

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 2025

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 Kourtellos

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 (2024–2025) + 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 (2024–2025)
- 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 & Kourtellos (2013)* : utiliser des données financières **quotidiennes** pour prévoir une variable macro **trimestrielle** (PIB réel US).
- Problème clé : intégrer linformation 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 (spécificité du papier : leads),
 - ② expliciter les choix opérationnels (facteurs ACP, lags, Discounted RMSFE...),
 - ③ tester une extension (MIDAS avec 1 coefficient pour leads et 1 coefficient pour lags).

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.

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 (2024–2025)
- 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 1 hyperparamètre.
- Modèle FADL-MIDAS(p_Y^Q, q_X^D, r_X^M) : ajout de facteurs macro (CFNAI : Chicago Fed National Activity Index)

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 lagréat quotidien pondéré.
- Intuition : mémoire décroissante (observations récentes plus informatives).
- Estimation : moindres carrés non-linéaires (NLS) sur $(\mu^h, \rho^h, \beta^h, \theta^h)$.
- Problématique du 0 informatique et incohérence du papier

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 (2024–2025)
- ⑦ Extension : MIDAS à deux paramètres (lags vs leads)

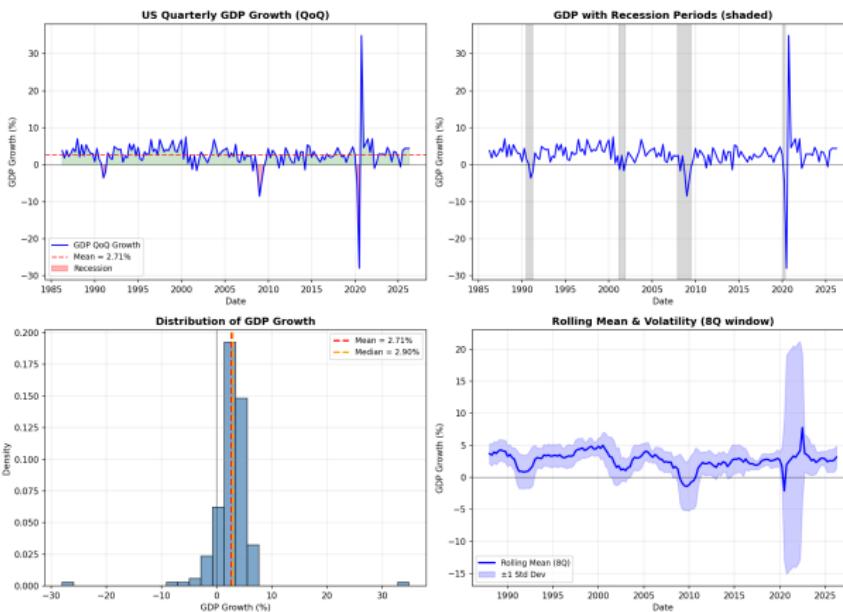
⑧ Conclusion

Données et pré-traitements

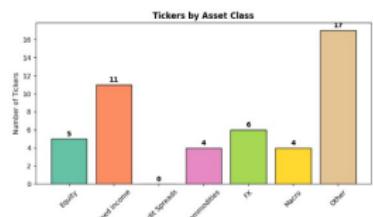
Sources, fréquences, périmètre

- Données Bloomberg export Excel, restructurées en panel *date* × *ticker*.
- 47 séries quotidiennes : 02/01/1986 – 22/01/2025 (14 631 dates).
- Variable cible : PIB réel US trimestriel (GDP CQOQ Index), 161 trimestres (1986–2025).
- Deux échantillons principaux : Long (1986-fin 2008) et Short (1999-fin 2008)

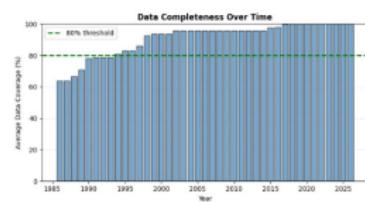
Target



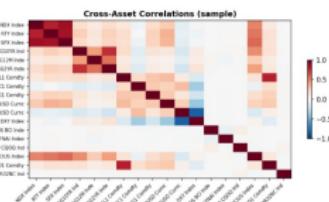
Composition & couverture



(a) Tickers par classe d'actifs



(b) Complétude dans le temps

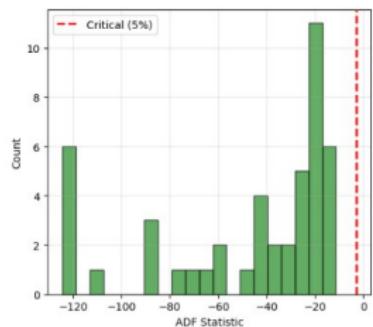


(c) Corrélation croisées

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).
- Test ADF (Augmented DickeyFuller) sur la racine unitaire ($\neq -$ > Stationnaire)
- PCA sur 03/01/1998–22/01/2025 \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 récursive (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)

- Profondeur historique inégale ⇒ échantillon effectif réduit si fenêtres longues.
- Crises (2008, COVID) : instabilité, pondérations concentrées ; winsorisation limitée.
- Mapping Bloomberg : proxys possibles.
- Difficulté : Pétrole négatif, trou dans les datas

- 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 (2024–2025)
- 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).
- Fenêtre = Période sur laquelle les poids MIDAS sont étalés
- Sensibilité : $m \in \{126, 189, 252\}$.
- Prévisions directes : une équation estimée pour chaque horizon h .

Modèles estimés (benchmarks & MIDAS)

- Random Walk (RW) : $y_{t+h} = y_t$ (prévision = dernier info disponible)
- Benchmark AR(1) : $y_{t+h} = \alpha + \rho y_t + u_{t+h}$.
- (F)ADL–MIDAS : retards macro + agrégat daily pondéré.
- MIDAS avec leads : nowcasting intra-trimestre (1 ou 2 mois).
- Facteurs DF1–DF5 (ACP): synthèse de l'info quotidienne.
- Combinaison de prévisions (Discounted RMSFE, option robustesse).

Combinaison de prévisions poids MSFE actualisés

Prévision combinée. La prévision h -pas-à-l'avance à l'instant t est une moyenne pondérée des M prévisions individuelles :

$$\hat{Y}_{c_M, t+h}^{Q,h} = \sum_{i=1}^M \omega_{i,t}^h \hat{Y}_{i,t+h}^{Q,h},$$

où $\omega_{i,t}^h$ représente le poids combiné (variable dans le temps).

Construction des poids. Les poids sont basés sur la performance historique des prévisions, en actualisant les erreurs passées :

$$\omega_{i,t}^h = \frac{(\lambda_{i,t}^{-1})^\kappa}{\sum_{j=1}^M (\lambda_{j,t}^{-1})^\kappa}, \quad \lambda_{i,t} = \sum_{\tau=T_0}^{t-h} \delta^{t-h-\tau} \left(Y_{\tau+h}^{Q,h} - \hat{Y}_{i,\tau+h}^{Q,h} \right)^2.$$

Rôle des paramètres

- δ (*facteur d'actualisation*, $\delta = 0.9$) : contrôle la vitesse à laquelle les erreurs passées sont actualisées. Plus δ est grand, plus les performances récentes des modèles sont importantes.
- κ (*paramètre de puissance*, $\kappa = 2$) : amplifie les différences de performance prédictive. Plus κ est grand, plus le poids se concentre sur les meilleurs modèles.

- 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 (2024–2025)
- 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

Résultats de réPLICATION (Short Sample)

Tableau 1 RMSFE des prévisions du PIB réel trimestriel (pas de leads)

Modèle	$h=1$	$h=4$
<i>Modèles univariés</i>		
RW	3.46	4.66
AR	1.13	1.00
<i>Modèles avec données macro</i>		
FAR (CFNAI)	0.94	0.98
<i>Modèles avec données financières (5 DF)</i>		
ADL	1.20	1.14
ADL-MIDAS($J_X^D = 0$)	1.24	1.13
<i>Modèles macro + financiers (5 DF)</i>		
FADL (CFNAI)	1.00	1.14
FADL-MIDAS($J_X^D = 0$) (CFNAI)	1.02	1.00

Tableau 3 RMSFE des modèles avec leads

Modèle	$h=1$	$h=4$
<i>Leads dans les données financières quotidiennes</i>		
ADL-MIDAS($J_X^D = 2$)	0.94	0.89
FADL-MIDAS($J_X^D = 2$)	0.70	0.62
<i>Leads dans macro mensuelles + financières quotidiennes</i>		
FADL-MIDAS($J_M = 1, J_X^D = 2$)	0.86	0.82
<i>Leads dans macro mensuelles</i>		
FAR($J_M = 1$)	0.84	0.72
FADL($J_M = 1$)	0.92	0.86
FADL-MIDAS($J_M = 1, J_X^D = 0$)	0.92	0.86

Tableau 5 RMSFE pour modèles avec leads dans ADS

Modèle	$h=1$	$h=4$
<i>Leads dans ADS quotidien</i>		
ADL-MIDAS($J_{ADS}^D = 2$)	0.56	0.42
FADL-MIDAS($J_M = 1, J_{ADS}^D = 2$)	0.60	0.46

- 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 (2024–2025)
- ⑦ Extension : MIDAS à deux paramètres (lags vs leads)

⑧ Conclusion

Analyse récente (OOS 2024–2025)

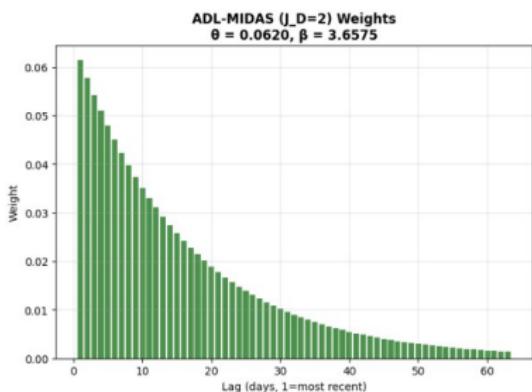
Contexte (OOS récent)

- Entraînement : 2020Q1–2023Q4 ; évaluation : 2024Q1–2025Q3.

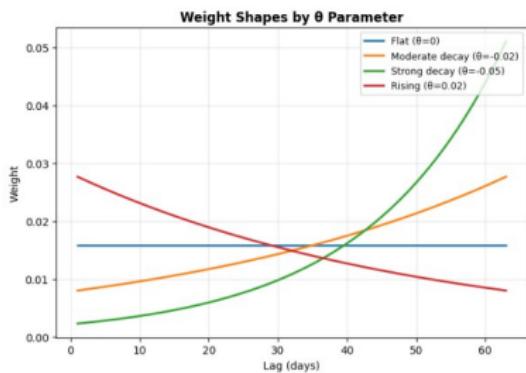
Classement modèles (OOB 2024–2025, $h = 1$)

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%

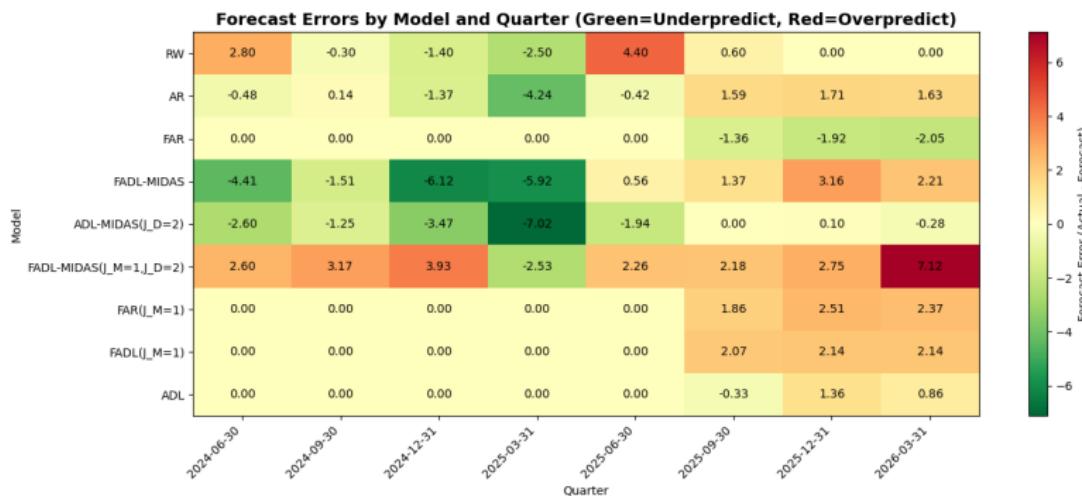
Diagnostics pondérations MIDAS



(a) Poids MIDAS estimés

(b) Formes selon θ

Carte des erreurs (2024–2025)



- 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 (2024–2025)
- ⑦ Extension : MIDAS à deux paramètres (lags vs leads)

⑧ Conclusion

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

Motivation

- MIDAS standard : un seul β 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 : séparer β_{lag} et β_{lead} (on ajoute 1 paramètre).

Spécification two- β ($h=1$)

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

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

Single- β : cas particulier $\beta_{\text{lag}} = \beta_{\text{lead}}$.

Résultats Two- β : Long Sample

- OOS 2024–2025, $h = 1$, $p_y = 1$, $m = 63$, leads = 2 mois.
- Two- θ vs Single- θ MIDAS.

Modèle	RMSFE	RMSFE / RW	vs RW
Two- θ MIDAS ($J_D = 2$)	2.3966	1.132	-13.2%
Single- θ MIDAS ($J_D = 2$)	3.2400	1.530	-53.0%

Amélioration Two- θ vs Single- θ :

- Single- θ RMSFE : 3.2400
- Two- θ RMSFE : 2.3966
- Gain : +26.0%

Résultats Two- β : Short Sample

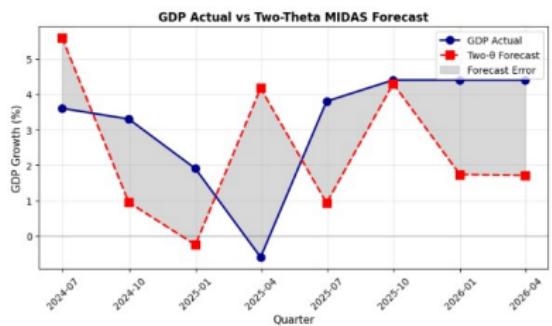
- OOS 2024–2025, $h = 1$, $p_y = 1$, $m = 63$, leads = 2 mois.
- Two- θ vs Single- θ MIDAS.

Modèle	RMSFE	RMSFE / RW	vs RW
Two- θ MIDAS ($J_D = 2$)	2.8268	1.335	-33.5%
Single- θ MIDAS ($J_D = 2$)	3.2400	1.530	-53.0%

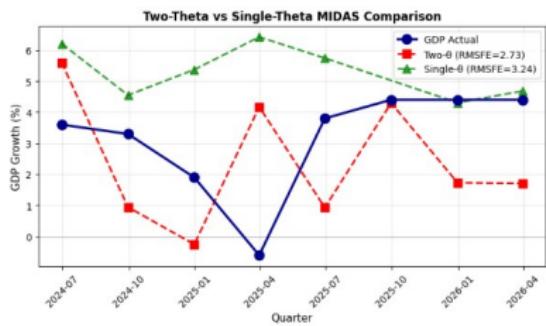
Amélioration Two- θ vs Single- θ :

- Single- θ RMSFE : 3.2400
- Two- θ RMSFE : 2.8268
- Gain : +12.8%

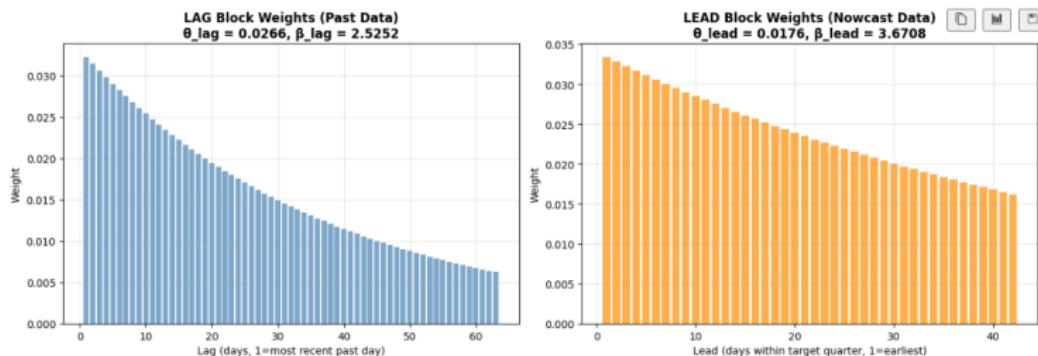
Figures (two- β)



(a) Trajectoires

(b) Two- β vs single

Extension des coefficients β avec leads et lags



- 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 (2024–2025)
- ⑦ Extension : MIDAS à deux paramètres (lags vs leads)

⑧ Conclusion

Conclusion

Messages clés

- Dans le papier, l'intuition est respectée (horizon, facteur macro, avec/sans leads) mais pas dans notre réplication (par exemple pour lh horizon).
- Les données financières quotidiennes contiennent de l'information utile pour prévoir une variable macro trimestrielle (valable avec leads).
- Sans leads, les résultats sont décevants, surtout sur la période récente.
- MIDAS : équilibre parcimonie / flexibilité via des pondérations paramétriques.
- Leads : nowcasting en forme réduite, sans états latents.
- Long sample : gains marqués (surtout avec leads et indicateurs macro HF)

Messages clés

- RéPLICATION sur données récentes peu convaincante impact du COVID ; meilleurs résultats sur la fenêtre sans COVID.
- Les leads ont un très fort impact sur notre RMSFE et améliorent grandement nos modèles. Une des causes pourrait être le nombre de séries temporelles utilisé, inférieur à celui du papier.
- Le facteur macro ADS (filtré par Kalman, publié quotidiennement) reste bien plus performant.