

Analyse et Modélisation SARIMA des Prix des Actions Apple (AAPL)

Période 1998-2025

[Senhaji Ahmed - Soufiane Oukessou]

28 mai 2025

Table des matières

1	Introduction	3
2	Analyse Descriptive des Données	3
2.1	Présentation des Données	3
2.2	Évolution Temporelle	3
2.3	Analyse de la Volatilité	4
3	Décomposition de la Série Temporelle	5
3.1	Décomposition Additive	5
4	Tests de Stationnarité	6
4.1	Test de Dickey-Fuller Augmenté sur la Série Originale	6
4.2	Transformation Logarithmique et Différenciation	6
4.3	Test ADF sur la Série Log-Différenciée	6
5	Identification du Modèle SARIMA	8
5.1	Analyse des Fonctions d'Autocorrélation	8
5.2	Justification des Paramètres SARIMA(2,1,1)(0,1,0) ₁₂	8
5.2.1	Partie Non-Saisonnière : ARIMA(2,1,1)	8
5.2.2	Partie Saisonnière : (0,1,0) ₁₂	9
6	Estimation et Validation du Modèle	9
6.1	Méthode de Validation Croisée	9
6.2	Résultats de l'Estimation	9
6.3	Graphiques de Validation	10
7	Prévisions et Intervalles de Confiance	11
7.1	Prévisions sur 30 Mois	11

8	Discussion et Limitations	11
8.1	Points Forts du Modèle	11
8.2	Limitations Identifiées	11
8.2.1	Hétéroscédasticité	11
8.2.2	Changements Structurels	11
9	Conclusion	12

1 Introduction

Cette étude présente une analyse complète des prix des actions Apple (AAPL) sur la période 1998-2025 en utilisant la méthodologie SARIMA (Seasonal AutoRegressive Integrated Moving Average). L'objectif principal est de modéliser la série temporelle des prix de clôture mensuels et d'évaluer la capacité prédictive du modèle $SARIMA(2,1,1)(0,1,0)_{12}$.

Les données analysées couvrent une période de plus de 20 ans, incluant plusieurs événements économiques majeurs qui ont impacté les marchés financiers, notamment la bulle internet (2000-2002), la crise financière de 2007-2008, et le krach boursier de 2020.

2 Analyse Descriptive des Données

2.1 Présentation des Données

Les données utilisées proviennent des prix historiques des actions Apple, avec un focus sur les prix de clôture mensuels. La série temporelle présente les caractéristiques suivantes :

- Période d'analyse : 1998-2025
- Fréquence : Mensuelle
- Variable d'intérêt : Prix de clôture (Close)

2.2 Évolution Temporelle

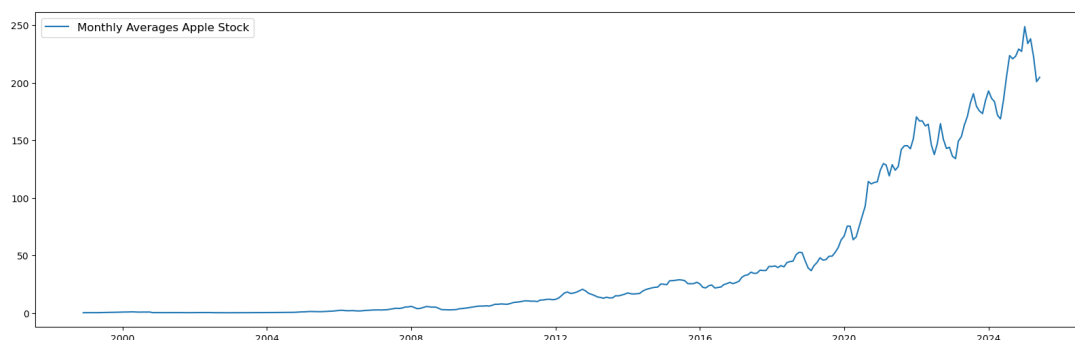


FIGURE 1 – Évolution des prix de clôture mensuels d'Apple (1998-2025)

L'analyse de l'évolution temporelle révèle :

- Une croissance quasi-exponentielle à partir de 2009
- Une période de relative stabilité entre 2000 et 2008
- Une forte accélération de la croissance après 2012
- Une volatilité croissante en fin de période

2.3 Analyse de la Volatilité

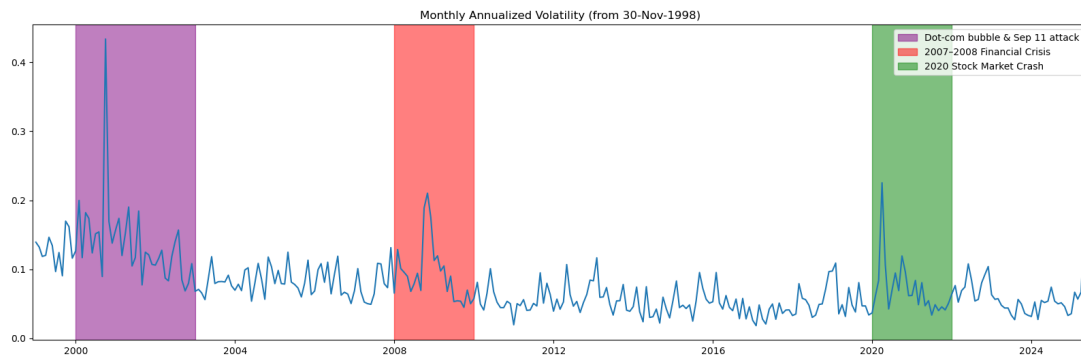


FIGURE 2 – Volatilité mensuelle annualisée avec événements majeurs

L'analyse de la volatilité mensuelle annualisée met en évidence trois périodes de haute volatilité correspondant aux crises économiques majeures :

- Bulle internet et attentats du 11 septembre (2000-2002)
- Crise financière mondiale (2007-2009)
- Krach boursier de 2020 (début de période)

3 Décomposition de la Série Temporelle

3.1 Décomposition Additive

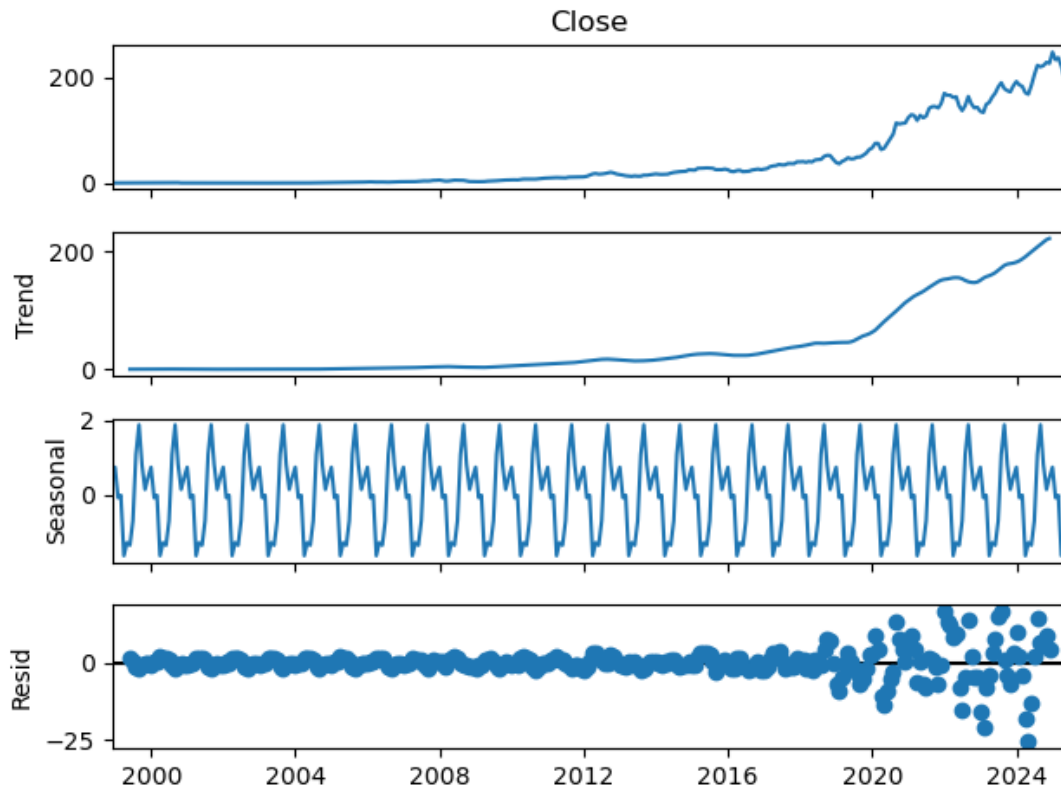


FIGURE 3 – Décomposition de la série temporelle (Tendance, Saisonnalité, Résidus)

La décomposition de la série révèle :

Tendance :

- Forme exponentielle caractéristique
- Croissance accélérée après 2012
- Justifie l'utilisation d'une transformation logarithmique

Composante Saisonnière :

- Cycles réguliers d'amplitude constante
- Stabilité sur toute la période d'observation
- Justifie le paramètre de différenciation saisonnière $D=1$

Résidus :

- Hétéroscédasticité évidente (variance non-constante)
- Augmentation significative de la variance après 2025
- Suggère des limitations du modèle SARIMA classique

4 Tests de Stationnarité

4.1 Test de Dickey-Fuller Augmenté sur la Série Originale

Les résultats du test ADF sur la série originale confirment la non-stationnarité :

- Statistique ADF : -2.607416
- p-value : 0.091474 > 0.05
- Conclusion : Série non-stationnaire (présence d'une tendance)

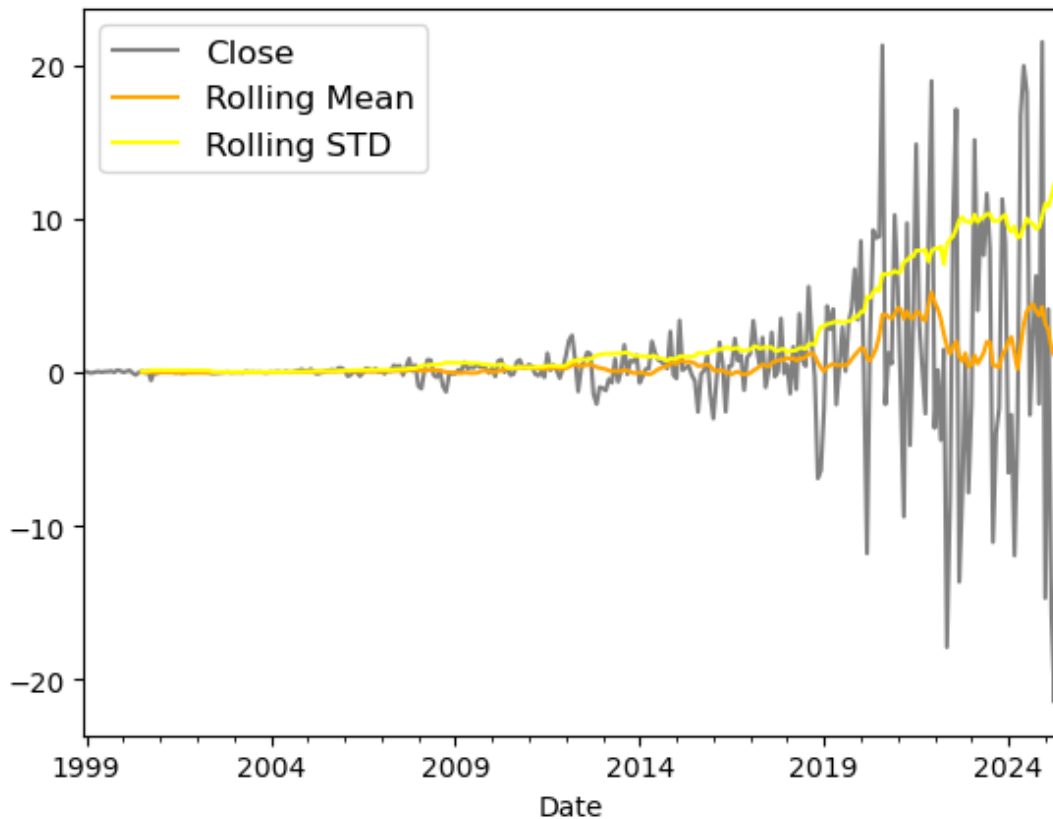


FIGURE 4 – Série après différenciation simple : toujours non-stationnaire

4.2 Transformation Logarithmique et Différenciation

La simple différenciation $X_t - X_{t-1}$ n'a pas suffi à rendre la série stationnaire : les tendances persistantes et la variance non-constante subsistaient. Par conséquent, une transformation logarithmique a été appliquée avant différenciation pour stabiliser la variance :

1. **Transformation logarithmique** : $\log(Close_t)$
2. **Différenciation première du log** : $\Delta \log(Close_t) = \log(Close_t) - \log(Close_{t-1})$

4.3 Test ADF sur la Série Log-Différenciée

Après cette transformation double, la série devient stationnaire, comme l'indique le test de Dickey-Fuller :

- Statistique ADF : -9.913556
- p-value : $0.000000 < 0.05$
- Conclusion : Stationnarité atteinte après transformation logarithmique et différenciation

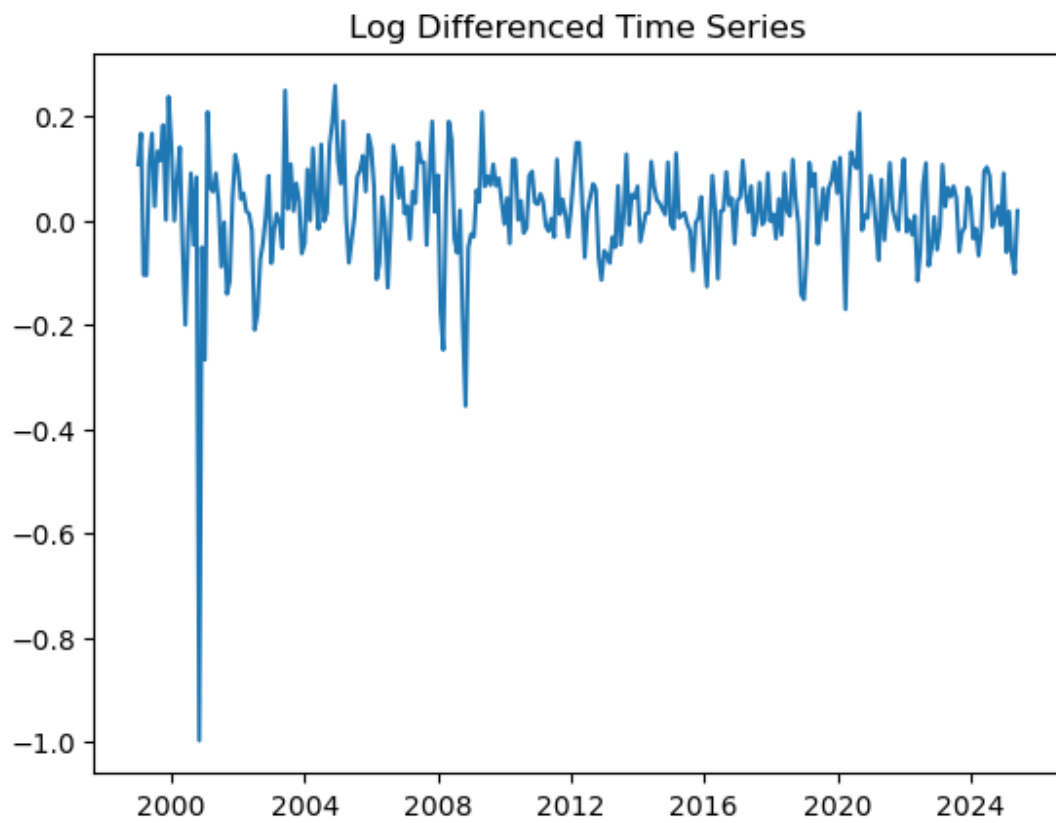


FIGURE 5 – Série log-différenciée : stationnaire (ADF significatif)

5 Identification du Modèle SARIMA

5.1 Analyse des Fonctions d'Autocorrélation

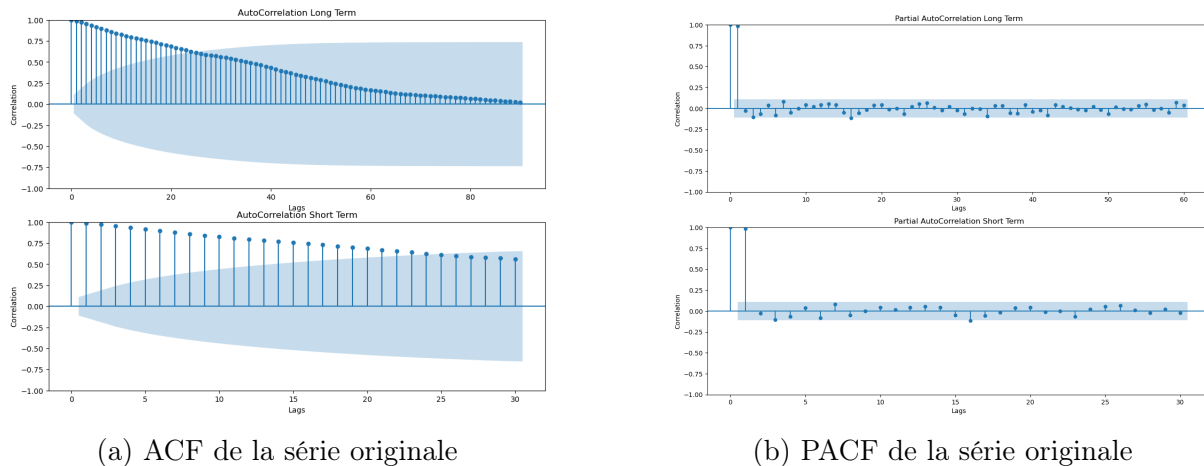


FIGURE 6 – Fonctions d'autocorrélation

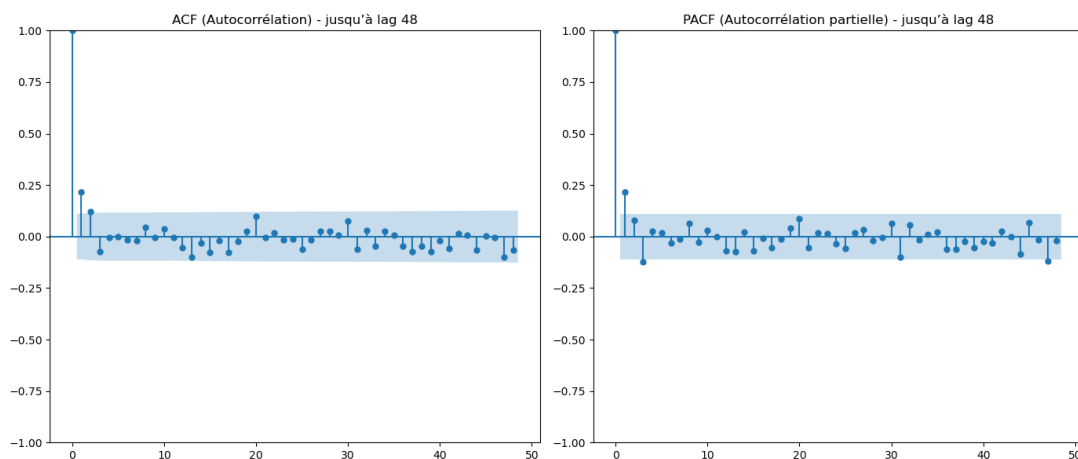


FIGURE 7 – ACF et PACF de la série log-différenciée

5.2 Justification des Paramètres SARIMA(2,1,1)(0,1,0)₁₂

5.2.1 Partie Non-Saisonnnière : ARIMA(2,1,1)

Paramètre d = 1 :

- Nécessaire pour éliminer la tendance non-stationnaire
- Confirmé par les tests ADF

Paramètre p = 2 :

- PACF montre deux pics significatifs aux lags 1 et 2
- Décroissance rapide après le lag 2
- Suggère un processus AR(2)

Paramètre $q = 1$:

- ACF présente une décroissance géométrique
- Nécessaire pour capturer la dépendance temporelle résiduelle

5.2.2 Partie Saisonnière : $(0,1,0)_{12}$ **Paramètre $D = 1$:**

- Différenciation saisonnière nécessaire (données mensuelles)
- Élimine la saisonnalité déterministe

Paramètres $P = 0$ et $Q = 0$:

- Absence de pics significatifs aux lags saisonniers multiples
- La différenciation saisonnière simple suffit

6 Estimation et Validation du Modèle

6.1 Méthode de Validation Croisée

L'évaluation du modèle a été réalisée using une approche de validation croisée temporelle (TimeSeriesSplit) avec :

- 53% des données pour l'entraînement
- 47% des données pour le test
- Période d'entraînement : 2007-2016 (avant 2007 la data n'est pas vraiment intéressante)
- Période de test : 2016-2025

6.2 Résultats de l'Estimation

Le modèle SARIMA(2,1,1)(0,1,0)₁₂ a été estimé avec les résultats suivants :

TABLE 1 – Métriques de Performance du Modèle SARIMA

Métrique	Entraînement	Test
R ² Score	0.976	0.979
MSE	1.565	75.539
MAE	0.918	6.535
MAPE	8.71%	6.65%

Conclusion : Le modèle SARIMA(2,1,1)(0,1,0)₁₂, entraîné sur les données mensuelles de stock de 2007 à 2016, a été testé sur la période de 2017 à 2024.

- **Le coefficient de détermination R^2** est très élevé pour les deux ensembles (0.976 en entraînement, 0.979 en test), ce qui signifie que le modèle capture efficacement la dynamique de la série, même sur une période non observée.
- **La MSE (erreur quadratique moyenne)** augmente fortement sur le test (75.539 contre 1.565), suggérant que certaines prédictions sont affectées par des écarts

importants, possiblement liés à des événements imprévus ou à une volatilité accrue sur la période récente.

- **La MAE (erreur absolue moyenne)** est également plus élevée sur le test (6.535 contre 0.918), traduisant une augmentation globale des erreurs, bien que cette valeur reste raisonnable dans un contexte boursier.
- **Le MAPE (erreur moyenne en pourcentage)** reste faible sur les deux périodes, et même meilleur sur le test (6.65% contre 8.71%), ce qui indique que les erreurs restent proportionnellement faibles par rapport aux valeurs réelles.

En résumé, le modèle SARIMA démontre une excellente capacité de généralisation sur la période 2017–2024. Malgré une hausse des erreurs absolues, les métriques relatives confirment la pertinence du modèle pour effectuer des prévisions mensuelles fiables sur le long terme, même dans un contexte économique potentiellement instable.

6.3 Graphiques de Validation



FIGURE 8 – Comparaison entre valeurs réelles et prédictions SARIMA

7 Prévisions et Intervalles de Confiance

7.1 Prévisions sur 30 Mois

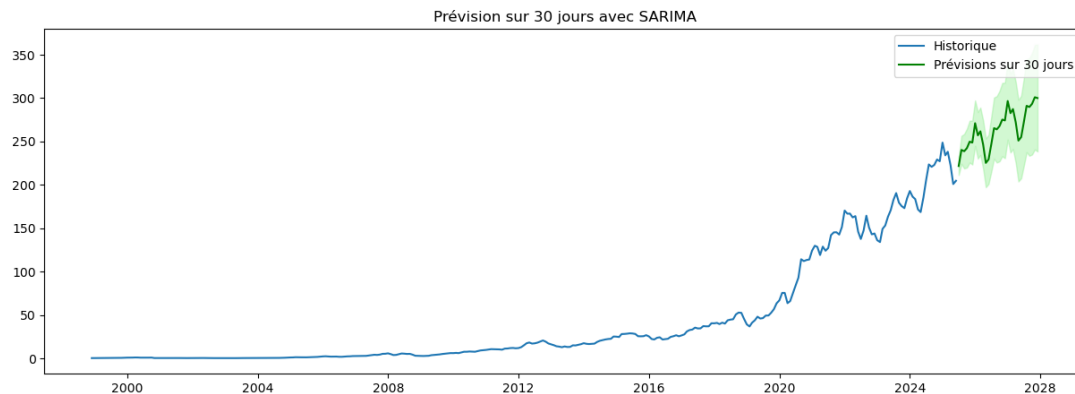


FIGURE 9 – Prévisions SARIMA sur 30 mois avec intervalles de confiance

Le modèle génère des prévisions pour les 30 prochains mois avec :

- Intervalles de confiance à 95%
- Tendance de croissance maintenue
- Incertitude croissante avec l'horizon temporal

8 Discussion et Limitations

8.1 Points Forts du Modèle

- Capture efficacement la tendance et la saisonnalité
- Performance acceptable sur la période d'entraînement
- Méthode statistiquement rigoureuse

8.2 Limitations Identifiées

8.2.1 Hétéroscédasticité

La variance non-constante des résidus, particulièrement visible après 2019, suggère que le modèle SARIMA classique pourrait être insuffisant pour capturer toute la complexité de la série.

8.2.2 Changements Structurels

La série présente des changements de régime évidents, notamment :

- Accélération de la croissance après 2012
- Augmentation de la volatilité récente
- Impact d'événements exceptionnels non modélisés

9 Conclusion

Cette analyse a démontré l'applicabilité du modèle SARIMA(2,1,1)(0,1,0)₁₂ pour la modélisation des prix des actions Apple sur la période 1998-2025. Les principales conclusions sont :

- La transformation logarithmique est essentielle pour stabiliser la série
- Le modèle capture correctement la tendance et la saisonnalité
- Les paramètres identifiés sont statistiquement justifiés
- La performance du modèle est acceptable mais limitée par l'hétéroscédasticité

La méthodologie SARIMA reste un outil fondamental pour l'analyse des séries temporelles financières, mais doit être complétée par des approches plus avancées pour capturer la complexité croissante des marchés financiers modernes.