State Space Models vs MIDAS Regressions

Réplication du papier
State Space Models and MIDAS Regressions
Jennie Bai, Eri Ghysels et Jonathan H. Wright

N. Liesse, H. Sericola, M. Jehannin, A. Greggio

31 janvier 2026

Contents

1 Objectif du papier

L'objectif de ce papier est de comparer les régressions MIDAS et le filtre de Kalman sur le forecast de séries basses fréquences avec des séries hautes fréquences. Pour cela, le GDP[1] est déterminé à l'aide de séries mensuelles (IP[2], Oil, etc.)

Pour cela, les auteurs comparent deux approches :

- **Filtre de Kalman** avec un système d'équations qui nécessite un certain nombre de paramètres, souvent conséquent
- **Modèles MIDAS**, régressions pour prévoir de la basse fréquence à l'aide de séries hautes et basses fréquences

Les attentes sont les suivantes : les régressions MIDAS ne peuvent pas mieux prédire que le filtre de Kalman si le SSM est correctement spécifié et que les paramètres sont connus. Mais si ces conditions ne sont pas respectées, alors MIDAS peut donner de meilleurs forecasts de séries basse fréquence.

2 State Space Model (SSM) – Filtre de Kalman

2.1 Le filtre de Kalman : cadre général

Le filtre de Kalman est un outil d'estimation de **variables latentes** à partir de données observées bruitées. Il repose sur un **modèle espace–état**, composé d'une équation d'état décrivant la dynamique du facteur latent et d'une équation de mesure reliant ce facteur aux variables observées.

Dans ce papier, le filtre de Kalman permet d'exploiter des **données de fréquences mixtes** (mensuelles et trimestrielles), de traiter la variable basse fréquence comme **manquante** aux dates infra-périodiques et de produire des **prévisions optimales** conditionnelles à l'information disponible.

2.2 Spécification du modèle SSM

Le papier considère un modèle à **facteur dynamique** F_t, inobservable. La dynamique du facteur est spécifiée à la fréquence haute selon le processus suivant :

$$F_{t+j/m} = \sum_{\ell=1}^{p} \Phi_\ell F_{t+(j-\ell)/m} + \eta_{t+j/m},$$

où F_t est un vecteur de dimension $n_f \times 1$, Φ_ℓ sont des matrices de dimension $n_f \times n_f$ et $\eta_{t+j/m}$ désigne un vecteur d'innovations gaussiennes centrées i.i.d. de variance $\sigma^2_{i,n}$. Les variables observées sont reliées au facteur latent à travers des **équations de mesure**. Le PIB trimestriel

[1] Gross Domestic Product
[2] Industrial Production Index

(basse fréquence) est donné par :

$$y^*_{t+j/m} = \gamma_1' F_{t+j/m} + u_{1,t+j/m}, \qquad j = 1, \ldots, m,$$

tandis que les indicateurs mensuels (haute fréquence) s'écrivent :

$$x_{i,t+j/m} = \gamma_i' F_{t+j/m} + u_{i,t+j/m}, \qquad i = 2, \ldots, n,\ j = 1, \ldots, m.$$

Les erreurs de mesure peuvent être autocorrélées et suivre un processus AR(k).

2.3 Mise en forme du SSM : équation de mesure

Le modèle peut être réécrit sous forme espace–état comme :

$$Y_t^j = Z_j\, \alpha_{t+j/m},$$

où le vecteur d'observation dépend de la position j dans le trimestre. Plus précisément,

$$Y_t^j = \begin{cases} (y_t,\ x_{2,t}, \ldots, x_{n,t})' & \text{si } j = m, \\ (x_{2,t+j/m}, \ldots, x_{n,t+j/m})' & \text{si } 1 \leq j \leq m-1. \end{cases}$$

La matrice Z_j correspond à la **matrice de mesure** et regroupe les coefficients associés aux différentes équations de mesure.

Quand j = m, on a :

$$Z_m = \begin{bmatrix} \gamma_1' & & & \\ \gamma_2' & \mathbf{0}_{n\times n_f(p-1)} & \mathbf{I}_n & \mathbf{0}_{n\times n(k-1)} \\ \vdots & & & \\ \gamma_n' & & & \end{bmatrix}$$

Quand j $\neq$ m, on a :

$$Z_j = \begin{bmatrix} \gamma_2' & & & \\ \gamma_3' & \mathbf{0}_{(n-1)\times n_f(p-1)} & \tilde{\mathbf{I}}_{(n-1)} & \mathbf{0}_{(n-1)\times n(k-1)} \\ \vdots & & & \\ \gamma_n' & & & \end{bmatrix}$$

2.4 Vecteur d'état et matrices de transition

Le **vecteur d'état** regroupe le facteur latent ainsi que les erreurs de mesure associées aux variables observées. Il s'écrit :

$$\alpha_{t+j/m} = \left(F'_{t+j/m}, \ldots, F'_{t+(j-p+1)/m},\ u'_{t+j/m}, \ldots, u'_{t+(j-k+1)/m}\right)',$$

où $F_{t+j/m}$ désigne le **facteur latent** et $u_{t+j/m}$ correspond au vecteur des erreurs de mesure des séries observées au temps $t + j/m$.

La dynamique du vecteur d'état est décrite par l'équation de transition suivante :

$$\alpha_{t+j/m} = G\alpha_{t+(j-1)/m} + R\zeta_{t+j/m},$$

où G est la **matrice de transition** et R la matrice associée aux **innovations structurelles** $\zeta_{t+j/m}$. Ces matrices déterminent respectivement la persistance du facteur latent et la propagation des chocs dans le système.

2.5 Matrice de variance conditionnelle

Soit $P_{j+1|j}$ la matrice de **variance conditionnelle** de l'état $\alpha_{t+(j+1)/m}$ sachant l'information disponible à la date $t + j/m$. Cette matrice vérifie une **équation de Riccati périodique**, donnée pour $j = 1, \ldots, m-1$ par :

$$P_{j+1|j} = R\Sigma_\zeta R' + GP_{j|j-1}G' - GP_{j|j-1}Z_j'\big(Z_jP_{j|j-1}Z_j'\big)^{-1}Z_jP_{j|j-1}G'.$$

À la fin du cycle intra-périodique, la matrice de variance conditionnelle satisfait :

$$P_{1|m} = R\Sigma_\zeta R' + GP_{m|m-1}G' - GP_{m|m-1}Z_m'\big(Z_mP_{m|m-1}Z_m'\big)^{-1}Z_mP_{m|m-1}G'.$$

À partir de cette matrice, on définit le **gain de Kalman stationnaire périodique** :

$$K_{j|j-1} = P_{j|j-1}Z_j'\big(Z_jP_{j|j-1}Z_j'\big)^{-1}, \qquad K_{j|j-1} = K_{j+m|j+m-1},$$

qui mesure l'importance relative accordée aux nouvelles observations dans la mise à jour de l'état latent.

2.6 Extraction de l'état latent

L'**état filtré** est défini comme l'espérance conditionnelle du vecteur d'état compte tenu de l'ensemble de l'information disponible jusqu'à la date $t + j/m$:

$$\hat{\alpha}_{(t+j/m)|(t+j/m)} = \mathbb{E}\Big[\alpha_{t+j/m} \;\Big|\; Y_t^j, Y_t^{j-1}, \ldots, Y_t^1, Y_{t-1}^m, Y_{t-2}^m, \ldots\Big].$$

Il vérifie la récurrence du filtre de Kalman :

$$\hat{\alpha}_{(t+j/m)|(t+j/m)} = A_{j|j-1}\hat{\alpha}_{(t+(j-1)/m)|(t+(j-1)/m)} + K_{j|j-1}Y_t^j,$$

où

$$A_{j|j-1} = G - K_{j|j-1}Z_jG.$$

Cette récurrence implique que l'état filtré admet une **représentation en MA(∞)** :

$$\hat{\alpha}_{t|t} = \sum_{j=0}^{\infty}\sum_{i=1}^{m}\left(\tilde{A}_1^m\right)^j \tilde{A}_{i+1}^m K_{i|i-1}Y_{t-j}^i.$$

2.7 Interprétation de la représentation MA

La représentation précédente implique que l'état filtré admet une **représentation en MA(∞)**, qui peut être écrite sous la forme suivante :

$$\hat{\alpha}_{t|t} = \sum_{j=0}^{\infty} \left(\tilde{A}_1^m\right)^j K_{m|m-1} \begin{pmatrix} y_{t-j} \\ x_{2,t-j} \\ \vdots \\ x_{n,t-j} \end{pmatrix} + \sum_{j=0}^{\infty} \sum_{i=1}^{m-1} \left(\tilde{A}_1^m\right)^j \tilde{A}_{i+1}^m K_{i|i-1} \begin{pmatrix} x_{2,t-1-j+i/m} \\ \vdots \\ x_{n,t-1-j+i/m} \end{pmatrix}.$$

Cette expression met en évidence que l'état latent est une combinaison linéaire de l'information observée à différentes fréquences. En particulier, il dépend à la fois du PIB trimestriel passé et des indicateurs mensuels présents et passés, pondérés par des coefficients qui reflètent la dynamique du système et la précision relative des observations.

La prévision de la variable basse fréquence à l'horizon h s'obtient alors naturellement à partir de l'état filtré. Elle est donnée par :

$$\mathbb{E}_t[y_{t+h}] = Z_{m,1}\, G^{mh}\, \hat{\alpha}_{t|t},$$

où $Z_{m,1}$ désigne la première ligne de la matrice de mesure Z_m. Cette expression montre que la prévision repose sur une propagation dynamique de l'état latent estimé à la date t.

2.8 Intuition du filtre de Kalman

Le filtre de Kalman est un algorithme récursif destiné à estimer un facteur latent non observable représentant l'état sous-jacent de l'économie. À chaque date, il combine l'information issue de la dynamique du modèle avec celle contenue dans des variables observées bruitées, éventuellement de fréquences différentes.

Son fonctionnement repose sur deux étapes successives. Dans un premier temps, une phase de **prévision** permet d'anticiper l'état latent à partir de son évolution passée, en utilisant uniquement l'équation de transition. Dans un second temps, une phase de **mise à jour** corrige cette prédiction à l'aide des nouvelles observations disponibles. La pondération entre ces deux sources d'information dépend de leur précision relative, telle qu'elle est capturée par les matrices de variance du modèle.

Ainsi, le filtre de Kalman permet d'exploiter de manière cohérente l'ensemble de l'information disponible en temps réel et d'actualiser efficacement l'estimation du facteur latent à chaque nouvelle observation, ce qui en fait un outil particulièrement adapté à la prévision avec des données de fréquences mixtes.

3 Modèles MIDAS

En économie et en finance, il est récurrent de devoir prévoir des variables **de basse fréquence** (PIB trimestriel) à partir de variables **de haute fréquence** (mensuelles, hebdomadaires).

Cependant, les approches standards posent problème. Faut-il effectuer une agrégation des données ? Cela conduit à une perte d'informations à travers une mesure telle que la moyenne, la dernière valeur ou le quantile. Faut-il estimer un VAR haute fréquence ? Cela conduit à un algorithme lourd avec beaucoup de paramètres à estimer. Faut-il alors construire un State Space Model ? Cela revient également à estimer beaucoup de paramètres et donc à obtenir un modèle lourd et structurel.

MIDAS exploite l'information haute fréquence sans l'agréger brutalement, via une pondération sur les retards.

3.1 Modèle DL-MIDAS

La forme réduite de Andreou et al. (2009) - Distributed Lag MIDAS (DL-MIDAS) regression

$$y_{t+h} = \beta \sum_{j=0}^{K} w_j \, x_{t-\frac{j}{m}} + \varepsilon_t, \tag{1}$$

où w_j désigne une fonction de pondération sur les retards haute fréquence.

Les poids suivent un **polynôme d'Almon exponentiel à deux paramètres** :

$$w_j(\theta_1, \theta_2) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{l=0}^{K} \exp(\theta_1 l + \theta_2 l^2)}, \tag{2}$$

avec $\sum_{j=0}^{K} w_j = 1$

Le modèle DL-MIDAS constitue une approche réduite permettant d'exploiter l'information de variables haute fréquence pour la prévision d'une variable basse fréquence, sans imposer de structure dynamique multivariée explicite. Le polynôme d'Almon exponentiel impose une structure lisse sur les poids, réduisant le nombre de paramètres à estimer.
En contrepartie, un DL-MIDAS ne permet pas une mise à jour de l'information passée mais uniquement une relation avec une autre série de fréquence différente.

3.2 Regular MIDAS Regression - ADL MIDAS

Par rapport au DL-MIDAS, la regression Regular MIDAS - ADL introduit une composante autorégressive à basse fréquence ainsi qu'un terme MIDAS reposant sur les observations de haute fréquence. La dynamique intra-période de la variable explicative haute fréquence n'est toutefois pas modélisée explicitement. L'ensemble des retards haute fréquence est directement intégré à l'équation de prévision à l'aide d'un unique polynôme d'Almon.

$$y_{t+h} = \beta_y \sum_{j=0}^{\bar{K}} w_j(\theta_y)\, y_{t-j} \;+\; \beta_x \sum_{j=0}^{3\bar{K}} w_j(\theta_x)\, x_{t-j/m} \;+\; \varepsilon_{t+h} \tag{3}$$

Cette spécification constitue une approche de forme réduire qui traite l'ensemble des observations de haute fréquence et ne distingue pas les effets intra- et inter-périodes.

3.3 Multiplicative MIDAS Regression - ADL MIDAS

À la différence de la spécification MIDAS régulière, la régression MIDAS multiplicative introduit une structure hiérarchique dans l'exploitation de l'information haute fréquence. Elle distingue explicitement l'agrégation intra-périodique des observations haute fréquence et la dynamique inter-périodes, en combinant deux polynômes de pondération distincts.

$$y_{t+h} = \beta_y \sum_{j=0}^{K_y} w_j(\theta_y)\, y_{t-j} \;+\; \beta_x \sum_{j=0}^{K_x} w_j(\theta_x^{(1)})\, x(\theta_x^{(2)})_{t-j} \;+\; \varepsilon_{t+h} \tag{4}$$

Avec $w_j(\theta_y)$ et $w_j(\theta_x^1)$ les poids d'Almon exponentiel et

$$x_t(\theta_x^{(2)})_t \equiv \sum_{k=0}^{m-1} w_k(\theta_x^{(2)})\, L^{k/m} x_{t-\frac{k}{m}}. \tag{5}$$

qui suit aussi un schéma d'Almon exponentiel

Cette formulation multiplicative permet de modéliser séparément l'importance relative des observations haute fréquence à l'intérieur d'une période de basse fréquence et leur propagation dynamique entre périodes. Cela implique une approximation plus flexible au prix d'une complexité paramétrique supplémentaire.

4 SSM vs MIDAS (Out Of Sample)

Dans cette section nous allons évaluer dans quelle mesure les régressions MIDAS peuvent reproduire le comportement du filtre de Kalman, qui constitue une référence optimale lorsque le SSM est correctement spécifié.

L'analyse repose sur une étude en population et s'articule autour de trois exercices.

- Comparer les poids de prévision MIDAS à ceux du filtre de Kalman dans un modèle à un facteur correctement spécifié.
- Évaluer la capacité de MIDAS à agréger efficacement plusieurs séries haute fréquence.
- Étudier la performance relative de MIDAS lorsque le filtre de Kalman est basé sur un modèle mal spécifié.

4.1 1-factor SSM vs MIDAS

Afin d'évaluer dans quelle mesure les régressions MIDAS peuvent reproduire les poids de prévision du filtre de Kalman lorsque le SSM est correctement spécifié, nous comparons les poids issus des deux approches à l'aide de la mesure suivante :

$$\mathbb{L}^2 \equiv \sum_{j=0}^{3\bar{K}} \left(w_{x,j}^{KF} - w_{x,j}^{Mds}\right)^2 \;+\; \sum_{j=0}^{\bar{K}} \left(w_{y,j}^{KF} - w_{y,j}^{Mds}\right)^2 \tag{6}$$

Les poids du filtre de Kalman sont obtenus à partir du SSM à un facteur latent correctement spécifié :

$$
\begin{aligned}
f_{t+j/m} &= \rho f_{t+(j-1)/m} + \eta_{t+j/m}, \\
y^*_{t+j/m} &= \gamma_1 f_{t+j/m} + u_{1,t+j/m}, \\
x_{t+j/m} &= \gamma_2 f_{t+j/m} + u_{2,t+j/m}, \\
u_{i,t+j/m} &= d\, u_{i,t+(j-1)/m} + \varepsilon_{i,t+j/m}, \qquad i = 1, 2.
\end{aligned}
$$

Le modèle espace d'état est d'abord résolu en régime stationnaire périodique à l'aide du filtre de Kalman, ce qui fournit une règle de prévision linéaire pour $\widehat{y}_{t+h|t}$ en fonction de l'ensemble des observations passées. Les poids du filtre de Kalman sont ensuite extraits comme les dérivées partielles de cette règle de prévision par rapport aux observations retardées de la variable basse fréquence et de la variable haute fréquence, soit

$$
w_{y,j}^{KF} = \frac{\partial \widehat{y}_{t+h|t}}{\partial y_{t-j}}, \qquad w_{x,k}^{KF} = \frac{\partial \widehat{y}_{t+h|t}}{\partial x_{t-k/m}}.
$$

Cette procédure revient à identifier les coefficients de la projection linéaire implicite associée au filtre de Kalman en régime stationnaire (résolution par méthode d'impulsion ici, plutôt que par méthode analytique comme fait dans le papier).

Les paramètres des régressions MIDAS sont ensuite choisis de manière à minimiser la distance $\mathbb{L}^2$ par rapport aux poids du filtre de Kalman.

Les paramètres suivants sont fixés :

$$
d = d_1 = d_2, \gamma_1 = \gamma_2 = 1
$$

et

$$
\sigma_\eta^2 = \sigma_{u_y}^2 = \sigma_{u_x}^2 = 1.
$$

Les résultats sont obtenus en faisant varier :

$$
d \in \{-0.9, -0.5, 0.0, 0.5, 0.95\} \quad \text{et} \quad \rho \in \{-0.9, -0.5, 0, 0.5, 0.95\},
$$

pour $m \in \{3, 13\}$ et $h \in \{1, 4\}$.

Table 1: Résultats d'approximation entre le modèle MIDAS et le SSM à un facteur. Les entrées reportent la mesure en norme L^2 définie à l'équation (6), comparant les spécifications du filtre de Kalman et du modèle MIDAS. Le schéma MIDAS multiplicatif correspond à la régression ADL-MIDAS (équation (4)), tandis que le schéma régulier renvoie à la régression MIDAS standard (équation (3)). Les tables A à D présentent différentes configurations de la fréquence d'échantillonnage relative m et de l'horizon de prévision h.

	Panel A: One-Quarter Horizon ($h = 1$)										**Panel B: Four-Quarters Horizon ($h = 4$)**									
	Regular MIDAS					Multiplicative MIDAS					Regular MIDAS					Multiplicative MIDAS				
$d\backslash\rho$	-0.9	-0.5	0.0	0.5	0.95	-0.9	-0.5	0.0	0.5	0.95	-0.9	-0.5	0.0	0.5	0.95	-0.9	-0.5	0.0	0.5	0.95
-0.9	0.000	0.004	0.011	0.013	0.039	0.000	0.036	0.082	0.013	0.039	0.000	0.001	0.002	0.001	0.000	0.000	0.008	0.012	0.001	0.000
-0.5	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.0	0.011	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
0.5	0.046	0.002	0.000	0.000	0.000	0.046	0.002	0.002	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000
0.95	0.440	0.113	0.046	0.013	0.000	0.852	0.098	0.114	0.046	0.000	0.012	0.034	0.018	0.007	0.000	0.013	0.034	0.045	0.004	0.000
	Panel C: One-Quarter Horizon ($h = 1$)										**Panel D: Four-Quarters Horizon ($h = 4$)**									
	Regular MIDAS					Multiplicative MIDAS					Regular MIDAS					Multiplicative MIDAS				
$d\backslash\rho$	-0.9	-0.5	0.0	0.5	0.95	-0.9	-0.5	0.0	0.5	0.95	-0.9	-0.5	0.0	0.5	0.95	-0.9	-0.5	0.0	0.5	0.95
-0.9	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.0	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.5	0.004	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.95	0.111	0.030	0.014	0.031	0.000	0.238	0.076	0.048	0.031	0.000	0.002	0.001	0.000	0.001	0.000	0.001	0.001	0.001	0.001	0.000

Étude des résultats :

La Table 1 montre que, lorsque le SSM à un facteur latent est correctement spécifié, les régressions MIDAS sont capables de reproduire très fidèlement la structure des poids de prévision du filtre de Kalman (la distance quadratique entre les poids Kalman et MIDAS étant faible pour une large gamme de valeurs et paramètres).

Ce résultat indique que, même en l'absence d'une modélisation explicite de l'état latent, MIDAS peut fournir une approximation précise de l'opérateur de prévision optimal associé au filtre de Kalman. La spécification multiplicative MIDAS tend en particulier à mieux capturer la forme des poids issus du modèle state-space, notamment lorsque la fréquence relative m est élevée.

4.2 Plusieurs séries haute fréquence : SSM vs MIDAS

Afin d'évaluer la capacité de MIDAS à combiner plusieurs indicateurs haute fréquence, nous comparons les poids de prévision issus du filtre de Kalman (optimaux) à ceux obtenus à partir de différentes spécifications MIDAS.

La proximité entre les deux méthodes est mesurée à l'aide de la distance quadratique suivante:

$$\mathbb{L}^2 \equiv \sum_{j=0}^{3\bar{K}} \left(w_{x_2,j}^{KF} - w_{x_2,j}^{Mds} \right)^2 + \sum_{j=0}^{3\bar{K}} \left(w_{x_3,j}^{KF} - w_{x_3,j}^{Mds} \right)^2 + \sum_{j=0}^{\bar{K}} \left(w_{y,j}^{KF} - w_{y,j}^{Mds} \right)^2 \tag{7}$$

Les poids de Kalman sont calculés à partir du même SSM que précédemment, tandis que les paramètres MIDAS sont choisis de manière à minimiser la distance $\mathbb{L}^2$ par rapport aux poids du filtre de Kalman. Le DGP repose sur un modèle à un facteur commun, avec deux séries haute fréquence. Les paramètres des modèles (d, γ, ρ, m, h et σ) sont fixés comme dans la table précédente, à l'exception des variances des erreurs de mesure associées aux séries haute fréquence.

Deux configurations sont alors considérées :

$$\text{(i) Bruits de même variance :} \qquad \mathrm{Var}(u_{2,t}) = \mathrm{Var}(u_{3,t}),$$

$$\text{(ii) Bruits de variance inégale :} \qquad \mathrm{Var}(u_{2,t}) = \frac{1}{10}\,\mathrm{Var}(u_{3,t}).$$

Cette seconde configuration introduit une asymétrie dans la qualité informationnelle des indicateurs haute fréquence, le filtre de Kalman attribuant alors un poids plus faible à la série la plus bruitée.

Table 2: Cas avec plusieurs séries haute fréquence : combinaisons de prévisions MIDAS et SSM à un facteur. Les entrées reportent la mesure en norme L^2 définie à l'équation (7), comparant les spécifications du filtre de Kalman et du modèle MIDAS. Le schéma MIDAS multiplicatif correspond à la régression ADL-MIDAS (équation (4)), tandis que le schéma régulier renvoie à la régression MIDAS standard (équation (3)). Deux séries de haute fréquence sont considérées, avec des cas de variances de bruit égales, $\mathrm{var}(u_{2,t+j/m}) = \mathrm{var}(u_{3,t+j/m})$, et inégales, $\mathrm{var}(u_{2,t+j/m}) = \mathrm{var}(u_{3,t+j/m})/10$. Les tables A–B correspondent à $m = 3$ et les tables C–D à $m = 13$, avec un horizon de prévision fixé à $h = 1$.

	Panel A: Two high frequency series equal noise variance										**Panel B: Two high frequency series unequal noise variance**									
	Regular MIDAS					Multiplicative MIDAS					Regular MIDAS					Multiplicative MIDAS				
$d\backslash\rho$	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95
-0.9	0.000	0.003	0.007	0.027	0.027	0.000	0.020	0.012	0.027	0.027	0.000	0.004	0.010	0.013	0.039	0.000	0.035	0.010	0.014	0.039
-0.5	0.001	0.000	0.000	0.000	0.001	0.027	0.000	0.001	0.000	0.017	0.002	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000	0.001
0.0	0.005	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000
0.5	0.023	0.001	0.000	0.000	0.000	0.023	0.005	0.001	0.000	0.000	0.039	0.002	0.000	0.000	0.000	0.040	0.002	0.002	0.000	0.001
0.95	0.238	0.065	0.029	0.008	0.000	0.219	0.111	0.027	0.028	0.000	0.376	0.098	0.041	0.012	0.000	0.403	0.144	0.071	0.039	0.000
	Panel C: Two high frequency series equal noise variance										**Panel D: Two high frequency series unequal noise variance**									
	Regular MIDAS					Multiplicative MIDAS					Regular MIDAS					Multiplicative MIDAS				
$d\backslash\rho$	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95
-0.9	0.000	0.004	0.001	0.002	0.001	0.000	0.007	0.001	0.002	0.027	0.000	0.000	0.000	0.000	0.001	0.000	0.007	0.010	0.000	0.001
-0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.0	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
0.5	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.001
0.95	0.056	0.017	0.009	0.003	0.000	0.056	0.035	0.024	0.017	0.000	0.093	0.026	0.012	0.028	0.000	0.093	0.026	0.012	0.026	0.000

Étude des résultats :

Les résultats montrent que MIDAS est capable de combiner efficacement l'information provenant de plusieurs indicateurs, de manière comparable au filtre de Kalman.

Lorsque les séries haute fréquence ont des variances de bruit identiques, MIDAS reproduit une combinaison symétrique proche de celle du Kalman optimal. Lorsque les variances de bruit sont inégales, le filtre de Kalman attribue naturellement moins de poids à la série la plus bruitée ; les résultats de la Table 2 indiquent que MIDAS parvient également à apprendre cette hiérarchie informationnelle.

Ainsi, même sans imposer explicitement une structure factorielle, MIDAS est capable d'agréger plusieurs sources d'information haute fréquence d'une manière proche de l'agrégation optimale fournie par un SSM correctement spécifié.

4.3 Mauvaise spécification : MIDAS vs Kalman

À présent nous allons étudier un environnement dans lequel le SSM est **mal spécifié**. Le DGP repose sur un modèle à **deux facteurs latents** :

$$\begin{aligned}
f^{(1)}_{t+j/m} &= \rho f^{(1)}_{t+(j-1)/m} + \eta_{1,t+j/m}, \\
f^{(2)}_{t+j/m} &= \rho f^{(2)}_{t+(j-1)/m} + \eta_{2,t+j/m}, \\
y^*_{t+j/m} &= \gamma_1 f^{(1)}_{t+j/m} + \gamma_2 f^{(2)}_{t+j/m} + u_{1,t+j/m}, \\
x_{t+j/m} &= \gamma_3 f^{(1)}_{t+j/m} + u_{2,t+j/m}, \\
u_{i,t+j/m} &= d\, u_{i,t+(j-1)/m} + \varepsilon_{i,t+j/m}, \qquad i = 1, 2.
\end{aligned}$$

Tandis que le filtre de Kalman est basé sur un modèle à **un seul facteur latent** :

$$\begin{aligned}
f_{t+j/m} &= \rho f_{t+(j-1)/m} + \eta_{t+j/m}, \\
y^*_{t+j/m} &= \gamma_1 f_{t+j/m} + u_{1,t+j/m}, \\
x_{t+j/m} &= \gamma_2 f_{t+j/m} + u_{2,t+j/m}, \\
u_{i,t+j/m} &= d\, u_{i,t+(j-1)/m} + \varepsilon_{i,t+j/m}, \qquad i = 1, 2.
\end{aligned}$$

Les paramètres du filtre de Kalman mal spécifiés sont choisis de manière à minimiser la variance de l'erreur de prévision sous le vrai DGP. De son côté, MIDAS est également calibrée afin de minimiser la variance de l'erreur de prévision.

La performance relative des deux approches est mesurée à l'aide du ratio suivant :

$$\frac{\mathbf{w}'_{KF}\,\mathbf{\Sigma}_{xy}\,\mathbf{w}_{KF}}{\mathbf{w}'_{Mds}\,\mathbf{\Sigma}_{xy}\,\mathbf{w}_{Mds}} \tag{8}$$

Les paramètres ρ et d varient sur la même grille que précédemment, pour $m \in \{3, 13\}$ et $h \in \{1, 4\}$.

Un ratio inférieur à 1 indique une meilleure performance du filtre de Kalman mal spécifié, tandis qu'un ratio supérieur signale une meilleure performance de MIDAS.

Table 3: Distances d'erreur de prévision : SSM à un facteur vs MIDAS. Les entrées correspondent aux ratios de variance de l'erreur de prévision définis à l'équation (8), les paramètres de MIDAS étant choisis de manière à minimiser $\mathbf{w}'_{Mds}\mathbf{\Sigma}_{xy}\mathbf{w}_{Mds}$. Le ratio de distance est calculé en comparant un filtre de Kalman à un facteur et un modèle MIDAS régulier équation (3). Le processus générateur de données repose sur un modèle à deux facteurs équation (10) avec $\rho_1 = \rho_2 = \rho$ et $d_1 = d_2 = d$. Les résultats sont reportés pour deux horizons de prévision, $h = 1$ et $h = 4$.

$d\backslash\rho$	-0.9	-0.5	0	0.5	0.95	-0.9	-0.5	0	0.5	0.95
PE distance: Sampling frequency $m = 3$										
	Panel A : One-Quarter Horizon ($h = 1$)					**Panel B : Four-Quarter Horizon** ($h = 4$)				
-0.9	0.864	0.967	0.967	0.726	0.738	0.763	0.847	0.838	0.822	0.768
-0.5	0.642	1.00	1.00	0.997	0.724	0.892	0.998	0.995	1.00	0.781
0.0	0.638	0.997	0.999	0.997	0.778	0.882	1.00	1.00	1.00	0.764
0.5	0.689	0.997	0.996	1.00	0.743	0.901	0.995	0.997	1.00	0.801
0.95	0.743	0.721	0.744	0.762	0.746	0.758	0.743	0.747	0.736	0.701
PE distance: Sampling frequency $m = 13$										
	Panel C : One-Quarter Horizon ($h = 1$)					**Panel D : Four-Quarter Horizon** ($h = 4$)				
-0.9	0.981	0.987	0.992	0.993	0.867	1.00	0.999	1.000	0.996	0.995
-0.5	0.988	0.998	0.999	1.00	0.853	1.00	0.991	1.00	1.00	0.995
0.0	0.989	0.995	0.996	1.00	0.838	1.00	0.993	1.00	0.999	0.982
0.5	0.994	1.00	1.00	1.00	0.842	0.992	1.00	1.00	0.994	0.983
0.95	0.864	0.852	0.833	0.830	0.761	0.995	0.991	0.982	0.982	0.978

Étude des résultats :

Pour une fréquence d'échantillonnage faible ($m = 3$, tables A et B), les résultats montrent que le filtre de Kalman mal spécifié peut conserver un avantage relatif lorsque la persistance est faible ou négative, en particulier à l'horizon court ($h = 1$). En revanche, pour des valeurs de ρ plus modérées ou positives, les ratios sont fréquemment proches ou supérieurs à 1, ce qui suggère que MIDAS est au moins aussi performante, voire légèrement supérieure, dans ces configurations.

Lorsque la fréquence d'échantillonnage augmente ($m = 13$, tables C et D), les ratios deviennent très proches de l'unité, et souvent égaux à 1, quelle que soit la valeur de ρ ou de d. Cela indique que, dans ces environnements, la mauvaise spécification du filtre de Kalman neutralise largement son avantage potentiel, et que MIDAS parvient à atteindre une performance prédictive comparable, même face à un modèle structurellement mal spécifié.

Dans l'ensemble, la Table 3 met en évidence que l'avantage du filtre de Kalman disparaît progressivement lorsque le degré de mauvaise spécification augmente ou lorsque la fréquence d'échantillonnage s'élève, confirmant que MIDAS constitue une alternative robuste dans des environnements où la structure latente du DGP est imparfaitement connue.

5 Simulations de Monte Carlo

5.1 Objectif et descriptions des simulations

Cette section présente une étude de simulations de Monte Carlo visant à évaluer la performance relative des modèles SSM et MIDAS dans des contextes de spécification correcte et erronée du modèle.

Les données sont générées à partir d'un modèle à facteur latent défini en haute fréquence, dans lequel les innovations du facteur et les erreurs de mesure sont supposées gaussiennes, centrées et mutuellement indépendantes.

Trois scénarios distincts sont considérés. Dans le premier, le modèle SSM estimé correspond exactement au processus générateur de données. Le second scénario introduit une erreur de spécification en estimant un modèle à deux facteurs alors que le DGP ne comporte qu'un seul facteur. Enfin dans le troisième scénario, le nombre de facteurs est sélectionné à l'aide de critères d'information AIC - BIC.
Ces trois scénarios sont comparés aux régressions DL-MIDAS et ADL-MIDAS.

Les simulations sont effectuées pour différentes tailles d'échantillon, avec $T = 40$ observations basse fréquence et Tm observations haute fréquence. Les paramètres d et ρ varient respectivement dans les ensembles $\{-0.9, -0.5, 0, 0.5, 0.95\}$ et plusieurs combinaisons de fréquences ($m \in \{3\}$) et d'horizon de prévision ($h \in \{1, 4\}$) sont considérées.

Les performances des modèles sont évaluées à l'aide de l'erreur quadratrique moyenne de prévision (RMSPE).

Afin de comparer directement les performances prédictives des approches SSM et MIDAS, les résultats sont présentés sous la forme d'un ratio des RMSPE obtenues avec le filtre de Kalman relativement aux modèles MIDAS. Un ratio inférieur à 1 indique une performance supérieure du SSM, tandis qu'un ratio supérieur à 1 favorise le modèle MIDAS.

5.2 Simulation 1 - SSM 1 facteur vs MIDAS

Dans cette première simulation, les données sont générées à partir d'un modèle à facteur latent défini à la fréquence haute. Le facteur latent $f_{t+j/m}$ suit un processus autorégressif d'ordre 1. Les variables observées sont reliées au facteur latent à travers des équations de mesure linéaires. La variable de basse fréquence $y^*_{t+j/m}$ et l'indicateur haute fréquence $x_{t+j/m}$ dépendent du facteur latent courant avec des coefficients de charge identiques, $\gamma_1 = \gamma_2 = 1$.

$$
\begin{aligned}
f_{t+j/m} &= \rho f_{t+(j-1)/m} + \eta_{t+j/m}, \\
y^*_{t+j/m} &= \gamma_1 f_{t+j/m} + u_{1,t+j/m}, \\
x_{t+j/m} &= \gamma_2 f_{t+j/m} + u_{2,t+j/m}, \\
u_{i,t+j/m} &= d\, u_{i,t+(j-1)/m} + \varepsilon_{i,t+j/m}, \qquad i = 1, 2.
\end{aligned} \tag{9}
$$

Les erreurs idiosyncratiques associées aux équations de mesure suivent des processus autorégressifs d'ordre un, avec un coefficient commun d. Les innovations du facteur latent et les chocs idiosyncratiques sont supposés indépendants et distribués selon des lois normales centrées et de variance unitaire.

L'objectif de cette simulation est de comparer les performances de prévision hors échantillon des approches MIDAS, ADL-MIDAS et du filtre de Kalman lorsque le modèle espace–état est correctement spécifié.

5.3 Simulation 2 - SSM mal spécifié vs MIDAS

La seconde simulation considère un environnement dans lequel le modèle SSM est volontairement mal spécifié. Les données sont générées à partir d'un processus à deux facteurs latents définis en haute fréquence, tandis que le filtre de Kalman est estimé sous l'hypothèse erronée d'un seul facteur latent.

Le processus générateur de données est donné par le système suivant :

$$\begin{aligned}
f^{(1)}_{t+j/m} &= \rho f^{(1)}_{t+(j-1)/m} + \eta_{1,t+j/m},\\
f^{(2)}_{t+j/m} &= \rho f^{(2)}_{t+(j-1)/m} + \eta_{2,t+j/m},\\
y^*_{t+j/m} &= \gamma_{1,1} f^{(1)}_{t+j/m} + \gamma_{1,2} f^{(2)}_{t+j/m} + u_{1,t+j/m},\\
x_{t+j/m} &= \gamma_{2,1} f^{(1)}_{t+j/m} + \gamma_{2,2} f^{(2)}_{t+j/m} + u_{2,t+j/m},\\
u_{i,t+j/m} &= d\, u_{i,t+(j-1)/m} + \varepsilon_{i,t+j/m}, \qquad i = 1, 2.
\end{aligned} \tag{10}$$

Les deux facteurs latents suivent des dynamiques autoregressives d'ordre 1. La variable basse fréquence dépend simultanément de deux facteurs latents, tandis que l'indicateur haute fréquence ne dépend que du premier facteur. Les coefficients sont fixés comme suit, $\gamma_{1,1} = \gamma_{2,2} = 0.9$ $\gamma_{1,2} = \gamma_{2,1} = 0.1$.

Les erreurs idiosyncratiques associées aux équations de mesure suivent des processus autorégressifs d'ordre 1 avec un coefficient commun d. Les innovations des facteurs latents et les chocs idiosyncratiques sont supposés $i.i.d$ selon des lois normales centrées réduites.

L'objectif de cette simulation est d'évaluer la robustesse des approches MIDAS et ADL-MIDAS relativement au filtre de Kalman dans un contexte de mauvaise spécification du modèle.

Table 4: RMSPE out of sample du filtre de Kalman à un facteur par rapport au MIDAS, et du filtre de Kalman à un facteur mal spécifié par rapport au MIDAS

	Panel A: One-Factor DGP, $h = 1$								**Panel B: One-Factor DGP,** $h = 4$							
	Regular MIDAS				Multiplicative MIDAS				Regular MIDAS				Multiplicative MIDAS			
$d\backslash\rho$	-0.9	-0.5	0.5	0.95	-0.9	-0.5	0.5	0.95	-0.9	-0.5	0.5	0.95	-0.9	-0.5	0.5	0.95
-0.9	1.05	1.05	1.04	0.92	1.07	1.09	1.05	0.91	1.12	1.10	1.10	1.15	1.13	1.13	1.11	1.13
-0.5	1.05	1.12	1.13	1.02	1.06	1.15	1.14	1.02	1.09	1.11	1.12	1.23	1.10	1.13	1.15	1.17
0.0	1.04	1.12	1.12	1.04	1.07	1.15	1.13	1.04	1.08	1.11	1.13	1.23	1.10	1.13	1.15	1.16
0.5	1.03	1.12	1.11	1.05	1.04	1.14	1.12	1.04	1.08	1.12	1.13	1.24	1.10	1.14	1.14	1.16
0.95	0.96	1.05	1.08	1.08	0.95	1.04	1.07	1.07	1.19	1.29	1.30	1.31	1.14	1.19	1.20	1.20
	Panel C: Two-Factor DGP, $h = 1$								**Panel D: Two-Factor DGP,** $h = 4$							
	Regular MIDAS				Multiplicative MIDAS				Regular MIDAS				Multiplicative MIDAS			
-0.9	0.92	0.97	0.98	1.47	0.92	0.97	0.98	1.51	1.13	1.15	1.15	1.28	1.12	1.14	1.15	1.25
-0.5	0.99	1.03	1.04	1.19	0.98	1.04	1.04	1.19	1.14	1.18	1.20	1.31	1.13	1.17	1.17	1.24
0.0	1.06	1.04	1.04	1.11	1.05	1.05	1.04	1.12	1.14	1.18	1.19	1.30	1.14	1.17	1.17	1.23
0.5	1.07	1.04	1.05	1.09	1.06	1.05	1.04	1.09	1.15	1.19	1.20	1.30	1.12	1.15	1.17	1.22
0.95	1.26	1.09	1.08	1.07	1.23	1.06	1.06	1.06	1.26	1.27	1.29	1.32	1.20	1.19	1.19	1.20

Étude des résultats :

Pour la **Simulation 1**, les ratios de RMSPE Kalman / MIDAS sont globalement proches de l'unité, ce qui indique que, en moyenne, les performances prédictives du filtre de Kalman et des régressions MIDAS sont très similaires sous les hypothèses considérées. Cette conclusion est robuste pour différentes valeurs de la persistance du facteur (ρ), de la persistance de l'erreur de mesure (d) et pour les horizons de prévision $h = 1$ et $h = 4$.

Ces résultats peuvent paraître surprenants dans la mesure où, dans cette configuration, le filtre de Kalman est correctement spécifié et bénéficie donc d'un avantage structurel. Toutefois, ils soulignent la forte capacité des régressions MIDAS à exploiter l'information.

Le nombre relativement limité de répétitions Monte Carlo ($N = 100$) peut contribuer à l'obtention de ces résultats. Dans le papier, où un nombre plus élevé de simulations est utilisé, les auteurs trouvent également des ratios proches de l'unité, avec des gains maximaux du filtre de Kalman pouvant atteindre environ 20 %.

Pour la **Simulation 2**, les ratios de RMSPE Kalman / MIDAS sont majoritairement supérieurs à 1. Cela indique que, dans ce cadre, les régressions MIDAS fournissent en moyenne des prévisions plus précises que le filtre de Kalman.

Ce résultat s'explique par la mauvaise spécification du filtre de Kalman : le processus générateur de données repose sur deux facteurs latents, alors que le modèle espace-état estimé n'en comporte qu'un seul. Dans cette situation, le filtre de Kalman ne parvient pas à exploiter correctement l'information contenue dans les données de haute fréquence.

À l'inverse, les régressions MIDAS, bien que plus réduites sur le plan structurel, apparaissent plus robustes à cette forme de mauvaise spécification du modèle. Ces résultats sont cohérents avec ceux rapportés dans le papier de référence, où les auteurs montrent que, sous mauvaise spécification du modèle espace-état et en échantillon fini et réduit, les régressions MIDAS peuvent surpasser le filtre de Kalman, avec des gains pouvant atteindre environ 15 % en termes de RMSPE.

5.4 Simulation 3 - SSM (AIC-BIC) vs MIDAS

La troisième simulation considère un environnement dans lequel la dimension de l'espace d'état n'est pas connue a priori et doit être sélectionnée à partir des données, à l'aide des critères AIC / BIC. Le processus générateur de données est identique à celui de la Simulation 2 et repose sur deux facteurs latents définis à la fréquence haute, tandis que le modèle estimé peut donc comporter un nombre variable de facteurs.

Contrairement à la Simulation 2, le nombre de facteurs du modèle SSM estimé n'est pas imposé. Il est sélectionné à l'aide de critères d'information standards, tels que l'AIC et le BIC. Les performances de prévision sont ensuite évaluées conditionnellement au modèle sélectionné.

L'objectif de cette simulation est d'évaluer l'impact de l'incertitude liée à la sélection du nombre de facteurs sur les performances prédictives du filtre de Kalman, et de comparer ces performances à celles des approches regular ADL-MIDAS et multiplicative ADL-MIDAS, qui ne nécessitent pas la spécification explicite de la dimension de l'état latent.

Le papier conclut sur le fait que le facteur AIC a tendance à surajuster et à choisir le modèle à 2 facteurs. Cela nuit aux performances du SSM. Quant au BIC, plus parcimonieux, choisit presque tout le temps le modèle à 1 facteur. Ainsi les performances avec le SSM sous le BIC sont pratiquement tout le temps supérieures aux performances des regressions MIDAS.

Table 5: RMSPE out of sample du filtre de Kalman - Choix AIC - BIC - par rapport au RMSPE

	Panel A: AIC, $h = 1$								**Panel B: BIC,** $h = 1$							
	Regular MIDAS				Multiplicative MIDAS				Regular MIDAS				Multiplicative MIDAS			
$d\backslash\rho$	-0.9	-0.5	0.5	0.95	-0.9	-0.5	0.5	0.95	-0.9	-0.5	0.5	0.95	-0.9	-0.5	0.5	0.95
-0.9	1.05	1.05	1.04	0.91	1.07	1.08	1.05	0.91	1.05	1.05	1.04	0.92	1.07	1.09	1.05	0.91
-0.5	1.05	1.12	1.13	1.02	1.06	1.15	1.14	1.02	1.05	1.12	1.13	1.02	1.06	1.15	1.14	1.02
0.0	1.04	1.12	1.12	1.04	1.07	1.15	1.12	1.04	1.04	1.12	1.12	1.04	1.07	1.15	1.13	1.04
0.5	1.03	1.11	1.11	1.05	1.04	1.13	1.12	1.04	1.03	1.12	1.11	1.05	1.04	1.14	1.12	1.04
0.95	0.93	1.05	1.08	1.08	0.93	1.03	1.07	1.06	0.93	1.05	1.08	1.08	0.93	1.03	1.07	1.07
	Panel C: AIC, $h = 4$								**Panel D: BIC,** $h = 4$							
	Regular MIDAS				Multiplicative MIDAS				Regular MIDAS				Multiplicative MIDAS			
-0.9	1.12	1.10	1.10	1.15	1.13	1.13	1.11	1.13	1.12	1.10	1.10	1.15	1.13	1.13	1.11	1.13
-0.5	1.09	1.11	1.12	1.23	1.10	1.13	1.15	1.17	1.09	1.11	1.12	1.23	1.10	1.13	1.15	1.17
0.0	1.08	1.11	1.13	1.23	1.10	1.13	1.15	1.16	1.08	1.11	1.13	1.23	1.10	1.13	1.15	1.16
0.5	1.08	1.12	1.13	1.24	1.10	1.14	1.14	1.15	1.08	1.12	1.13	1.24	1.10	1.14	1.14	1.16
0.95	1.19	1.28	1.29	1.31	1.14	1.19	1.20	1.19	1.19	1.29	1.30	1.31	1.14	1.19	1.20	1.20

Étude des résultats Les résultats obtenus nous informe que les régressions MIDAS sont plus précises que le filtre de Kalman dans cette configuration. Nous nous attendions à ce résultat. Toutefois, nous remarquons également que le choix du critère d'information AIC ou BIC n'a pas énormement d'impact dans nos simulations. Cela s'explique par le faible nombre de simulations effectuées ($N = 100$). Nous remarquons durant l'exécution du code que le modèle à deux facteurs est très minoritairement choisi par les critères AIC - BIC, presque jamais pour BIC. Ce choix varie si le nombre de simulation augmente. Le papier conclut bien sur le fait que le BIC est plus conservateur et choisi le plus souvent 1 seul facteur donnant une meilleure performance du filtre de Kalman.

6 SSM vs MIDAS (empirical study)

Les données utilisées sont des séries temporelles. Elles ont été transformées afin d'obtenir une stationnarité. Elles sont également normalisées pour pouvoir les comparer. Le PIB suit une fréquence trimestrielle tandis que les autres séries sont au pas de temps mensuel. Veuillez trouver les séries mensuelles, leur période et transformation dans le tableau ci-dessous :

Name	Description	Period	Transformation
TERM	Term spread (10yr T-Bond – 1yr T-bond)	1959:01 – 2023:12	lv
SP500	Stock Price Index: Standard & Poor's 500 Composite	1959:01 – 2023:12	Δ ln
IP	Industrial Production Index (SA)	1959:01 – 2023:12	Δ ln
Emply	All Employees: Total Nonfarm Payroll (SA, Thous)	1959:01 – 2023:12	Δ ln
Exptn	Consumer Expectations (Q1–66 = 100)	1978:01 – 2023:12	ln
PI	Personal Income less Transfer Payments (SAAR)	1959:02 – 2023:12	Δ ln
LEI	Leading index, percent change from previous month	1959:02 – 2023:12	lv
Manu	Real Manufacturing & Trade Inventories: All Industries (SA)	1967:02 – 2023:12	Δ ln
Oil	Crude Oil Spot Price: WTI Cushing	1982:01 – 2023:12	Δ ln

6.1 Les forecasts et le critère d'évaluation

La précision des prévisions est évaluée à l'aide de la racine de l'erreur quadratique moyenne (*Root Mean Squared Error*, RMSE).

$$\text{RMSE}(h) = \sqrt{\frac{1}{T_2 - T_1 - h + 1} \sum_{t=T_1}^{T_2-h} \left(\widehat{Y}_{t+h} - Y_{t+h}\right)^2}$$

Le modèle est estimé sur la période $t \in [1, T_1]$; Les prévisions sont évaluées sur $t \in [T_1 + h, T_2]$. L'horizon de prévision h est estimé en trimestres.

La RMSE mesure l'écart moyen entre les valeurs observées et les prévisions correspondantes. Une valeur plus faible de la RMSE indique une meilleure performance de prévision.
Toutes les séries étant normalisées, les RMSE sont exprimées en unités d'écart-type.

6.2 Résultats de prévision

La Table 7 compare les performances de prévision hors échantillon entre différents modèles à fréquences mixtes. Nous considérons les modèles suivants : **modèle espace d'états** (m0), **le MIDAS régulier** (m1) et **le MIDAS multiplicatif** (m2). Les forecasts se font sur un horizon de prévision d'un trimestre à deux ans ($h \in \{1, 2, 3, 4, 5, 6, 7, 8\}$. Les résultats correspondent aux RMSE que l'on trouve suivant le modèle ainsi que le ratio entre le modèle m0 et les modèles m1, m2. Ces résultats sont présentés pour les différents indicateurs mensuels que nous avons présenté en partie 6.

Les ratios ($m0/m1$) et ($m0/m2$) permettent de comparer directement les performances relatives de nos modèles:

- un ratio inférieur à 1 indique une meilleure performance du SSM
- Ex : un ratio égal à 0.80 correspond à un gain de précision de 20 %

Table 6: RMSE des forecasts de chaque modèle par série

h (Quarters)	1	2	3	4	5	6	7	8
Term Spread								
State Space (m0)	1.18	1.17	1.17	1.17	1.16	1.16	1.16	1.17
Regular Midas (m1)	1.24	1.16	1.12	1.12	1.12	1.11	1.10	1.11
Multiple Midas (m2)	1.24	1.16	1.12	1.12	1.12	1.10	1.10	1.12
Ratio (m0/m1)	0.95	1.01	1.04	1.04	1.04	1.05	1.06	1.05
Ratio (m0/m2)	0.95	1.01	1.04	1.04	1.04	1.06	1.06	1.05
SP 500								
State Space (m0)	1.27	1.21	1.19	1.18	1.18	1.17	1.16	1.17
Regular Midas (m1)	1.20	1.11	1.08	1.09	1.08	1.09	1.10	1.11
Multiple Midas (m2)	1.21	1.11	1.09	1.10	1.11	1.12	1.10	1.14
Ratio (m0/m1)	1.06	1.10	1.10	1.08	1.09	1.07	1.05	1.05
Ratio (m0/m2)	1.04	1.10	1.10	1.08	1.06	1.04	1.06	1.03
Industrial Production								
State Space (m0)	1.28	1.25	1.22	1.21	1.20	1.18	1.17	1.17
Regular Midas (m1)	1.22	1.14	1.11	1.10	1.11	1.12	1.14	1.15
Multiple Midas (m2)	1.15	1.13	1.14	1.11	1.10	1.14	1.10	1.12
Ratio (m0/m1)	1.05	1.10	1.10	1.09	1.07	1.05	1.02	1.02
Ratio (m0/m2)	1.11	1.11	1.08	1.09	1.09	1.03	1.06	1.04
Employment								
State Space (m0)	1.29	1.20	1.19	1.18	1.17	1.17	1.16	1.17
Regular Midas (m1)	1.23	1.13	1.12	1.12	1.12	1.12	1.31	1.14
Multiple Midas (m2)	1.21	1.13	1.11	1.24	1.13	1.31	1.12	1.13
Ratio (m0/m1)	1.05	1.06	1.06	1.05	1.05	1.04	0.89	1.03
Ratio (m0/m2)	1.07	1.06	1.07	0.95	1.04	0.89	1.03	1.04
Expectations								
State Space (m0)	1.18	1.17	1.17	1.17	1.16	1.16	1.16	1.17
Regular Midas (m1)	1.30	1.10	1.08	1.08	1.09	1.10	1.09	1.11
Multiple Midas (m2)	1.29	1.09	1.08	1.08	1.09	1.11	1.09	1.12
Ratio (m0/m1)	0.91	1.06	1.08	1.08	1.07	1.06	1.07	1.06
Ratio (m0/m2)	0.91	1.07	1.08	1.08	1.07	1.05	1.06	1.05
Personal Income								
State Space (m0)	1.25	1.20	1.18	1.17	1.17	1.17	1.16	1.17
Regular Midas (m1)	1.14	1.11	1.10	1.18	1.10	1.12	1.11	1.13
Multiple Midas (m2)	1.14	1.11	1.09	1.10	1.10	1.12	1.10	1.11
Ratio (m0/m1)	1.10	1.08	1.07	1.00	1.06	1.04	1.04	1.04
Ratio (m0/m2)	1.10	1.08	1.09	1.07	1.06	1.04	1.06	1.05
Crude Oil Price								
State Space (m0)	1.54	1.55	1.56	1.56	1.59	1.53	1.50	1.52
Regular Midas (m1)	1.35	1.15	1.11	1.13	1.12	1.14	1.14	1.14
Multiple Midas (m2)	1.36	1.16	1.13	1.13	1.12	1.15	1.13	1.14
Ratio (m0/m1)	1.14	1.35	1.40	1.39	1.42	1.34	1.32	1.33
Ratio (m0/m2)	1.13	1.34	1.38	1.38	1.42	1.34	1.32	1.34
Leading Index								
State Space (m0)	1.27	1.22	1.20	1.20	1.18	1.17	1.17	1.17
Regular Midas (m1)	1.14	1.08	1.07	1.08	1.10	1.10	1.11	1.11
Multiple Midas (m2)	1.14	1.09	1.07	1.08	1.10	1.10	1.11	1.11
Ratio (m0/m1)	1.11	1.13	1.13	1.10	1.08	1.07	1.05	1.05
Ratio (m0/m2)	1.11	1.13	1.12	1.10	1.08	1.07	1.05	1.06
Manufacturing								
State Space (m0)	1.26	1.22	1.20	1.19	1.18	1.17	1.16	1.17
Regular Midas (m1)	1.21	1.11	1.09	1.10	1.10	1.11	1.13	1.15
Multiple Midas (m2)	1.22	1.12	1.10	1.11	1.11	1.12	1.12	1.14
Ratio (m0/m1)	1.04	1.11	1.10	1.08	1.07	1.06	1.03	1.02
Ratio (m0/m2)	1.03	1.10	1.09	1.07	1.07	1.05	1.03	1.02

Étude des résultats :

Les résultats du Tableau 6 montrent que les modèles MIDAS obtiennent dans la plupart des cas des performances de prévision comparables, voire supérieures, au modèle espace d'état à un facteur, comme l'indiquent les ratios de RMSE généralement supérieurs à l'unité. Cette supériorité est plus marquée pour certains indicateurs macroéconomiques (production industrielle, emploi, prix du pétrole), tandis que les écarts sont plus faibles pour d'autres séries, telles que le term spread ou le SP 500.

Les résultats ne coïncident toutefois pas exactement avec ceux rapportés dans l'article original. Cette différence peut s'expliquer par l'horizon d'évaluation hors échantillon : alors que l'étude originale s'arrête en 2009, notre analyse prolonge la période de prévision jusqu'en 2023. L'inclusion de phases macroéconomiques récentes, marquées par des chocs importants et des changements structurels, est susceptible d'affecter les performances relatives des modèles et de générer des écarts par rapport aux résultats initiaux.

6.3 Meilleurs estimateurs

Table 7: Meilleurs estimateurs par modèle et RMSE associées

h (Quarter)	model	1	2	3	4	5	6	7	8
Best Predictor	State Space	Term	Term	Term	Term	Exptn	Term	Emply	Term
	Regular MIDAS	PI	LEI	LEI	Exptn	SP	SP	Exptn	Exptn
	Multiplicative MIDAS	PI	LEI	LEI	Exptn	Exptn	Term	Exptn	LEI
RMSE	State Space	1.18	1.17	1.17	1.17	1.16	1.16	1.16	1.17
	Regular MIDAS	1.14	1.08	1.07	1.08	1.08	1.09	1.09	1.11
	Multiplicative MIDAS	1.14	1.09	1.07	1.08	1.09	1.10	1.09	1.11

Étude des résultats :

Le Tableau 7 présente, pour chaque horizon de prévision, les meilleurs prédicteurs sélectionnés par modèle ainsi que les RMSE associées. Les résultats indiquent que les spécifications MIDAS, régulière et multiplicative, sélectionnent fréquemment des variables de type prix ou anticipations, et atteignent des niveaux de RMSE systématiquement inférieurs à ceux du modèle espace d'état à un facteur. Les différences entre MIDAS régulier et multiplicatif restent toutefois limitées, leurs performances étant très proches sur l'ensemble des horizons.

Comme pour le tableau précédent, nos résultats ne reproduisent pas exactement ceux de l'étude originale. Ces écarts peuvent s'expliquer par l'extension de l'échantillon de prévision jusqu'en 2023, alors que l'article initial s'arrête en 2009. L'intégration de périodes récentes, caractérisées par des chocs macroéconomiques majeurs et des changements structurels, est susceptible d'influencer la sélection des prédicteurs et les performances relatives des modèles.

7 Conclusion

Ce papier avait pour objectif de reproduire et de discuter les résultats centraux de *Bai Ghysels Wright 2013*, qui comparent deux manières d'exploiter des données à fréquences mixtes pour prévoir une variable basse fréquence : (i) un SSM estimé via filtre de Kalman et (ii) des régressions MIDAS, sous forme régulière et multiplicative. L'enjeu principal est de déterminer dans quelle mesure des spécifications MIDAS (plus simples à mettre en œuvre) peuvent approcher l'opérateur de prévision optimal fourni par le filtre de Kalman lorsque le SSM est correctement spécifié.

Les exercices de simulation reproduits dans ce rapport confirment le message principal des auteurs. Dans le cas d'un modèle à un facteur correctement spécifié, les poids de prévision induits par les régressions MIDAS sont très proches de ceux du filtre de Kalman sur une large grille de paramètres (d, ρ), ce qui se traduit par des valeurs faibles de la distance quadratique en norme L^2. Cette proximité est particulièrement marquée lorsque la fréquence relative des données haute fréquence augmente, ce qui est cohérent avec l'interprétation théorique selon laquelle les régressions MIDAS peuvent être vues comme une approximation paramétrée d'une représentation en moyenne mobile des prévisions de Kalman.

La comparaison des spécifications MIDAS met également en évidence, comme dans l'article original, un avantage récurrent des formes multiplicatives (ADL-MIDAS) sur les formes régulières, dans la mesure où la structure multiplicative capture plus fidèlement la périodicité et la dynamique intra-période implicites dans les gains de Kalman. Toutefois, l'ordre de grandeur des écarts demeure généralement modéré : dans la plupart des configurations, les deux familles de régressions MIDAS fournissent une approximation très précise des poids de Kalman, ce qui renforce l'idée d'une grande robustesse de ces modèles.

L'extension au cas de plusieurs séries haute fréquence confirme également les conclusions des auteurs : les schémas MIDAS de combinaison de prévisions parviennent à agréger efficacement l'information contenue dans plusieurs indicateurs, y compris lorsque la qualité informationnelle des séries diffère (variance de bruit inégale). Les résultats montrent que MIDAS peut apprendre une hiérarchie de pondérations analogue à celle induite par le filtre de Kalman, sans imposer explicitement une structure factorielle complète.

Au total, cette reproduction corrobore l'argument de *Bai Ghysels Wright 2013* selon lequel les régressions MIDAS constituent une alternative empirique crédible aux SSM, surtout dans des contextes appliqués où la spécification d'un SSM complet, la dimension de l'état latent et l'estimation de nombreux paramètres peuvent être coûteuses ou incertaines. Si le filtre de Kalman demeure la référence lorsque le modèle est correctement spécifié et que ses paramètres sont connus, les régressions MIDAS (et en particulier l'ADL-MIDAS) apparaissent comme des outils flexibles et performants pour la prévision à fréquences mixtes, capables d'approcher de très près l'opérateur de prévision optimal.