

Test Séries Temporelles - Niveau 2 (Avancé)

UniBank Haiti - Data Analyst

Durée: 45 minutes

Questions: 25

Niveau: Avancé

Sujets: SARIMA, Prophet, validation, diagnostics avancés, cas bancaires complexes

Q1. Vous analysez les dépôts mensuels d'UniBank sur 5 ans. Le test ADF donne $p = 0.45$ et le test KPSS donne $p = 0.02$. Que concluez-vous?

- A) La série est stationnaire
- B) La série est non stationnaire (les deux tests convergent)
- C) Résultats contradictoires - besoin d'analyse supplémentaire
- D) Les tests sont invalides

Réponse: B) La série est non stationnaire (les deux tests convergent)

ADF: $H_0 = \text{non stationnaire}$, $p = 0.45 \rightarrow \text{ne pas rejeter} \rightarrow \text{non stationnaire}$. KPSS: $H_0 = \text{stationnaire}$, $p = 0.02 \rightarrow \text{rejeter} \rightarrow \text{non stationnaire}$. Les deux tests indiquent une série non stationnaire.

Q2. Après différenciation d'ordre 1 ET différenciation saisonnière (période 12), votre série devient stationnaire. Quel modèle SARIMA utiliseriez-vous?

- A) SARIMA(p,0,q)(P,0,Q)₁₂
- B) SARIMA(p,1,q)(P,1,Q)₁₂
- C) ARIMA(p,2,q)
- D) SARIMA(p,1,q)(P,0,Q)₁₂

Réponse: B) SARIMA(p,1,q)(P,1,Q)₁₂

$d=1$ (une différenciation simple) et $D=1$ (une différenciation saisonnière de période 12). Le modèle est donc SARIMA(p,1,q)(P,1,Q)₁₂.

Q3. L'ACF de votre série différenciée montre une décroissance exponentielle, tandis que le PACF coupe après le lag 2. Quels ordres AR et MA suggère ce pattern?

- A) AR(2), MA(0)
- B) AR(0), MA(2)
- C) AR(2), MA(2)

D) AR(1), MA(1)

Réponse: A) AR(2), MA(0)

ACF décroissance exponentielle + PACF coupe nettement → processus AR. PACF coupe après lag 2 → AR(2). Donc ARIMA(2,d,0).

Q4. Vous utilisez TimeSeriesSplit avec 5 folds sur 60 mois de données. Quelle est la taille approximative du dernier fold de test?

A) 12 mois

B) 10 mois

C) 60 mois

D) 5 mois

Réponse: B) 10 mois

TimeSeriesSplit divise progressivement. Avec 5 folds sur 60 mois, le dernier fold utilise ~50 mois pour l'entraînement et ~10 mois pour le test ($60/6 \approx 10$ par fold incrémental).

Q5. Le MAPE de votre modèle de prévision des transactions ATM est de 8%. Comment l'interprétez-vous?

A) Le modèle se trompe de 8 transactions en moyenne

B) L'erreur moyenne est de 8% par rapport aux valeurs réelles

C) Le modèle capture 8% de la variance

D) 8% des prévisions sont exactes

Réponse: B) L'erreur moyenne est de 8% par rapport aux valeurs réelles

MAPE = Mean Absolute Percentage Error. Un MAPE de 8% signifie qu'en moyenne, les prévisions s'écartent de 8% des valeurs observées. Généralement, $MAPE < 10\%$ est considéré comme bon.

Q6. Vous comparez deux modèles: Modèle A (AIC=450, MAPE=12%) et Modèle B (AIC=480, MAPE=9%). Lequel choisir pour la production?

A) Modèle A car AIC plus faible

B) Modèle B car MAPE plus faible sur données de test

C) Les deux sont équivalents

D) Ni l'un ni l'autre

Réponse: B) Modèle B car MAPE plus faible sur données de test

L'AIC est calculé sur les données d'entraînement. Pour la performance en production, le MAPE sur données de test (out-of-sample) est plus pertinent. Modèle B avec 9% sera plus précis.

Q7. Dans Prophet, que représente le paramètre “changepoint_prior_scale”?

- A) La saisonnalité annuelle
- B) La flexibilité de la tendance (capacité à détecter des changements brusques)
- C) Le niveau de bruit
- D) La fréquence des données

Réponse: B) La flexibilité de la tendance (capacité à détecter des changements brusques)

changepoint_prior_scale contrôle la flexibilité de la tendance. Une valeur plus élevée permet plus de changements de pente (utile si la tendance change souvent, risque d'overfitting si trop élevé).

Q8. Vos résidus de modèle ARIMA montrent une autocorrélation significative au lag 12. Que suggère ce résultat?

- A) Le modèle est parfait
- B) Il manque une composante saisonnière - passer à SARIMA
- C) Il faut supprimer la saisonnalité des données
- D) Le modèle capture trop de signal

Réponse: B) Il manque une composante saisonnière - passer à SARIMA

Une autocorrélation résiduelle au lag 12 (données mensuelles) indique que le modèle n'a pas capturé la saisonnalité annuelle. Il faut utiliser SARIMA avec une composante saisonnière de période 12.

Q9. Pour prévoir les retraits DAB du vendredi de paye (fin de mois), quelle approche est la plus appropriée?

- A) ARIMA simple sur moyennes mensuelles
- B) Modèle avec régresseurs externes incluant les indicateurs de jour de paye
- C) Lissage exponentiel simple
- D) Moyenne mobile sur 7 jours

Réponse: B) Modèle avec régresseurs externes incluant les indicateurs de jour de paye

Les pics de fin de mois liés aux salaires sont des événements calendaires prévisibles. Un modèle avec régresseurs externes (dummy pour jour de paye, jour avant fête, etc.) capturera mieux ces patterns.

Q10. Quelle est la différence entre lissage exponentiel simple (SES) et Holt-Winters additif?

- A) Aucune différence
- B) SES capture seulement le niveau, Holt-Winters capture niveau + tendance + saisonnalité
- C) Holt-Winters est toujours plus rapide
- D) SES nécessite plus de données

Réponse: B) SES capture seulement le niveau, Holt-Winters capture niveau + tendance + saisonnalité

SES: seulement le niveau (1 paramètre α). Holt: niveau + tendance (2 paramètres). Holt-Winters: niveau + tendance + saisonnalité (3 paramètres ou plus).

Q11. Vous prévoyez les flux de liquidité pour les 30 prochains jours. Le RMSE est en HTG et le MAPE en %. Laquelle de ces métriques est plus facile à communiquer au management?

- A) RMSE car en unité monétaire
- B) MAPE car en pourcentage, plus intuitif
- C) Les deux sont également difficiles
- D) Aucune n'est appropriée

Réponse: B) MAPE car en pourcentage, plus intuitif

Le MAPE en % est plus facile à communiquer ("notre prévision s'écarte en moyenne de 5%") que le RMSE en unité absolue ("erreur moyenne de 2.3 millions HTG"). Le management comprend mieux les pourcentages.

Q12. Dans une décomposition STL, vous observez que le résidu contient encore un pattern répétitif. Que suggère cela?

- A) La décomposition est parfaite
- B) La période saisonnière spécifiée est incorrecte ou il y a une saisonnalité multiple
- C) Il n'y a pas de tendance
- D) Les données sont stationnaires

Réponse: B) La période saisonnière spécifiée est incorrecte ou il y a une saisonnalité multiple

Si le résidu montre encore des patterns répétitifs, la saisonnalité n'est pas entièrement capturée. Peut-être la période est mauvaise, ou il y a plusieurs saisonnalités (ex: hebdomadaire + annuelle).

Q13. Pour un modèle ARIMA(1,1,1), combien de paramètres sont estimés (sans compter la variance)?

- A) 1
- B) 2
- C) 3
- D) 4

Réponse: C) 3

ARIMA(1,1,1): ϕ_1 (AR), $d=1$ n'est pas un paramètre estimé (juste différenciation), θ_1 (MA), plus potentiellement une constante (drift). Sans la constante: 2 paramètres. Avec constante: 3 paramètres.

Q14. Vous utilisez auto_arma et obtenez ARIMA(2,1,2). Le modèle a un AIC de 500 mais des résidus autocorrélés. Que faites-vous?

- A) Accepter le modèle car auto_arma est optimal
- B) Investiguer manuellement d'autres ordres ou ajouter une composante saisonnière
- C) Réduire les ordres à (1,1,1)
- D) Ignorer l'autocorrélation des résidus

Réponse: B) Investiguer manuellement d'autres ordres ou ajouter une composante saisonnière

auto_arma optimise l'AIC mais ne garantit pas des résidus blancs. Si l'autocorrélation persiste, il faut ajuster manuellement (SARIMA, ordres différents, ou modèle alternatif).

Q15. Quelle transformation appliquer à une série de volumes de transactions qui montre une variance croissante avec le niveau?

- A) Différenciation
- B) Transformation logarithmique ou Box-Cox
- C) Standardisation
- D) Ajout de constante

Réponse: B) Transformation logarithmique ou Box-Cox

Une variance qui croît avec le niveau indique une hétéroscédasticité multiplicative. Log ou Box-Cox stabilise la variance en “compressant” les grandes valeurs.

Q16. Dans Prophet, comment modéliser l’effet des jours fériés haïtiens (1er janvier, Carnaval, etc.)?

- A) Ajouter manuellement dans la tendance
- B) Utiliser le paramètre holidays avec un DataFrame des jours fériés
- C) Ignorer car Prophet les détecte automatiquement
- D) Utiliser uniquement la saisonnalité hebdomadaire

Réponse: B) Utiliser le paramètre holidays avec un DataFrame des jours fériés

Prophet permet d’ajouter des effets de jours fériés via un DataFrame contenant les dates et noms. C’est important pour le contexte haïtien avec ses fêtes spécifiques (Carnaval, Jour des Morts, etc.).

Q17. Vous prévoyez les défauts de paiement mensuels. La série montre une tendance croissante après le séisme de 2021. Comment modéliser cette rupture structurelle?

- A) Ignorer et utiliser toutes les données
- B) Utiliser seulement les données post-séisme
- C) Ajouter une variable indicatrice (intervention) pour la période post-séisme
- D) Appliquer une différenciation d’ordre 3

Réponse: C) Ajouter une variable indicatrice (intervention) pour la période post-séisme

Une rupture structurelle (changement de niveau permanent) se modélise par une variable indicatrice (0 avant, 1 après). Cela permet d’utiliser toutes les données tout en capturant le changement.

Q18. Le test de Ljung-Box sur les résidus donne $p = 0.02$ au lag 12. Que concluez-vous?

- A) Les résidus sont du bruit blanc
- B) Il reste de l’autocorrélation dans les résidus - le modèle est inadéquat
- C) Le modèle est parfait
- D) Il faut plus de données

Réponse: B) Il reste de l’autocorrélation dans les résidus - le modèle est inadéquat

Ljung-Box teste si les résidus sont du bruit blanc (H_0). Avec $p = 0.02 < 0.05$, on rejette H_0 - les résidus sont autocorrélés, indiquant que le modèle n'a pas capturé toute la structure.

Q19. Quelle est la principale différence entre ARIMA et modèles de lissage exponentiel (ETS)?

- A) ARIMA est toujours meilleur
- B) ETS est basé sur la décomposition état-espace, ARIMA sur l'autocorrélation
- C) Ils sont mathématiquement identiques
- D) ETS ne peut pas faire de prévisions

Réponse: B) ETS est basé sur la décomposition état-espace, ARIMA sur l'autocorrélation

ETS (Error, Trend, Seasonal) utilise une approche état-espace avec des équations de mise à jour. ARIMA modélise les dépendances temporelles via autocorrélation et moyenne mobile. Approches conceptuellement différentes.

Q20. Pour prévoir les transactions du prochain trimestre (90 jours) avec des données quotidiennes sur 2 ans, quel horizon est raisonnable?

- A) 1 an (365 jours)
- B) 90 jours maximum, avec incertitude croissante
- C) 5 ans
- D) Illimité

Réponse: B) 90 jours maximum, avec incertitude croissante

Avec 2 ans (~730 jours) de données, prévoir 90 jours ($\approx 12\%$ de l'historique) est raisonnable. L'incertitude des intervalles de prévision augmente avec l'horizon. Au-delà de 90 jours, la précision décroît significativement.

Q21. Vous observez que votre série mensuelle a une forte saisonnalité de période 12 mais aussi un pic chaque fin de trimestre (période 3). Comment modéliser?

- A) Utiliser uniquement période 12
- B) Utiliser SARIMA avec deux composantes saisonnières ou modèle à saisonnalités multiples (TBATS/Prophet)
- C) Ignorer la période 3
- D) Faire deux modèles séparés

Réponse: B) Utiliser SARIMA avec deux composantes saisonnières ou modèle à saisonnalités multiples (TBATS/Prophet)

Les saisonnalités multiples (annuelle + trimestrielle) nécessitent des modèles spécialisés: TBATS, Prophet (qui gère nativement plusieurs saisonnalités), ou des régresseurs externes pour les effets trimestriels.

Q22. Dans le contexte de prévision de liquidité bancaire, pourquoi est-il important de fournir des intervalles de prévision plutôt qu'une seule valeur?

- A) Pour impressionner le management
- B) Pour quantifier l'incertitude et permettre une planification prudente (buffers de liquidité)
- C) Les intervalles sont plus faciles à calculer
- D) C'est une exigence purement académique

Réponse: B) Pour quantifier l'incertitude et permettre une planification prudente (buffers de liquidité)

En gestion de liquidité, connaître l'intervalle (ex: "entre 50M et 70M HTG avec 95% de confiance") permet de dimensionner les réserves de sécurité. Une seule valeur donnerait une fausse précision.

Q23. Votre modèle Prophet donne un MAPE de 5% sur l'entraînement mais 15% sur le test. Quel problème cela indique-t-il?

- A) Le modèle est parfait
- B) Overfitting - le modèle a trop appris les spécificités de l'entraînement
- C) Underfitting
- D) Les données de test sont incorrectes

Réponse: B) Overfitting - le modèle a trop appris les spécificités de l'entraînement

Une différence importante entre performance train (5%) et test (15%) est le signe classique d'overfitting. Il faut réduire la flexibilité (changepoint_prior_scale plus bas, régularisation).

Q24. Pour combiner les prévisions de 3 modèles différents (ARIMA, ETS, Prophet), quelle méthode simple est souvent efficace?

- A) Choisir aléatoirement un modèle
- B) Moyenne simple ou pondérée des prévisions
- C) Multiplier les prévisions
- D) Utiliser uniquement le plus complexe

Réponse: B) Moyenne simple ou pondérée des prévisions

L'ensemble (combinaison) de modèles réduit souvent l'erreur de prévision. Une moyenne simple ou pondérée (par performance historique) est une approche efficace et robuste.

Q25. Dans un contexte de reporting réglementaire BRH, quelle fréquence de mise à jour des modèles de prévision de liquidité recommanderiez-vous?

- A) Une fois par an
- B) Jamais, le modèle initial suffit
- C) Mensuelle ou trimestrielle, avec re-calibration complète annuelle
- D) Quotidienne

Réponse: C) Mensuelle ou trimestrielle, avec re-calibration complète annuelle

Les conditions économiques changent. Une mise à jour fréquente (mensuelle/trimestrielle) des prévisions avec re-calibration complète annuelle (ou après événements majeurs) assure la pertinence des modèles.

Résumé des Concepts Clés

Tests de Stationnarité

- **ADF:** H_0 = non stationnaire (rejeter si $p < 0.05$)
- **KPSS:** H_0 = stationnaire (rejeter si $p < 0.05$)
- **Convergence:** Les deux tests devraient converger

Identification ARIMA

- **ACF décroît + PACF coupe:** AR(p)
- **ACF coupe + PACF décroît:** MA(q)
- **Les deux décroissent:** ARMA(p,q)

Validation

- **TimeSeriesSplit:** Respecte l'ordre temporel
- **MAPE < 10%:** Généralement bon
- **Ljung-Box $p > 0.05$:** Résidus OK

Saisonnalité

- **SARIMA:** Une saisonnalité
 - **TBATS/Prophet:** Saisonnalités multiples
 - **Régresseurs externes:** Événements calendaires
-

Score: ___/25