	1,05%	3 227 574 $	0,975
linéaire	0,760	577 005 $	693 850 $	12,08%	4,49%	3 216 334 $	0,917
log-linéaire	0,835	512 511 $	739 625 $	7,07%	1,04%	3 014 145 $	0,937
polynomial	0,935	292 285 $	428 241 $	4,56%	-0,06%	3 253 449 $	0,97
Forêt aléatoire	0,953	525 621 $	642 905 $	8,80%	-15,42%	3 557 603 $	0,94
GLM Gamma	-	503 138 $	694 032 $	7.25%	-1,64%	3 100 157 $	0,937

Résultats des métriques appliquées à la base test pour différents modèles
pour une moyenne des revenus réels égale à 3 259 333 $



Indice en fonction du revenu

Institut des ACTUAIRES

4.1 • La méthode séries temporelles : génération

1

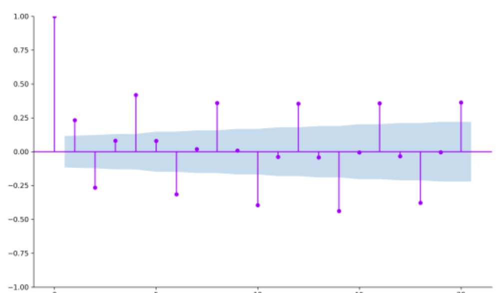
Analyse de la stationnarité des séries temporelles à la maille saisonnière

Test augmenté de Dickey-Fuller et test KPSS

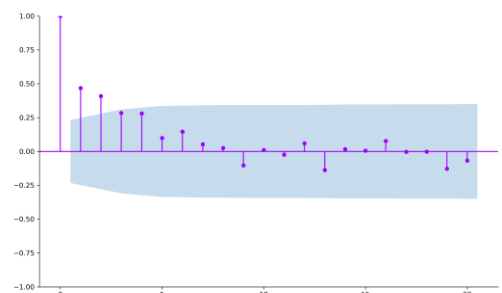
	Afflux		Reflux		Demande		Prix	
	Stat	p-valeur	Stat	p-valeur	Stat	p-valeur	Stat	p-valeur
Test ADF	-4,18	7,11e-04	-3,99	1,48e-03	-5,04	1.84e-05	-4,67	9,19e-05
Test KPSS	0,15	> 0,1	0,22	> 0,1	0,24	> 0,1	0,18	> 0,1

2

Génération de données synthétiques saisonnières



ACF de l'afflux d'eau saisonnier

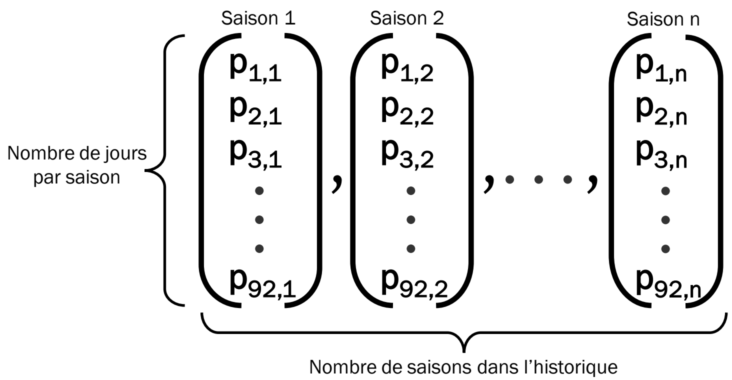


ACF du prix cumulé

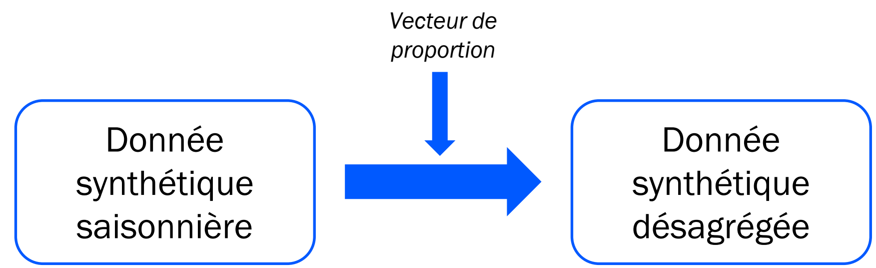
Institut des ACTUAIRES

4.2 • La méthode séries temporelles : désagrégation

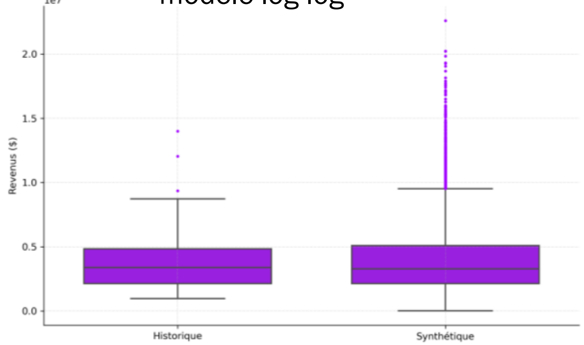
3 Création de vecteurs de proportion à partir de l'historique



4 Désagrégation des données synthétiques à l'aide d'un algorithme des K plus proches voisins



5 Calcul des revenus et calibration du modèle log-log



6 Répétition de la démarche

Base de calibration
Base de tarification
Base de validation

Institut A

5.1 • Les bases de la tarification pour l'assurance paramétrique

Calcul des indemnités

$$I(i_T) = \max(K - i_T, 0)$$

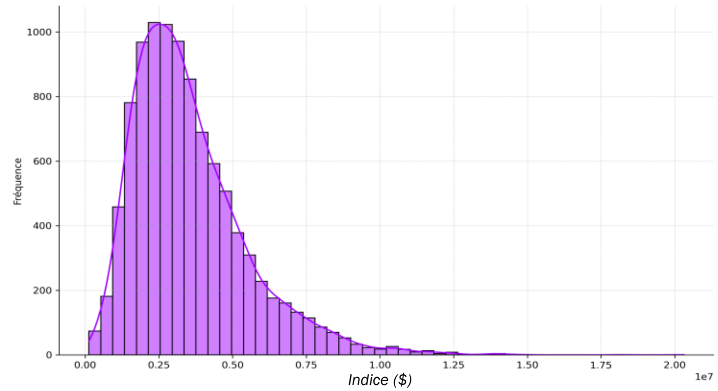
- i_T l'indice sur la période T
- $I(i_T)$ l'indemnité pour la période T
- K le seuil

Principe de prime

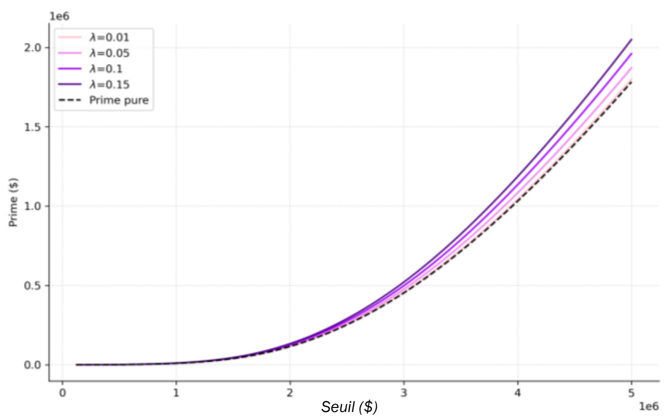
Valeur espérée : $\Pi(X) = (1+\lambda)E[X], \lambda$ un chargement de sécurité

Institut des ACTUAIRES

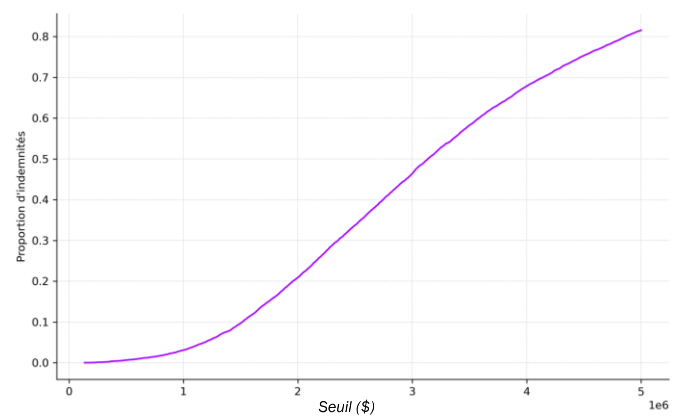
5.2 • Un premier aperçu des propriétés de la tarification



Distribution de l'indice Séries Temporelles (10 000 valeurs)



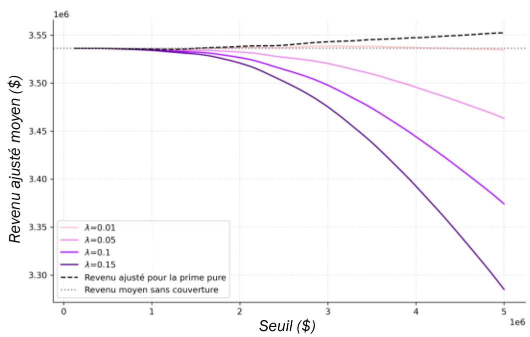
Prime de la valeur espérée via Séries Temporelles



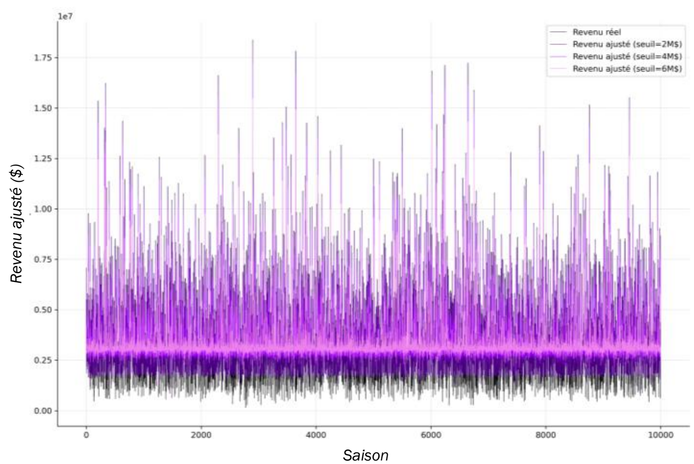
Fréquence d'indemnisation en fonction du seuil

5.3 • Un produit profitabilité pour l'assureur et intéressant pour l'assuré

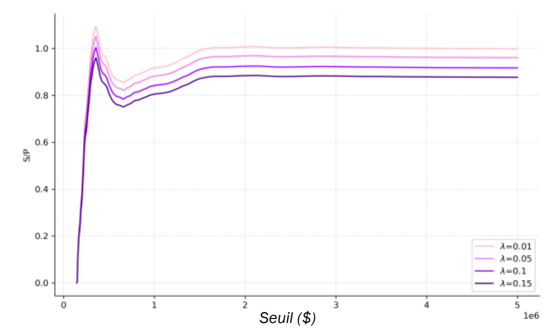
$$\text{Revenu ajusté} = \text{revenu sans couverture} + \text{indemnité} - \text{prime}$$



Revenu moyen ajusté



Volatilité du revenu ajusté



S/P

$\lambda = 10\%$	Revenu moyen ajusté	Volatilité ajustée	Revenu minimal	Revenu ajusté minimal
Seuil de 2M \$	99,8%	92,9%	144 756 \$	693 356 \$
Seuil de 3M \$	98,9%	80,9%	144 756 \$	1 172 813 \$
Seuil de 4M \$	96%	59,7%	144 756 \$	1 522 610 \$
Seuil de 5M \$	92,9%	52,3%	144 756 \$	1 734 012 \$

Impact de l'application du produit par rapport au cas sans couverture, selon plusieurs seuils

Optimiser un seuil en fonction des souhaits de l'assureur et de l'assuré :

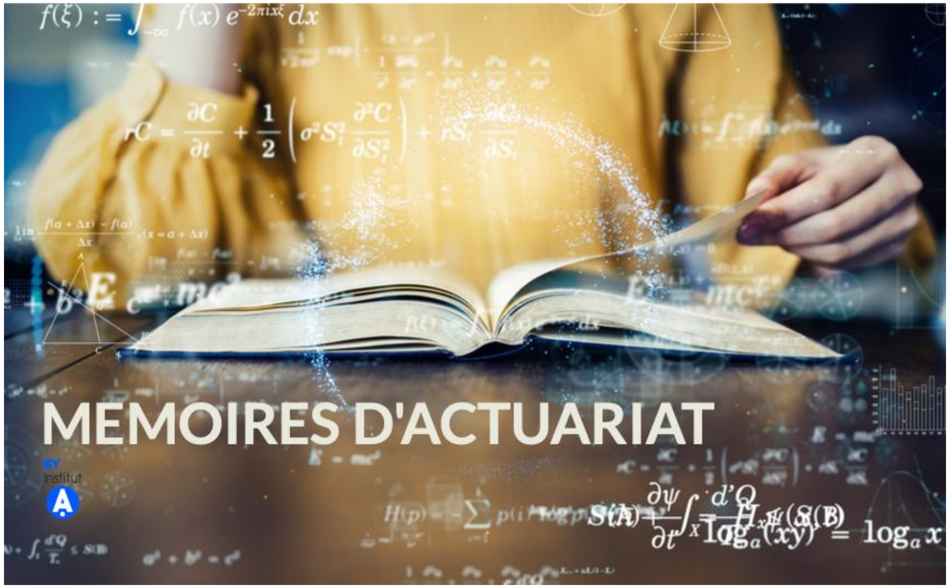
- L'assureur propose un intervalle dans lequel tout seuil lui convient
- L'assuré choisit le seuil optimal dans cet intervalle selon une fonction de coût qui combine moyenne et volatilité du revenu ajusté, par exemple, de la forme :

$$\text{coût} = \alpha \cdot \text{volatilité ajustée} - \beta \cdot \text{revenu ajusté moyen}$$

Paramètre	**Scénario 1**	**Scénario 2**	**Scénario 3**
α	0,3	0,2	0,1
β	1	1	1

$\lambda = 10\%$	**Seuil Optimal ($)**
Scénario 1	5 009 109,25
Scénario 2	3 897 496,22
Scénario 3	2 824 435,96

- Etude non prospective et prise en compte de la déformation du risque
- Les faux positifs et faux négatifs, un risque incontournable
- Un modèle réplicable, mais à quel point ?
- Construction d'un indice fondé sur des projections climatiques
- Construction d'un portefeuille diversifié



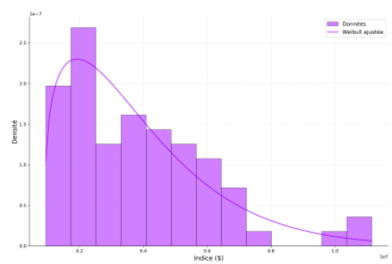
MERCI POUR VOTRE ATTENTION

Annexe

La méthode Monte Carlo : simuler l'indice

1 — Indice construit sur l'historique

2 — Ajustement de la loi Weibull à la distribution de l'indice



3 — Tests de vérification de l'adéquation de l'ajustement

	KS	CvM	AD
Stat	0,577	0,036	0,235
P-val	0,976	0,966	-

4 — Génération de données synthétiques

Annexe

La méthode de désagrégation, plus en détail

3 Création de vecteurs de proportion à partir de l'historique

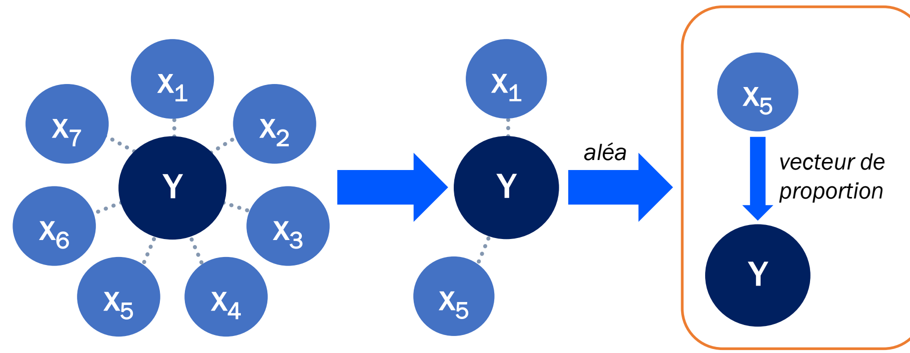
Afflux saisonnier = 10 000 ft^3

1000 ft^3 le jour 1
4000 ft^3 le jour 2
2000 ft^3 le jour 3
1500 ft^3 le jour 4
1500 ft^3 le jour 5

Saison i

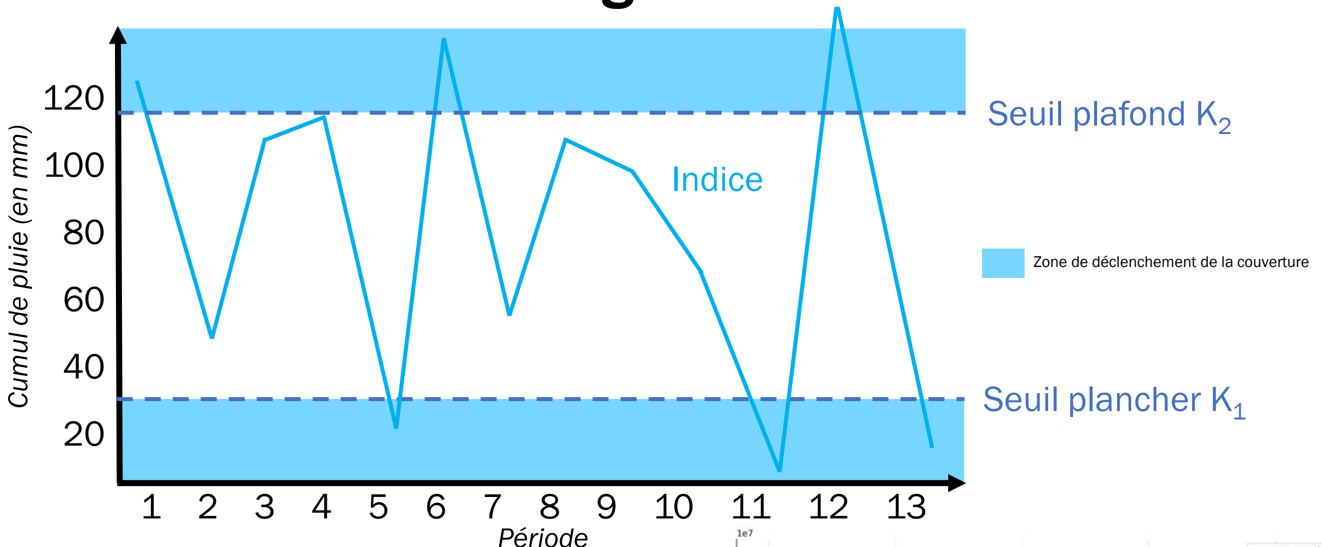
$$\begin{pmatrix} 0,1 \\ 0,4 \\ 0,2 \\ 0,15 \\ 0,15 \end{pmatrix}$$

4 Désagrégation des données synthétiques à l'aide d'un algorithme des K plus proches voisins



Annexe

Le *collar* : une solution de lissage des revenus



Analogie financière :

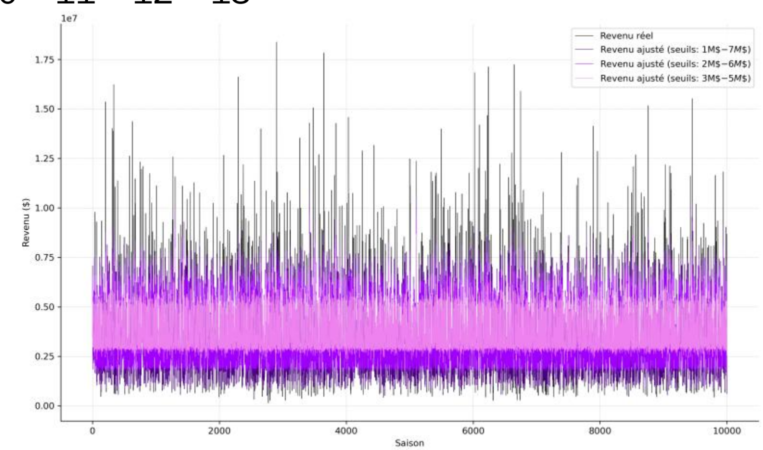
Position longue sur l'actif sous-jacent

+

Achat d'une option de vente

+

Vente d'une option d'achat



Annexe

Quelques chiffres sur le marché de l’assurance paramétrique

Poids du marché de l’assurance paramétrique

11,7 milliards de dollars en 2021

9,9% de croissance annuelle

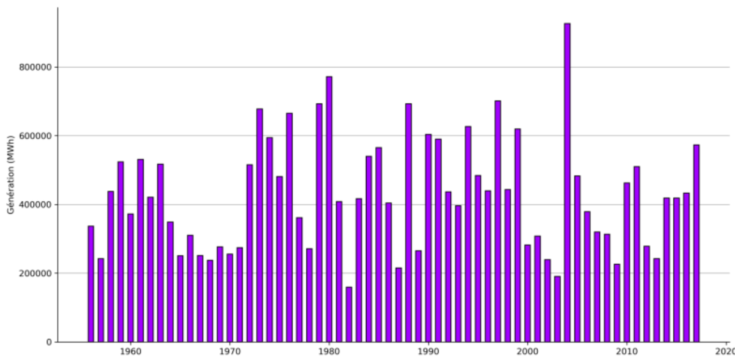
30 milliards de dollars en 2031

Source : Allied Market Reasearch

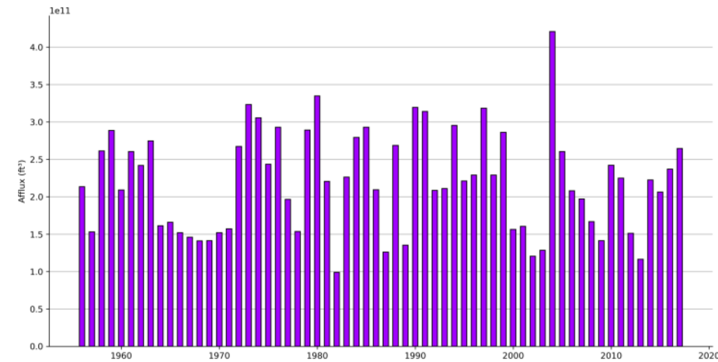
Le marché américain est le premier marché d’assurance paramétrique mondial.

Annexe

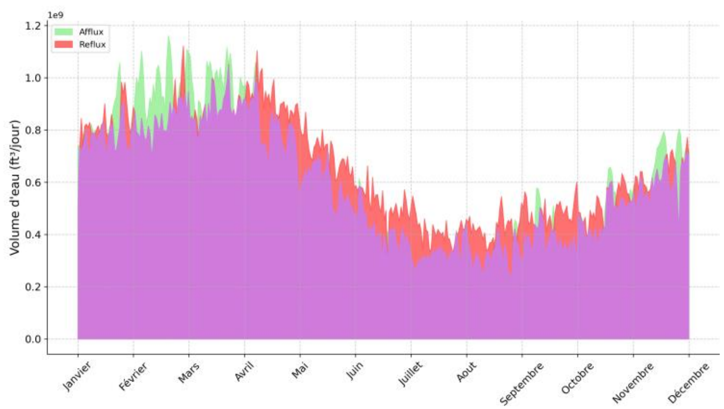
Statistiques sur la centrale de John H. Kerr



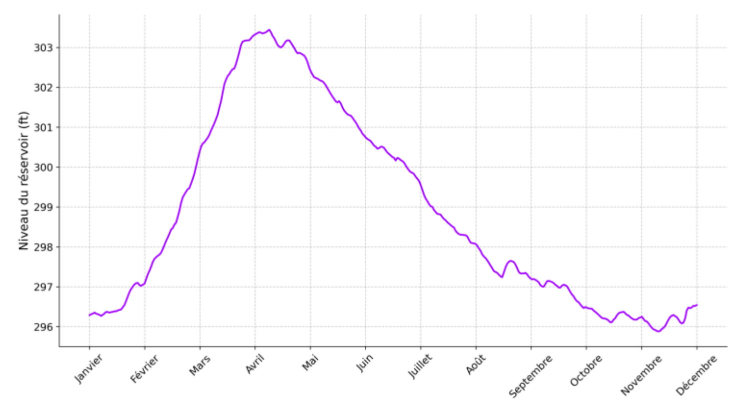
Production annuelle entre 1953 et 2016



Afflux annuel entre 1953 et 2016



Afflux et reflux moyens sur 1953-2022



Evolution du niveau du réservoir sur 1953-2022

Prix de l'électricité

Variation des prix temporelle et spatiale

$$LMP = SEP + CP + MLP$$

avec :

- *LMP : Locational Marginal Pricing*
- *SEP : System Energy Price*
- *CP : Congestion Price*
- *MLP : Marginal Loss Price*

Annexe

Séries temporelles

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \epsilon_t$$

Processus autorégressif

Ajout de termes croisés et de retards saisonniers

$$y_{1,t} = \alpha_1 + \sum_{j=1}^{p} \varphi_{1,j} y_{1,t-j} + \sum_{j=1}^{p} \psi_{1,j} y_{2,t-j} + \sum_{k=1}^{P} \theta_{1,k} y_{1,t-ks} + \sum_{k=1}^{P} \lambda_{1,k} y_{2,t-ks} + \epsilon_{1,t}$$

$$y_{2,t} = \alpha_2 + \sum_{j=1}^{p} \varphi_{2,j} y_{1,t-j} + \sum_{j=1}^{p} \psi_{2,j} y_{2,t-j} + \sum_{k=1}^{P} \theta_{2,k} y_{1,t-ks} + \sum_{k=1}^{P} \lambda_{2,k} y_{2,t-ks} + \epsilon_{2,t}$$

Vecteur autorégressif périodique

s la périodicité *p le*
nombre de retards non-saisonniers *P le*
nombre de retards saisonniers

Annexe

Métriques d’évaluation de la prédiction

On note $\begin{cases} \hat{y}_i \text{ les valeurs prédites} \\ y_i \text{ les valeurs réelles} \\ n \text{ le nombre total d'observations} \end{cases}$

$$\text{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}$$

$$\text{MPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i} \times 100$$

Erreur en pourcentage

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Erreur absolue

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Part de la variance expliquée

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \cdot \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}$$

Relation linéaire entre la prédiction et la variable à prédire

Annexe

Critères d'information

On note

k le nombre de paramètres de la loi L le
$\hat{L}$ maximum de la fonction de vraisemblance n le
nombre d'observations de l'échantillon

Critère d'information d'Akaike

$$\mathrm{AIC} = 2k - 2\ln(\hat{L})$$

Critères à minimiser

Critère d'information bayésien

$$\mathrm{BIC} = k\ln(n) - 2\ln(\hat{L})$$

Annexe

Tests d'adéquation

H_0 : l'échantillon suit la loi théorique considérée H_1 : l'échantillon ne suit pas la loi théorique considérée

Hypothèses de test

On note $F_n(x)$ la fonction de répartition empirique et
$F(x)$ la fonction de répartition théorique

Kolmogorov-Smirnov :

$$D_n = \sup_x |F_n(x) - F(x)|$$

p-valeur > 5% ✓

Cramér-von Mises :

$$W^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, dF(x)$$

p-valeur > 5% ✓

Anderson-Darling:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[\log(F(X_{(i)})) + \log(1 - F(X_{(n+1-i)})) \right]$$

Statistiques de test

Annexe

Analyse des résidus

H_0 : les résidus présentent une autocorrélation H_1 :
les résidus ne présentent pas d'autocorrélation

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

avec e_t le résidu

d proche de 2 indique une absence d'autocorrélation

Test de Durbin-Watson

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

avec $\hat{\rho}_k$ l'autocorrélation au retard k

p-valeur > 5%

Test de Ljung-Box

H_0 : les données suivent une distribution normale H_1 :
les données ne suivent pas une distribution normale

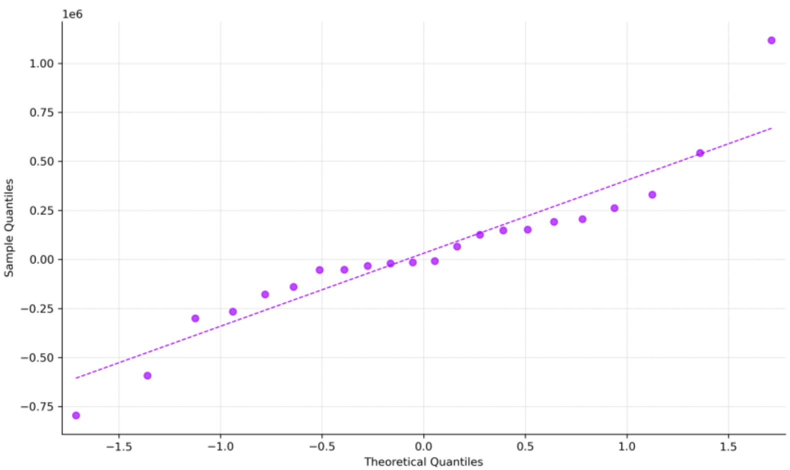
$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

p-valeur > 5%

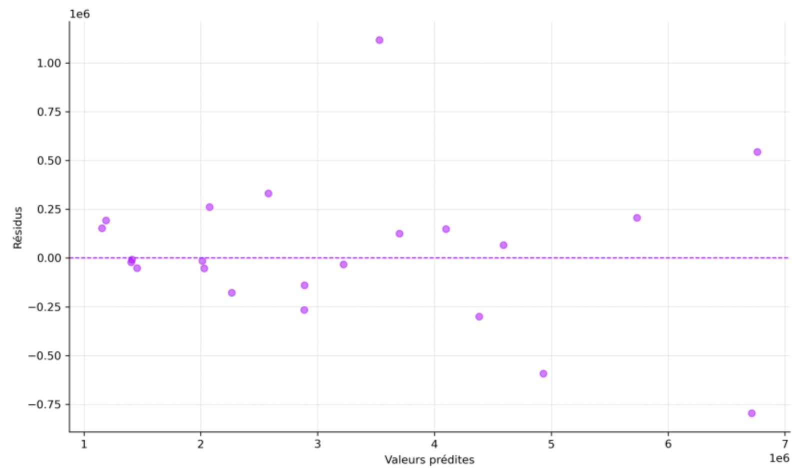
Test de Shapiro-Wilk

Annexe

Analyse des résidus standard



Graphe quantile-quantile des résidus



Graphe residuals vs fitted

Annexe

Stationnarité des séries temporelles

H_0 : la série est stationnaire
H_1 : la série n'est pas stationnaire

Hypothèses du test KPSS

H_0 : la série a une racine unitaire
H_1 : la série n'a pas de racine unitaire

Hypothèses du test ADF

$$y_t = r_t + \beta t + \epsilon_t$$

- *y_t la série temporelle*
- *r_t un processus stationnaire*
- *$\beta.t$ une tendance déterministe*
- *ε_t le terme d'erreur*

Le test KPSS vérifie si la variance de la tendance est nulle.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p - 1 \Delta y_{t-p+1} + \epsilon_t$$

où y_t est la série temporelle étudiée, $\Delta y_t = y_t - y_{t-1}$, δ_i sont les coefficients associés, α est une constante, βt est le coefficient associé à une tendance temporelle, γ est le coefficient d'autorégression, p est le nombre de retards et ϵ_t est l'erreur.

Le test ADF vérifie si γ st nul, signe de la présence d'une racine unitaire.