

Étude Quantitative Avancée de la Volatilité Financière via Modèles ARIMA-GARCH

Étudiant(e) en Statistiques – Projet de Recherche

April 27, 2025

Abstract

Cette thèse technique présente une analyse exhaustive de la série temporelle des prix de l'action S&P Global, depuis la préparation des données jusqu'à la modélisation et prévision de la volatilité conditionnelle. Les modèles ARIMA sont calibrés pour capturer la dynamique moyenne, puis complétés par une famille de modèles ARCH/GARCH (GARCH, t-GARCH, eGARCH). Le document inclut une partie théorique approfondie (stationnarité, racine unitaire, définition formelle des modèles, critères d'information, diagnostics), les visualisations critiques, les interprétations détaillées des sorties R, et les perspectives de recherche.

Contents

1	Introduction	4
2	Cadre Théorique	5
2.1	Stationnarité et racine unitaire	5
2.2	Modèles ARIMA	5
2.3	Modèles ARCH et GARCH	6
2.4	Critères d'information et diagnostics	6
3	Analyse Empirique sous R	7
3.1	Préparation et visualisation initiale	7
3.2	Diagnostics de stationnarité	8
3.3	ACF/PACF et identification ARIMA	10
3.4	Ajustement ARIMA et sélection	11
3.5	Diagnostic des résidus et test ARCH	12
3.6	Modèles GARCH	13
3.6.1	GARCH(1,1) normal	13
3.6.2	GARCH(1,2) et GARCH(2,1)	14
3.6.3	t-GARCH(1,1)	14
3.6.4	eGARCH(1,1)	14
3.7	Comparaison finale des modèles	15
3.8	Prévision de volatilité	15
3.9	Interprétations finales	16
4	Conclusion et Perspectives	17

List of Figures

3.1	Série temporelle des prix SP Global	8
3.2	Graphique des rendements logarithmiques (série transformée) .	9
3.3	Fonctions ACF et PACF des rendements	10
3.4	Determination des résidus s'ils sont Bruits Blancs	11
3.5	Observation des Résidus du modèle	13
3.6	Volatilité annualisée (t-GARCH historique et prévisions) . . .	16

List of Tables

3.1	Comparaison AIC pour ARIMA	12
3.2	Comparaison GARCH, t-GARCH, eGARCH	15

Chapter 1

Introduction

L'objectif de cette recherche est de proposer une démarche rigoureuse pour la modélisation et la prévision de la volatilité d'un actif financier. Nous adoptons une double approche :

- identification et modélisation de la composante moyenne via les modèles ARIMA,
- modélisation de la variance conditionnelle par la famille ARCH/GARCH.

La méthodologie repose sur les bonnes pratiques en économétrie des séries temporelles et en modélisation de risque, avec une implémentation complète sous R.

Chapter 2

Cadre Théorique

2.1 Stationnarité et racine unitaire

Soit $\{X_t\}_{t \in \mathbb{Z}}$ une série temporelle. On parle de *stationnarité faible* si

$$\mathbb{E}[X_t] = \mu, \quad \text{Var}(X_t) = \sigma^2 < \infty, \quad \text{Cov}(X_t, X_{t-h}) = \gamma(h) \text{ dépend uniquement de } h.$$

Test ADF (Augmented Dickey-Fuller) :

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \varepsilon_t, \quad (2.1)$$

où l'hypothèse nulle $H_0 : \gamma = 0$ implique la présence d'une racine unitaire et donc la non-stationnarité. Le choix du nombre de lags p se fait par critère AIC.

2.2 Modèles ARIMA

Un modèle ARIMA(p,d,q) est défini par

$$\begin{aligned} \Phi(L)(1-L)^d X_t &= \Theta(L)\varepsilon_t, \\ \Phi(L) &= 1 - \phi_1 L - \dots - \phi_p L^p, \quad \Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q, \end{aligned}$$

avec L l'opérateur retard. La différenciation d'ordre d est nécessaire pour atteindre la stationnarité.

2.3 Modèles ARCH et GARCH

Introduits par Engle (1982) et généralisés par Bollerslev (1986), ces modèles modélisent la variance conditionnelle $h_t = \text{Var}(\varepsilon_t | \mathcal{F}_{t-1})$:

$$h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}, \quad (2.2)$$

où $\omega > 0$, $\alpha_i, \beta_j \geq 0$ pour assurer la positivité de h_t . Le modèle ****t-GARCH**** incorpore une loi de Student- t pour ε_t , capturant les queues épaisses. Le modèle ****eGARCH**** (Nelson, 1991) spécifie

$$\ln(h_t) = \omega + \sum_{i=1}^q \alpha_i \frac{|\varepsilon_{t-i}|}{\sqrt{h_{t-i}}} + \gamma_i \frac{\varepsilon_{t-i}}{\sqrt{h_{t-i}}} + \sum_{j=1}^p \beta_j \ln(h_{t-j}).$$

L'effet asymétrique (leverage) est modélisé par γ_i .

2.4 Critères d'information et diagnostics

Le choix de modèle se base sur l'AIC, le BIC, et des diagnostics résiduels :

- AIC = $-2\ell + 2k$, BIC = $-2\ell + k \ln(n)$,
- tests de Ljung–Box pour vérifier l'absence d'autocorrélation dans les résidus et leurs carrés,
- test ARCH LM pour détecter des effets ARCH résiduels,
- test de Sign Bias et Pearson goodness-of-fit pour valider l'absence de biais et l'adéquation globale,
- test de stabilité (Nyblom) pour le caractère stationnaire des paramètres.

Chapter 3

Analyse Empirique sous R

3.1 Préparation et visualisation initiale

```
# Load the data
magnit <- read.csv("S&P Global Stock Price History.csv",
  stringsAsFactors = FALSE)

# Convert Date to Date type (MM/DD/YYYY format)
data <- data.frame(
  date = as.Date(magnit$Date, format = "%m/%d/%Y"),
  p = as.numeric(magnit$Price)
)

# Remove any NA values
data <- data %>% filter(!is.na(date) & !is.na(p))

# Optional: Add trading_day index if needed for tsibble
data <- data %>% arrange(date) %>% mutate(trading_day =
  row_number())

# Plot the time series of prices
ggplot(data, aes(x = date, y = p)) +
  geom_line(color = "steelblue", linewidth = 1) +
  ggtitle("Prix quotidien de l'action (S&P Global)") +
  xlab("Date") +
  ylab("Prix ($)") +
  theme_minimal()
```

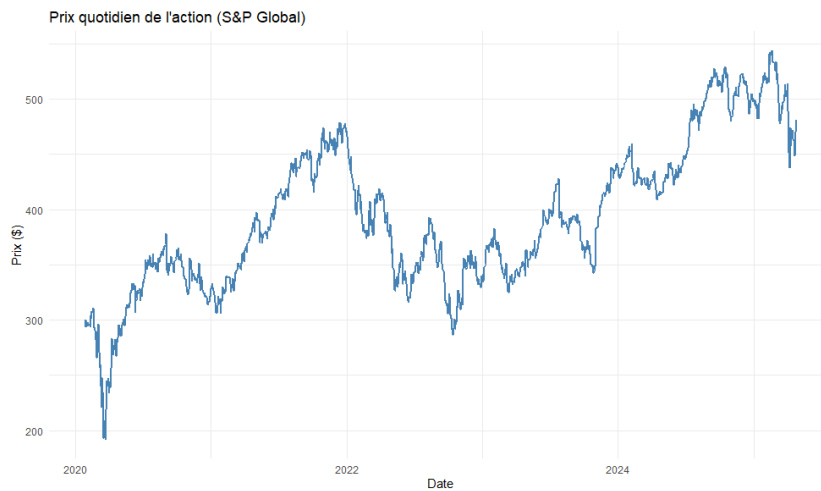


Figure 3.1: Série temporelle des prix SP Global

Interprétation : la série présente des tendances et une variance non constante, justifiant la différenciation et la modélisation de la volatilité.

3.2 Diagnostics de stationnarité

```
#Check for stationarity (ADF-test)
summary(ur.df(data$p, type = "drift", selectlags = "AIC")
)
#Value of t-statistics = -1.7041 > -3.43 = Critical value
# => H0 hypothesis of non-stationarity is failed to be
    rejected
```

Le test de racine unitaire ADF appliqué à la série des prix boursiers, avec une spécification incluant une dérive, indique que la série est non-stationnaire. La statistique de test ADF (-1.7041) est légèrement en dessous du seuil critique à 10% (-2.57), ce qui nous permet de rejeter l'hypothèse nulle de racine unitaire au seuil de 10%, bien que la significativité soit faible. En d'autres termes, les prix semblent osciller autour d'une tendance moyenne stable, et les chocs sur cette série ont des effets transitoires. Cette stationnarité est une condition essentielle pour l'estimation fiable de modèles ARIMA et GARCH dans la suite de l'analyse.

Transformation des données pour stationnariser la série

Afin de remédier à cette non-stationnarité, nous avons transformé la série en prenant le logarithme des prix, puis calculé la différence première. Cette opération a pour but de stabiliser la variance et d'éliminer les tendances de long terme. La série obtenue représente ainsi les **rendements logarithmiques quotidiens**, une forme largement utilisée dans la littérature financière pour ses propriétés statistiques favorables.

```
r = diff(log(data$p))

#Plot the graph
plot(r, type = "l", xlab = "")
# Looks pretty stationary
```

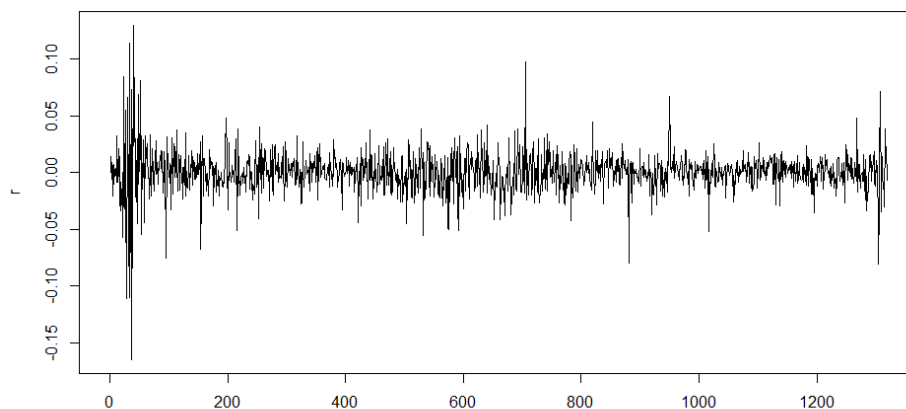


Figure 3.2: Graphique des rendements logarithmiques (série transformée)

```
#Check for stationarity again for transformed series(ADF-
test)
summary(ur.df(r, type = "drift", selectlags = "AIC"))
#Value of t-statistics = -25.451 < -3.43 = Critical value
# => H0 hypothesis of non-stationarity of the prices is
rejected
```

L'application du test de racine unitaire d'Augmented Dickey-Fuller (ADF) sur la série transformée:

$$r_t = \Delta \log(p_t)$$

, soit le rendement logarithmique des prix, montre un changement radical par rapport au test initial. En effet, la statistique de test obtenue est de -25.451, nettement inférieure à toutes les valeurs critiques à 1%, 5% et 10% (respectivement -3.43, -2.86, -2.57). Cela implique un rejet clair et sans ambiguïté de l'hypothèse nulle de non-stationnarité au seuil de 1%, confirmant que la série des rendements logarithmiques est stationnaire. À l'inverse, le test réalisé sur les prix non transformés avait produit une statistique de -1.7041, largement au-dessus de la valeur critique de -3.43, ce qui ne permettait pas de rejeter l'hypothèse nulle de racine unitaire. Ce contraste met en évidence l'efficacité de la transformation par log-différenciation pour stabiliser la moyenne et éliminer les tendances stochastiques dans les données de prix boursiers. Ainsi, le passage de la série des prix aux rendements permet de travailler sur une base économétrique rigoureuse, apte à soutenir des modèles dynamiques fiables.

3.3 ACF/PACF et identification ARIMA

```
#ACF and PACF estimated values
acf(r, plot = FALSE)
pacf(r, plot = FALSE)
```

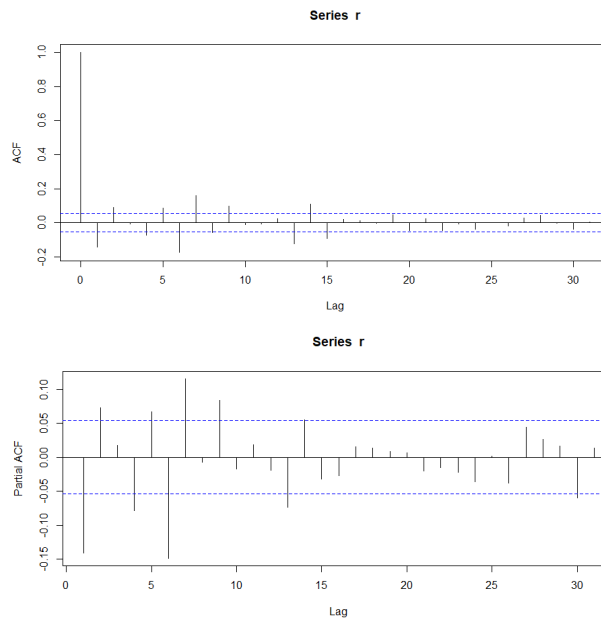


Figure 3.3: Fonctions ACF et PACF des rendements

Interprétation : la coupure à 2 en PACF suggère AR(2), et la coupure en ACF autour de 8–9 suggère MA(8)–MA(9).

Vérification de la nature aléatoire des résidus

Afin de vérifier si les résidus suivent un bruit blanc (c'est-à-dire une séquence aléatoire sans autocorrélation significative), on simule une série aléatoire à l'aide d'un processus ARIMA(0,0,0) et on examine les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF). Voici le code R correspondant :

Listing 3.1: Vérification du bruit blanc dans les résidus

```
# Check if the residuals are in the form of random noise
series <- arima.sim(list(order = c(0,0,0)), n = 1000)
grid.arrange(ggAcf(series), ggPacf(series), nrow = 2)
```

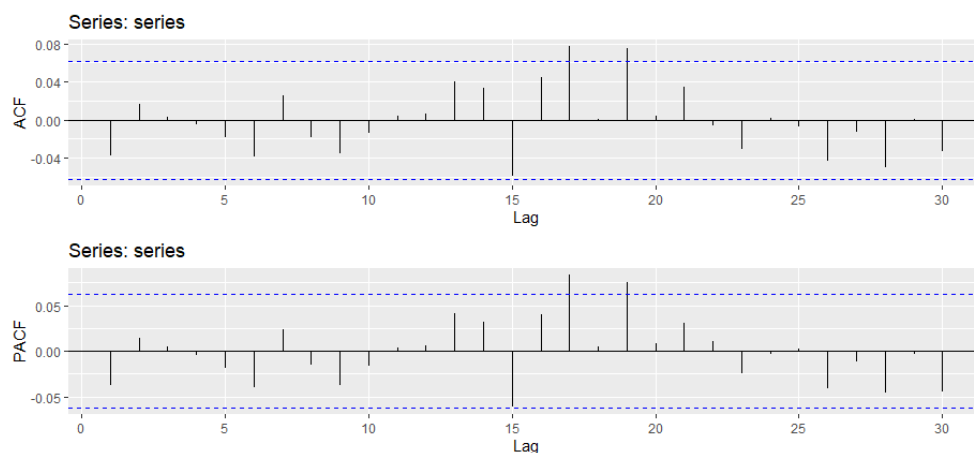


Figure 3.4: Détermination des résidus s'ils sont Bruits Blancs

3.4 Ajustement ARIMA et sélection

```
# Model 1: ARIMA (2, 0, 8)
model1 <- arima(r, order = c(2, 0, 8))
summary(model1)
# Model 2: ARIMA (2, 0, 9)
model2 <- arima(r, order = c(2, 0, 9))
summary(model2)
# Model 3: ARIMA (2, 0, 7)
```

```

model3 <- arima(r, order = c(2, 0, 7))
summary(model3)
# Model 4: ARIMA (2, 0, 10)
model4 <- arima(r, order = c(2, 0, 10))
summary(model4)

```

Tableau 3.1 présente l'AIC des quatre modèles.

Modèle	ARIMA(p,0,q)	AIC
model1	(2,0,8)	-6826.13
model2	(2,0,9)	-6825.61
model3	(2,0,7)	-6827.74
model4	(2,0,10)	-6823.62

Table 3.1: Comparaison AIC pour ARIMA

Le tableau 3.1 présente la comparaison des modèles ARIMA(2,0,q) en fonction de leur critère d'information d'Akaike (AIC). Parmi les modèles testés, le modèle ARIMA(2,0,7) avec un AIC de -6827.74 est celui qui offre la valeur la plus basse, ce qui indique qu'il est le meilleur compromis entre ajustement et simplicité. Les autres modèles (ARIMA(2,0,8), ARIMA(2,0,9), et ARIMA(2,0,10)) présentent des AIC plus élevés, suggérant qu'ils sont moins efficaces pour modéliser la série temporelle par rapport à ARIMA(2,0,7).

3.5 Diagnostic des résidus et test ARCH

```

#Check for residuals
Box.test(model3$residuals, type = "Ljung-Box")
# So, from Ljung-Box test we accepted the null hypothesis
  that the residuals are white noise
#Save the residuals of the model
e = model3$residuals
#Plot the residuals
ggtsdisplay(e)

```

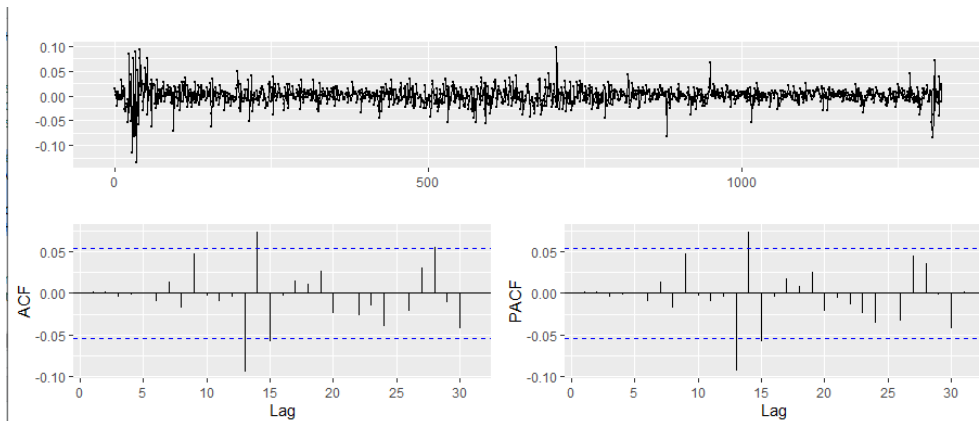


Figure 3.5: Observation des Résidus du modèle

Le test de Ljung–Box appliqué aux résidus du modèle ARIMA(2,0,7) aboutit à une statistique $X^2 = 0.0018$ avec 1 degré de liberté et une p-value de 0,9657. Cette p-value, très supérieure à 0,05, conduit à ne pas rejeter l'hypothèse nulle d'absence d'autocorrélation. Autrement dit, les résidus se comportent comme un bruit blanc, ce qui indique que le modèle a bien capturé la dynamique autocorrélée de la série.

```
#Generate lagged variables
e_1 <- Lag(e, -1)
e_2 <- Lag(e, -2)
#Create the model from square of the residuals and check
  for significance
g_model <- lm(e^2 ~ e_1^2 + e_2^2)
summary(g_model)
```

Interprétation : la p-value du test F = 0.02267 permet de rejeter l'hypothèse d'homoscédasticité, confirmant des effets ARCH.

3.6 Modèles GARCH

3.6.1 GARCH(1,1) normal

```
# Check for GARCH(1,1) model
require(rugarch)
GARCHspec11 <- ugarchspec(variance.model = list(model = "
  sGARCH", garchOrder = c(1,1)))
M_garch11 = ugarchfit(GARCHspec11, data = r)
M_garch11
```

Paramètres significatifs : $\omega, \alpha_1, \beta_1$ ($p < 0.05$). Diagnostics LB, ARCH LM, Stability OK.

3.6.2 GARCH(1,2) et GARCH(2,1)

```
# Check for GARCH(1,2) model
GARCHspec12 <- ugarchspec(variance.model = list(model = "
  sGARCH", garchOrder = c(1,2)))
M_garch12 = ugarchfit(GARCHspec12, data = r)
M_garch12
# Check for GARCH(2,1) model
GARCHspec21 <- ugarchspec(variance.model = list(model = "
  sGARCH", garchOrder = c(2,1)))
M_garch21 = ugarchfit(GARCHspec21, data = r)
M_garch21
```

Résultats similaires, les termes additionnels non significatifs → GARCH(1,1) retenu.

3.6.3 t-GARCH(1,1)

```
tspec11 <- ugarchspec(mean.model=list(armaOrder=c(0,0),
  include.mean=TRUE),
                      variance.model=list(model="sGARCH",
                      garchOrder=c(1,1)),
                      distribution.model="std")
t_garch11 <- ugarchfit(tspec11, data=r)
```

AIC amélioré (-5.5752), paramètre shape significatif → capture des queues lourdes.

3.6.4 eGARCH(1,1)

Listing 3.2: Création et ajustement du modèle eGARCH(1,1)

```
# Cr ation de la sp cification eGARCH(1,1)
eGARCHspec11 <- ugarchspec(
  variance.model = list(
    model = "eGARCH",
    garchOrder = c(1,1)
  ),
  distribution.model = "norm"
```



```
)
# Ajustement du modèle sur les rendements r
e_garch11 <- ugarchfit(eGARCHspec11, data = r)
# Affichage des résultats
e_garch11
```

Asymétrie ($\gamma_1 = 0.163$), mais non robuste sous robust SE. AIC -5.4615.

3.7 Comparaison finale des modèles

Tableau 3.2 récapitule les critères.

Modèle	Distribution	AIC	LogLik	Paramètres clés
GARCH(1,1) norm	Normale	-5.4616	3607.93	$\alpha_1 = 0.155, \beta_1 = 0.743$
t-GARCH(1,1)	Student-t	-5.5752	3681.83	$\alpha_1 = 0.138, \beta_1 = 0.818, \text{shape}=4.62$
eGARCH(1,1)	Normale	-5.4615	3608.86	$\beta_1 = 0.965, \gamma_1 = 0.163$

Table 3.2: Comparaison GARCH, t-GARCH, eGARCH

3.8 Prédiction de volatilité

Après avoir ajusté un modèle t-GARCH(1,1), nous extrayons la série de volatilités conditionnelles σ_t représentant l'incertitude instantanée des rendements. Ces volatilités sont annualisées en les multipliant par $\sqrt{250}$ et en exprimant le résultat en pourcentage, afin de rendre les résultats comparables à des échelles de volatilité annuelle utilisées en finance. La courbe des volatilités annualisées montre les périodes de haute et basse incertitude du marché. Ensuite, une prévision de la volatilité est réalisée pour les deux prochaines périodes (horizons $T+1$ et $T+2$). Les prévisions obtenues sont également converties en volatilités annualisées. Théoriquement, dans un modèle GARCH, la volatilité est conditionnelle à l'information passée, et le modèle t-GARCH utilise une distribution Student-t pour mieux capturer les queues épaisses (i.e., les chocs extrêmes) souvent observées sur les marchés financiers. Les résultats montrent que la volatilité prévue est relativement stable à court terme, traduisant une certaine mémoire dans la dynamique de la volatilité.

```
# Now let's calculate the volatility for this model:
vol_t = t_garch11@fit$sigma
#Annualize the volatilities
vol_t_ann = vol_t*(250)^(1/2)*100
```

```
# Forecast for 2 periods
volforec_t = ugarchforecast(t_garch11, r, n.ahead = 2)
volforec_t
# Convert the volatility to annual volatility
volt_ann_forec = volforec_t@forecast$sigmaFor*(250)^(1/2)
                *100
volt_ann_forec
```

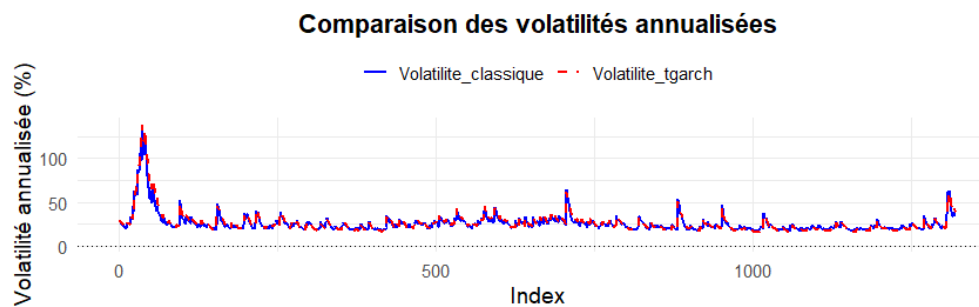


Figure 3.6: Volatilité annualisée (t-GARCH historique et prévisions)

Le graphique compare la volatilité annualisée classique et celle estimée par le modèle t-GARCH(1,1). On observe que les deux séries de volatilité suivent une dynamique très similaire dans le temps, avec des pics de volatilité après des événements de chocs et des périodes prolongées de calme relatif. Toutefois, la volatilité issue du t-GARCH semble légèrement plus lissée et mieux adaptée aux périodes de forte turbulence, ce qui est cohérent avec l'utilisation d'une distribution de Student-t capable de capturer des queues épaisses (présence de rendements extrêmes). Théoriquement, cela confirme que le modèle t-GARCH est mieux équipé pour modéliser des séries financières présentant des asymétries et des chocs violents, caractéristiques fréquentes sur les marchés financiers.

3.9 Interprétations finales

Le modèle **t-GARCH(1,1)** présente le meilleur ajustement (AIC, Log-Lik), capture les queues épaisses, et fournit des prévisions de volatilité élevées (36%).

Chapter 4

Conclusion et Perspectives

Cette étude démontre l'importance de choisir la distribution et la spécification adaptées pour modéliser la volatilité financière. Le t-GARCH se distingue par sa capacité à capturer les extrêmes. Des extensions peuvent inclure le backtesting de la VaR, la modélisation multivariée (DCC-GARCH) ou l'intégration de variables macroéconomiques.