

Should Macroeconomic Forecasters Use Daily Financial Data and How?

Gestion Quantitative

Léo Renault
Nicolas Annon
Théo Verdelhan
Arthur Le Net

Master 2 – Ingénierie Économique et Financière
Parcours Finance Quantitative
Université Paris-Dauphine – PSL

31 janvier 2026

Auteurs de l'article

Elena Andreou

Département d'économie, Université de Chypre, CY 1678 Nicosie, Chypre
elena.andreou@ucy.ac.cy

Eric Ghysels

Département d'économie, Université de Caroline du Nord, Chapel Hill, NC 27599-3305,
et Département de finance, Kenan-Flagler Business School, Chapel Hill, NC 27599
eghysels@unc.edu

Andros Kourtellis

Département d'économie, Université de Chypre, CY 1678 Nicosie, Chypre
andros@ucy.ac.cy

Plan

- 1 Contexte et motivation
- 2 Modèles MIDAS : intuition et équations
- 3 Données & pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication (Tables 1, 3, 5)
- 6 Analyse récente (2024–2026) + diagnostics graphiques
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Discussion & conclusion

- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Contexte, motivation et contribution de l'étude

Introduction (réplication + objectifs)

- Réplication de *Andreou, Ghysels & Kourtellis (2013)* : utiliser des données financières **quotidiennes** pour prévoir une variable macro **trimestrielle** (PIB réel US).
- Problème clé : intégrer l'information haute fréquence sans perdre la dynamique intra-période (agrégation naïve) ni exploser le nombre de paramètres (régression HF).
- Trois objectifs :
 - ① reproduire les résultats empiriques (logique OOS),
 - ② expliciter les choix opérationnels (facteurs, alignement, métriques),
 - ③ tester une extension parcimonieuse (MIDAS à deux θ).

Problématique : fréquence mixte & information forward-looking

- **Décalage de fréquence** : macro (mensuel/trimestriel) vs finance (quotidien).
- Les prix dactifs incorporent rapidement les anticipations \Rightarrow signal utile pour **prévision** et **nowcasting**.
- Défi économétrique : **pondérer** l'information intra-trimestre (récent \gg ancien) sans sur-paramétriser.

Approches classiques : pourquoi ça coince

- **(1) Agrégation temporelle** (moyenne/somme/dernière valeur)

$$X_t = \frac{1}{m} \sum_{i=1}^m x_{t,i}, \quad Y_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1}.$$

Limite : pondération uniforme, perte de dynamique intra-période.

- **(2) Régression naïve en haute fréquence**

$$Y_{t+1} = \alpha + \sum_{i=1}^m \beta_i x_{t,i} + \varepsilon_{t+1}.$$

Limites : trop de paramètres ($m \approx 60$ / trimestre), multicolinéarité, sur-apprentissage.

Nowcasting : utile mais lourd

- Nowcasting (Kalman / état-espace) : gère *ragged edge*, données manquantes, calendrier de publication.
- Mais : spécification lourde, nombreux paramètres, sensible aux erreurs de modélisation quand beaucoup de séries quotidiennes.
- **But de MIDAS** : alternative **forme réduite**, plus simple et parcimonieuse, tout en capturant un filtrage statique via pondérations.

Objectif et contribution

- Évaluer l'apport informationnel des données financières quotidiennes pour la prévision de la croissance trimestrielle du PIB réel.
- MIDAS : exploiter la dynamique quotidienne sans agrégation arbitraire.
- Alternative à un filtre de Kalman : forme réduite, estimation plus simple, robuste.
- Identifier les classes dactifs / facteurs les plus informatifs.

- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Modèles de régression MIDAS

ADL-MIDAS : formulation

- Cible trimestrielle : $Y_{t+h}^{Q,h}$; prédicteur quotidien : $X_{m-i,t-j}^D$.
- Modèle ADL-MIDAS(p_Y^Q, q_X^D) :

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{j=0}^{p_Y^Q-1} \rho_{j+1}^h Y_{t-j}^Q + \beta^h \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{m-1} w_{i+jm}^h(\theta) X_{m-i,t-j}^D + u_{t+h}^h.$$

- Les poids $w(\theta)$ résument m observations quotidiennes via peu d'hyperparamètres.

Pondérations : Almon exponentiel (1 paramètre)

$$w_j(\theta) = \frac{\exp(\theta j^2)}{\sum_{k=1}^m \exp(\theta k^2)} \quad (j = 1, \dots, m)$$

- Contraintes : $w_j(\theta) > 0$ et $\sum_j w_j(\theta) = 1$.
- Interprétation : β^h = effet marginal global de lagrégat quotidien pondéré.
- Intuition : mémoire décroissante (observations récentes plus informatives).
- Estimation : moindres carrés non-linéaires (NLS) sur $(\mu^h, \rho^h, \beta^h, \theta^h)$.

MIDAS comme agrégation endogène (lien ADL)

- MIDAS peut s'écrire comme une agrégation guidée par les données :

$$X_t^Q(\tilde{\theta}) = \sum_{i=0}^{m-1} w_i(\tilde{\theta}) X_{m-i,t}^D.$$

- Extension ADL-MIDAS-M :

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{k=0}^{p_Y^Q-1} \rho_k^h Y_{t-k}^Q + \sum_{k=0}^{q_X^Q-1} \beta_k^h X_{t-k}^Q(\tilde{\theta}^h) + u_{t+h}^h.$$

- Lecture : filtrage statique (forme réduite) vs filtrage optimal (Kalman, états latents).

Nowcasting avec avances (leads) : idée

- Objectif : utiliser l'information **intra-trimestre** disponible avant la publication du PIB.
- Exemple : fin du 2^e mois du trimestre cible $\Rightarrow \sim 42\text{--}44$ jours de bourse de leads.
- Les leads traitent le *ragged edge* sans modèle état-espace.

ADL-MIDAS avec avances : équation

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{k=0}^{p_Y^Q-1} \rho_k^h Y_{t-k}^Q + \beta^h \left[\sum_{i=(3-J_X^D)m/3}^{m-1} w_{i-m}^{\theta^h} X_{m-i,t+1}^D + \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{m-1} w_{i+jm}^{\theta^h} X_{m-i,t-j}^D \right] + u_{t+h}^h.$$

- (1) composante **leads** : $X_{\cdot,t+1}^D$ (info trimestre courant)
- (2) composante **lags** : trimestres passés

Nowcasting Kalman vs MIDAS-leads : 2 différences

- (1) **Prévisions directes vs itérées**
 - Kalman/état-espace : souvent itéré (erreurs se propagent avec h).
 - MIDAS : **direct** (une équation par horizon h) \Rightarrow robustesse.
- (2) **Ragged edge / calendrier**
 - Kalman : calendrier explicite, données manquantes, révisions.
 - MIDAS-leads : exploite directement ce qui est disponible via avances ; pas de variable latente.

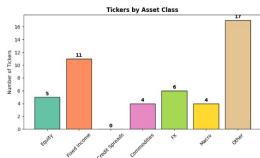
- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements**
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Données et pré-traitements

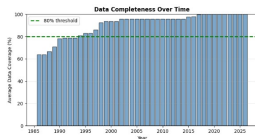
Sources, fréquences, périmètre

- Données Bloomberg export Excel, restructurées en panel *date* × *ticker*.
- 47 séries quotidiennes : 02/01/1986 – 22/01/2026 (14 631 dates).
- Variable cible : PIB réel US trimestriel (GDP CQOQ Index), 161 trimestres (1986–2026).
- Sur l'échantillon : moyenne 2.71, écart-type 4.22, min -28.0, max 34.9 (ruptures : COVID).

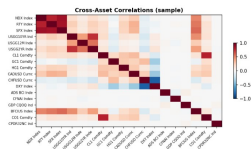
Composition & couverture



(a) Tickers par classe dactifs



(b) Complétude dans le temps



(c) Corrélations croisées

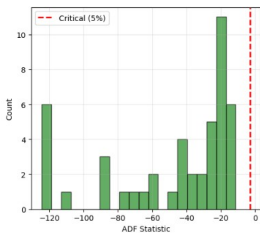
Alignement & construction des séries

- Index commun (DatetimeIndex), pivot en format large (colonnes=tickers).
- Pas d'interpolation ; infinis → manquants.
- PIB aligné fin de trimestre (dernière obs intra-trimestre).
- Prédicteurs quotidiens utilisés par blocs de m jours avant chaque date trimestrielle.

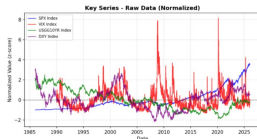
Transformations des données (règles)

Data Type	Transformation	Formula	Rationale
Prices/Indices	Log-returns	$r_t = \ln(P_t) - \ln(P_{t-1})$	Prices are I(1), returns are I(0)
Interest Rates	First differences	$\Delta x_t = x_t - x_{t-1}$	Rates can be ≤ 0 , so log impossible
Spreads	First differences	$\Delta x_t = x_t - x_{t-1}$	Already in % but often I(1)
GDP Growth	None (kept in levels)	—	Already a growth rate, stationary

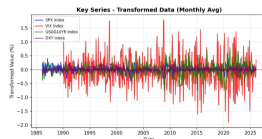
Transformations : diagnostics



(a) ADF



(b) Brut



(c) Transformé

PCA : pipeline

- Séries transformées + standardisées avant PCA (centrage-réduction).
- PCA sur 03/01/1998–22/01/2026 \Rightarrow DF1–DF5 ($10\,247 \times 5$).
- Variance expliquée : total 57.5% (DF1 20.0%, DF2 16.5%, DF3 8.6%, DF4 8.0%, DF5 4.4%).
- Séries retenues si couverture $\geq 70\%$.
- Limite : PCA idéalement réursive (éviter *look-ahead*).

Analyse PCA : variance expliquée

Facteur	Variance expliquée (%)	Corr. avec le PIB (CQOQ)
DF1	20.0	0.165
DF2	16.5	-0.050
DF3	8.6	-0.034
DF4	8.0	0.067
DF5	4.4	0.186
Cumul (DF1–DF5)	57.5	

TABLE 1 – Variance expliquée par les facteurs PCA et corrélation contemporaine avec le PIB.

Analyse PCA : interprétation

- DF1 : facteur global risk-on / risk-off (actions + taux).
- DF2 : composante liée au change (corrélations négatives FX) + taux.
- DF3–DF4 : dimensions plus spécifiques (equity, taux/FX).
- DF5 : corrélation maximale avec le PIB (0.186) dans l'échantillon.

Classe d'actifs	DF1	DF2	DF3	DF4	DF5
Equity	0.405	-0.084	0.373	-0.117	-0.024
Fixed Income	0.558	0.252	-0.135	0.187	0.041
Commodities	0.262	-0.247	-0.138	-0.148	0.020
FX	-0.080	-0.339	0.025	0.221	0.016

TABLE 2 – Corrélation moyenne des facteurs avec les séries de chaque classe d'actifs.

Limites (données)

- Multi-fréquences : choix d'ancrage (fin de trimestre), taille m , jours manquants.
- Profondeur historique inégale \Rightarrow échantillon effectif réduit si fenêtres longues.
- Crises (2008, COVID) : instabilité, pondérations concentrées ; winsorisation limitée.
- Mapping Bloomberg : proxys possibles.
- Leads : alignement temps réel strict.

- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Design de réplication et protocole de prévision

Échantillons, horizons, fenêtre m

- Fenêtre daily par défaut : $m = 63$ jours (1 trimestre de bourse).
- Sensibilité : $m \in \{126, 189, 252\}$.
- Prévisions directes : une équation estimée pour chaque horizon h .

Modèles estimés (benchmarks & MIDAS)

- Benchmark AR(1) : $y_{t+h} = \alpha + \rho y_t + u_{t+h}$.
- ADL-MIDAS : retards macro + agrégat daily pondéré.
- MIDAS avec leads : nowcasting intra-trimestre (2 mois).
- Facteurs DF1-DF5 : synthèse de linfo quotidienne.
- Combinaison de prévisions (option robustesse).

Procédure pseudo hors-échantillon

- OOS : estimation initiale puis prévisions séquentielles.
- Fenêtre expanding : ré-estimation à chaque date t sur toutes les données jusqu'à t .
- Option rolling : taille fixe pour tester la stabilité sous ruptures.

Métriques dévaluation

- RMSFE : $\sqrt{\frac{1}{T} \sum e_{t+h}^2}$
- MAE : $\frac{1}{T} \sum |e_{t+h}|$
- Gains relatifs : $100 \left(1 - \frac{\text{RMSFE}_{mod}}{\text{RMSFE}_{bench}} \right)$

- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication**
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Résultats de réplication (Long sample)

Table 1 : sans avances (no leads)

Modèle	Long $h = 1$	Long $h = 4$
<i>Univariate models</i>		
RW (absolute)	2.69	3.18
AR	1.01	0.91
<i>Models with macro data</i>		
FAR (CFNAI)	0.91	0.90
<i>Models with financial data (5 DF)</i>		
ADL (5 DF)	1.09	1.12
ADL-MIDAS (5 DF) ADL-MIDAS($J_X^D = 0$)	1.11	1.11
<i>Models with macro and financial data (CFNAI, 5 DF)</i>		
FADL (CFNAI, 5 DF)	0.96	1.12
FADL-MIDAS (CFNAI, 5 DF) FADL-MIDAS($J_X^D = 0$)	1.07	0.86

Lecture : valeurs < 1 = amélioration vs RW (RW en niveau).

Table 3 : avec avances (leads)

Modèle	Long $h = 1$	Long $h = 4$
<i>Models with leads in daily financial data</i>		
ADL-MIDAS($J_X^D = 2$)	0.97	0.87
FADL-MIDAS($J_X^D = 2$)	0.77	0.73
<i>Models with leads in monthly macro and daily financial data</i>		
FADL-MIDAS($J_M = 1, J_X^D = 2$)	0.93	0.81
<i>Models with leads in monthly macro data</i>		
FAR($J_M = 1$)	0.87	0.73
FADL($J_M = 1$)	0.90	0.88
FADL-MIDAS($J_M = 1, J_X^D = 0$)	0.97	0.82

Table 5 : ADS (macro quotidien)

Modèle	$h = 1$	$h = 4$
ADL-MIDAS($J_{ADS}^D = 2$)	0.57	0.48
FADL-MIDAS($J_M = 1, J_{ADS}^D = 2$)	0.60	0.52

- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)**
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Analyse récente (OOS 2024–2026)

Contexte (OOS récent)

- Entraînement : 2020Q1–2023Q4 ; évaluation : 2024Q1–2026Q1.
- Certains modèles ne produisent pas de forecast au premier trimestre (calendrier / date d'information).
- Sur cette fenêtre : benchmarks simples dominent plusieurs variantes MIDAS testées.

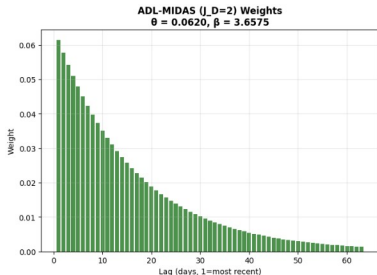
Classement modèles (OOS 2024–2026, $h = 1$)

Rang	Modèle	RMSFE	Rel. à RW	vs RW
1	ADL(flat)	0.9511	0.449	+55.1%
2	FAR(CFNAI)	1.8016	0.851	+14.9%
3	AR	1.8844	0.890	+11.0%
4	FADL($J_M = 1$)	2.1136	0.998	+0.2%
5	RW	2.1172	1.000	Baseline
6	FAR($J_M = 1$)	2.2609	1.068	-6.8%
7	ADL-MIDAS($J_D = 2$)	3.2400	1.530	-53.0%
8	FADL-MIDAS($J_M = 1, J_D = 2$)	3.6527	1.725	-72.5%
9	FADL-MIDAS	3.7317	1.763	-76.3%

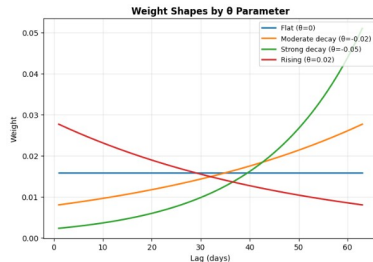
Paramètres MIDAS estimés (période récente)

Modèle	θ	β
ADL-MIDAS($J_D = 2$)	0.061963	3.6582
FADL-MIDAS	-1.000000	-1.1788

Diagnostics pondérations MIDAS

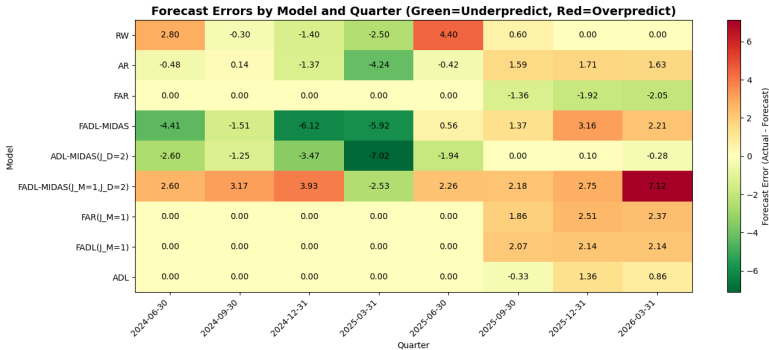


(a) Poids MIDAS estimés



(b) Formes selon θ

Carte des erreurs (2024–2026)



- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)
- 8 Conclusion

Extension : MIDAS à deux paramètres θ_{lag} vs θ_{lead}

Motivation

- MIDAS standard : un seul θ pilote les poids sur l'ensemble du bloc (lags + leads).
- Or l'information **historique** (lags) et l'information **intra-trimestre** (leads) peuvent nécessiter des profils de poids différents.
- Extension parcimonieuse : séparer θ_{lag} et θ_{lead} (on ajoute 1 paramètre).

Spécification two- θ ($h=1$)

$$y_{t+1} = \alpha + \rho y_t + \beta_{\text{lag}} \sum_{k=1}^m B(k; \theta_{\text{lag}}) x_{t-k} + \beta_{\text{lead}} \sum_{j=1}^{m_L} B(j; \theta_{\text{lead}}) x_{t+j} + \varepsilon_{t+1},$$

$$B(k; \theta) = \frac{\exp(\theta k)}{\sum_{\ell} \exp(\theta \ell)}.$$

Single- θ : cas particulier $\theta_{\text{lag}} = \theta_{\text{lead}}$.

Résultats two- θ : performance OOS

- OOS 2024–2026 (8 prévisions), $h = 1$, $p_y = 1$, $m = 63$, leads = 2 mois.
- RMSFE = 2.7267 \Rightarrow ratio vs RW = 1.288 (dégradation 28.8%).
- Mais : amélioration vs single- θ (RMSFE 3.2400).

Modèle	RMSFE	RMSFE / RW	vs RW
Two- θ MIDAS ($J_D = 2$)	2.7267	1.288	-28.8%
Single- θ MIDAS ($J_D = 2$)	3.2400	1.530	-53.0%

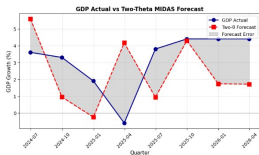
Prévisions OOS two- θ (par trimestre)

Trimestre	Réalisé	Prévu	Erreur	θ_{lag}	θ_{lead}
2024-Q2	3.600	5.595	-1.995	0.0145	0.5085
2024-Q3	3.300	0.947	2.353	0.0149	0.5205
2024-Q4	1.900	-0.251	2.151	0.0142	0.5358
2025-Q1	-0.600	4.171	-4.771	0.0130	0.5295
2025-Q2	3.800	0.932	2.868	0.0124	0.5295
2025-Q3	4.400	4.296	0.104	0.0130	0.5251
2025-Q4	4.400	1.734	2.666	0.0130	0.5248
2026-Q1	4.400	1.710	2.690	0.0124	0.5142
RMSFE : 2.7267					

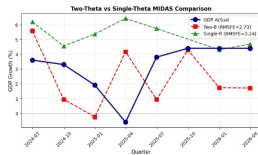
Classement global (avec \star two- θ)

Rang	Modèle	RMSFE	RMSFE / RW	vs RW
1	ADL(flat)	0.9511	0.449	+55.1%
2	FAR(CFNAI)	1.8016	0.851	+14.9%
3	AR	1.8844	0.890	+11.0%
4	FADL($J_M = 1$)	2.1136	0.998	+0.2%
5	RW	2.1172	1.000	Baseline
6	FAR($J_M = 1$)	2.2609	1.068	-6.8%
7	\star MIDAS-2 θ ($J_D = 2$)	2.7267	1.288	-28.8%
8	ADL-MIDAS($J_D = 2$)	3.2400	1.530	-53.0%
9	FADL-MIDAS($J_M = 1, J_D = 2$)	3.6527	1.725	-72.5%
10	FADL-MIDAS	3.7317	1.763	-76.3%

Figures (two- θ) du rapport



(a) Trajectoires



(b) Two- θ vs single

Discussion & limites

- Two- θ améliore le single- θ (baisse RMSFE 15.8%).
- Ne bat pas RW sur 2024–2026 :
 - OOS court (8 trimestres) \Rightarrow RMSFE sensible à quelques grosses erreurs.
 - Instabilité post-COVID possible (changement de régime).

- 1 Contexte, motivation et contribution de l'étude
- 2 Modèles de régression MIDAS
- 3 Données et pré-traitements
- 4 Design de réplication et protocole de prévision
- 5 Résultats de réplication
- 6 Analyse récente (2024–2026)
- 7 Extension : MIDAS à deux paramètres (lags vs leads)

8 Conclusion

Conclusion

Messages clés

- Les données financières quotidiennes contiennent de l'information utile pour prévoir une macro variable trimestrielle.
- MIDAS : équilibre parcimonie/flexibilité via pondérations paramétriques.
- Leads : nowcasting en forme réduite, sans états latents.
- Long sample : gains marqués (surtout leads / indicateurs macro HF), mais performances instables sur fenêtre récente.
- Réplication sur données récentes peu convaincante -> impact du COVID, meilleur résultat sur fenêtre sans COVID
- Les leads ont un très fort impact sur notre RMSFE et améliore grandement nos modèles, la cause serait un nombre de séries temporelles utilisées inférieur au nombre utilisé par le papier