

Estimation en temps réel de la tendance-cycle : apport de l'utilisation des filtres asymétriques dans la détection des points de retournement

Alain QUARTIER-LA-TENTE

(*) Insee, Département des études économiques (DEE), et Laboratoire d'Économie et de Management de Nantes-Atlantique (LEMNA)

alain.quartier-la-tente@insee.fr

Mots-clés. séries temporelles, tendance-cycle, désaisonnalisation, *nowcasting*.

Domaines. 10.2 Désaisonnalisation — 10.3 Nowcasting.

Résumé

L'analyse des cycles économiques, et en particulier la détection précoce des points de retournement, est un sujet majeur dans l'analyse de la conjoncture. Les moyennes mobiles, ou les filtres linéaires, sont omniprésents dans les méthodes d'extraction de la tendance-cycle et d'ajustement saisonnier. Au centre de la série, des moyennes mobiles symétriques sont appliquées¹. Cependant, en raison du manque d'observations futures, les estimations en temps réel doivent s'appuyer sur des moyennes mobiles asymétriques. C'est ce qui est par exemple fait dans les méthodes de désaisonnalisation les plus utilisées, TRAMO-SEATS et X-13ARIMA, qui prolongent la série sur 1 an par un modèle ARIMA. Les prévisions étant des combinaisons linéaires du passé, cela revient en réalité à utiliser des moyennes mobiles asymétriques dont les coefficients sont optimisés par rapport à la prévision avec une longueur d'avance.

La construction de moyennes mobiles asymétriques performantes en termes de fidélité (préservation du signal), de révision, de lissage et de déphasage (délais dans la détection de points de retournement) est un sujet de recherche toujours ouvert. Cette étude décrit et compare des approches récentes pour la construction de moyennes mobiles asymétriques, utilisées pour l'estimation en temps réel de la tendance-cycle : filtres polynomiaux locaux (PROIETTI et LUATI 2008 ; GRAY et THOMSON 1996) ; méthodes basées sur une optimisation sous contrainte d'une somme pondérée de critères de qualité des moyennes mobiles (GRUN-REHOMME, GUGGEMOS et LADIRAY 2018 ; WILDI et MCELROY 2019) ; et filtres basés sur les espaces de Hilbert à noyau reproduisant (RKHS) DAGUM et BIANCONCINI (2008). Elle montre également comment les filtres polynomiaux locaux peuvent être étendus pour inclure un critère de temporalité afin

1. C'est-à-dire que pour estimer la tendance-cycle à la date t , on utilise autant de points dans le passé que dans le futur et le même poids est associé aux observations passées et futures.

de minimiser le déphasage. Enfin, cette étude montre qu'il est possible d'établir une approche unificatrice générale qui permet de reproduire l'ensemble des méthodes étudiées.

La comparaison des méthodes sur séries simulées montre qu'il est important d'adapter la longueur du filtre à la variabilité de la série. Même si certains filtres RKHS pourraient donner des résultats satisfaisant en termes de délais dans la détection des points de retournement, leur utilisation doit être limitée aux premières estimations de la tendances-cycles. Lorsque l'on se rapproche de l'estimation finale (i.e., lorsque beaucoup de points dans le futur sont connus), d'autres méthodes devraient être utilisées afin de limiter les révisions. Les filtres RKHS sont par ailleurs sujets à problèmes d'optimisation. Enfin, lorsque la longueur du filtre est adapté à la variabilité de la série, chercher à conserver des tendances polynomiales de degré supérieur à un semble introduire de la variance dans les estimations (et donc plus de révisions) sans gain significatif en termes de détection de point de retournement.

Cette étude est reproductible. En particulier, toutes les méthodes décrites sont implémentées dans le *package*  `rjdfilters` (<https://github.com/palatej/rjdfilters>) et tous les codes utilisés sont disponibles sous <https://github.com/AQLT/articles>.

Abstract

This paper describes and compares different approaches to build asymmetric filters: local polynomials filters, methods based on an optimization of filters' properties (Fidelity-Smoothness-Timeliness, FST, approach and a data-dependent filter) and filters based on Reproducing Kernel Hilbert Space. It also describes how local polynomials filters can be extended to include a timeliness criterion to minimize phase shift. All these methods can be seen as a special case of a general unifying framework to derive linear filters.

This paper shows that, when the length of the filter is adapted to the variability of the series, constraining asymmetric filters to preserve constant trends (and not necessarily polynomial ones) reduce revision error and time lag. Therefore, future studies on the subject can focus on these filters. Moreover, with RKHS filters some optimisation issues can occurs and they might lead to erratic estimation. They might be able to produce satisfying results in terms of phase-shift and revisions, but they should be avoid for the last estimates of the trend-cycle component: other methods should then be preferred to reduce revisions with the final estimates.

All the methods are implemented in the  *package* `rjdfilters` and the results can be easily reproduced. The programs used, and a web version of this report, are available at <https://github.com/AQLT/articles>.

Table des matières

Résumé	1
Abstract	2
Introduction	4
1 De la désaisonnalisation à l'estimation tendance-cycle	5

2	Quelques propriétés sur les moyennes mobiles	6
2.1	Gain et fonction de déphasage	7
2.2	Propriétés souhaitables d'une moyenne mobile	7
2.2.1	Préservation de tendances	8
2.2.2	Réduction du bruit	8
2.3	Estimation en temps réel et moyennes mobiles asymétriques	8
2.3.1	Moyennes mobiles asymétriques et prévision	9
2.3.2	Indicateurs de qualité des moyennes mobiles asymétriques	10
3	D'une théorie générale sur la construction des filtres asymétriques à l'approche FST	10
3.1	Théorie générale de construction des filtres asymétriques	10
3.2	Approche <i>Fidelity-Smoothness-Timeliness</i> (FST)	11
4	Filtres dépendant des données : trilemme ATS	13
5	Régression non paramétrique et régression polynomiale locale	16
5.1	Régression polynomiale : approche de Proietti et Luati	16
5.1.1	Filtres symétriques	16
5.1.2	Les différents noyaux	18
5.1.3	Quelques filtres symétriques particuliers	19
5.1.4	Filtres asymétriques	19
5.2	Extension avec le critère de <i>timeliness</i>	22
5.3	Régression polynomiale : Gray et Thomson	23
5.3.1	Filtres symétriques	23
5.3.2	Filtres asymétriques	24
5.4	Reproducing Kernel Hilbert Space (RKHS) : approche de Dagum et Bianconcini	25
5.5	Liens entre les différentes méthodes	29
5.5.1	Critères de Gray et Thomson et ceux de Grun-Rehomme <i>et alii</i>	29
5.5.2	Équivalence avec les moindres carrés pondérés	30
5.5.3	RKHS et polynômes locaux	31
6	Comparaison des différentes méthodes	31
6.1	Méthodologie	32
6.2	Séries simulées	33
6.2.1	Comparaison des filtres polynomiaux locaux et des filtres RKHS	33
6.2.2	Comparaison avec l'approche FST	35
6.3	Série réelle	36
	Conclusion	40
A	Synthèse des liens entre les différentes méthodes de construction de moyennes mobiles	42
B	Coefficients, fonctions de gain et de déphasage	44
C	Équivalence entre l'approche FST et les filtres polynomiaux locaux	48
D	Prévisions implicites pour séries RETAILx	49

Introduction

L'analyse du cycle économique, et en particulier la détection rapide des points de retournement d'une série, est un sujet de première importance dans l'analyse de la conjoncture économique. Pour cela, les indicateurs économiques sont généralement corrigés des variations saisonnières. Toutefois, afin d'améliorer leur lisibilité, il peut être nécessaire d'effectuer un lissage supplémentaire afin de réduire le bruit, et ainsi analyser la composante tendance-cycle. Par construction, les méthodes d'extraction de tendance-cycle sont étroitement liées aux méthodes de désaisonnalisation puisqu'elles sont généralement appliquées sur des séries corrigées des variations saisonnières.

Les moyennes mobiles, ou les filtres linéaires, sont omniprésents dans les méthodes d'extraction du cycle économique et d'ajustement saisonnier². Ainsi, la méthode de désaisonnalisation X-13ARIMA-SEATS utilise des moyennes mobiles de Henderson et des moyennes mobiles composites pour estimer les principales composantes d'une série chronologique, tandis que TRAMO-SEATS utilise des filtres de Wiener-Kolmogorov. Au centre de la série, des filtres symétriques sont appliqués. En revanche, en raison du manque d'observations futures, pour estimer les points les plus récents, toutes ces méthodes doivent s'appuyer sur des filtres asymétriques. Par exemple, même si X-13ARIMA-SEATS et TRAMO-SEATS appliquent des moyennes symétriques aux prévisions obtenues à partir d'un modèle ARIMA, cela revient à appliquer des filtres asymétriques en fin de série, car les valeurs prédites sont des combinaisons linéaires de valeurs passées.

Si ces moyennes mobiles asymétriques ont de bonnes propriétés concernant la taille des révisions futures induites par le processus de lissage³, elles induisent également des déphasages qui retardent en général la détection en temps réel des points de retournement.

L'objectif de cette étude est de décrire et de comparer les approches récentes permettant l'extraction de tendance-cycle : filtres polynomiaux locaux, méthodes basées sur une optimisation des propriétés des filtres et filtres basés sur les espaces de Hilbert à noyau reproduisant (RKHS). Toutes ces méthodes sont implémentées dans le *package* `R` `rjdfilters`⁴, tous les résultats de cette étude sont facilement reproductibles. En raison du lien entre la désaisonnalisation et l'extraction de tendance-cycle (section 1), nous nous concentrons sur les méthodes non paramétriques qui peuvent être incluses dans X-13ARIMA-SEATS. Après une description des propriétés générales d'un filtre linéaire (section 2), nous décrivons une approche générale qui permet d'englober les différentes méthodes développées par PROIETTI et LUATI (2008), GRAY et THOMSON (1996), GRUN-REHOMME, GUGGEMOS et LADIRAY (2018), WILDI et MCELROY (2019) et DAGUM et BIANCONCINI (2008) (sections 3 à 5) et nous mettrons également en exergue les liens théoriques entre ces différentes variables. Enfin, dans la section 6, nous comparons les différentes méthodes en termes de déphasage et de révisions, en les appliquant sur des séries simulées et réelles.

2. Une moyenne mobile est une méthode statistique qui consiste à appliquer une moyenne pondérée glissante à une série temporelle : à chaque date t on calcule une moyenne pondérée de p points passés et q points futurs où $p, q \geq 0$ dépend de la moyenne mobile.

3. Voir par exemple PIERCE (1980).

4. <https://github.com/palatej/rjdfilters>.

1 De la désaisonnalisation à l'estimation tendance-cycle

La plupart des indicateurs macroéconomiques (PIB, production, consommation, etc.) sont affectés par des effets saisonniers et des effets jours ouvrables qui perturbent l'analyse des évolutions infra-annuelles et les comparaisons spatiales. C'est pourquoi les séries chronologiques sont généralement corrigées des variations saisonnières et des jours ouvrables, la *désaisonnalisation* étant le processus consistant à supprimer leurs effets.

Pour effectuer la désaisonnalisation, les méthodes de désaisonnalisation les plus populaires sont TRAMO-SEATS, une méthode paramétrique basée sur les modèles ARIMA (voir par exemple MARAVALL et CAPORELLO (2004)), et X-13ARIMA-SEATS, une méthode non-paramétrique basée sur les moyennes mobiles (voir par exemple LADIRAY et QUENNEVILLE (2011)). Ces méthodes supposent que toute série temporelle X_t peut se décomposer en quatre composantes :

1. Une composante saisonnière S_t .
2. Une composante jours ouvrables D_t .
3. Une composante tendance-cycle TC_t qui contient la tendance (qui représente les évolutions de long terme) et le cycle (qui représente les évolutions cycliques autour de la tendance). Dans ce cas, la tendance et le cycle sont estimés conjointement : les deux composantes peuvent en effet être difficiles à séparer (comment différencier les cycles courts des cycles longs ?) et les méthodes étudiées sont souvent appliquées sur des courtes (les cycles longs sont alors difficilement estimables).
4. Une composante irrégulière I_t qui contient toutes les autres fluctuations.

Toutes ces composantes étant inobservées, l'estimation de l'une dépend de l'estimation des autres. Ainsi, même si dans ce rapport, nous nous intéresserons aux méthodes d'extraction de tendance-cycle, celles-ci ne peuvent s'étudier indépendamment du processus de désaisonnalisation. Ce lien explique également que toutes les méthodes utilisées dans ce rapport sont implémentées dans les bibliothèques de Java de JDemetra+⁵, le logiciel de désaisonnalisation recommandé par Eurostat. Une interface  à l'ensemble des a été développée autour des méthodes présentées dans cette étude, grâce au *package* `rjdfilters`⁶.

Les filtres linéaires (ou moyennes mobiles) sont omniprésents dans la désaisonnalisation et l'estimation des différentes composantes. Au centre de la série, des filtres dits symétriques sont appliqués (pour estimer une composante à la date t , on utilise autant de points après t qu'avant t). Pour extraire la tendance-cycle, le filtre symétrique le plus connu est celui de HENDERSON (1916), notamment utilisé dans l'algorithme X-13ARIMA. Cependant, en raison du manque d'observations futures, les estimations en temps réel doivent s'appuyer sur des moyennes mobiles asymétriques. Les moyennes mobiles asymétriques classiques minimisent les erreurs de révision mais introduisent des retards dans la détection des points de retournement (appelé déphasage, voir section 2).

Dans la littérature, différentes approches ont été envisagées pour l'extraction de tendance-cycle en temps réel⁷. Parmi les plus récentes, on peut citer :

5. Voir <https://github.com/jdemetra/jdemetra-core>.

6. Disponible sur <https://github.com/palatej/rjdfilters>.

7. Voir par exemple ALEXANDROV et alii (2012) pour une revue de la littérature sur les méthodes d'extraction de tendance.

- Les *Model-Based Approach* — approches basées sur les modèles — supposent la spécification d’un modèle stochastique pour la tendance (modèle ARIMA, modèle d’espace d’état, etc.) et les estimations sont obtenues en minimisant une fonction de pénalité, généralement l’erreur quadratique moyenne. C’est par exemple le cas du filtre de Kalman, du filtre de Wiener-Kolmogorov (utilisé dans TRAMO-SEATS) et de l’Approche par Filtre Direct de WILDI et MCELROY (2019) (section 4).
- Les méthodes d’extraction non paramétriques ne supposent pas que la structure d’un modèle est fixe et peuvent être facilement appliquées à n’importe quelle série temporelle. C’est par exemple le cas des filtres d’Henderson et de MUSGRAVE (1964) (utilisés dans X-13ARIMA). Les méthodes classiques peuvent être vues comme des régressions polynomiales locales, approche généralisée par PROIETTI et LUATI (2008) (section 5.1). Les estimateurs non paramétriques peuvent également être reproduits en exploitant la méthodologie de l’espace de Hilbert du noyau reproducteur (RKHS), comme cela est fait par DAGUM et BIANCONCINI (2008) (section 5.4).

Bien que ces auteurs aient proposé des approches générales pour construire des filtres linéaires, ils ne font référence qu’aux méthodes les plus classiques (Henderson, Musgrave, Hodrick-Prescott, etc.) sans faire le lien avec les autres méthodes récentes. Dans cette étude, nous proposons une approche unificatrice générale qui permettrait de reproduire l’ensemble de ces méthodes. Cela a un double intérêt. D’une part, cela permet de faire une première revue de la littérature sur les méthodes de construction des filtres linéaires pour l’analyse conjoncturelle. D’autre part, cela permet de montrer les liens entre toutes ces approches et de les comparer en utilisant une même méthodologie.

Pour cela nous utiliserons l’approche générale de GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) (section 3.1) qui ont également proposé une procédure globale pour construire des moyennes mobiles asymétriques permettant de minimiser les effets de déphasage (section 3.2). Des diagrammes synthétiques des liens entre les différentes méthodes étudiées sont présentés dans l’annexe A.

Nous nous intéresserons uniquement aux estimations intermédiaires de la tendance-cycle faites en temps réels. Le filtre symétrique utilisé pour les estimations finales ne sera pas remis en cause et sera celui d’Henderson. Nous nous concentrons également sur les méthodes qui pourraient être implémentées dans X-13ARIMA. Pour maintenir la cohérence avec l’approche non paramétrique de X-13ARIMA, nous nous concentrons sur les méthodes d’extraction non paramétriques. C’est pourquoi ni les filtres de l’approche de WILDI et MCELROY (2019) (section 4), ni d’autres approches basées sur des modèles ne sont, pour l’instant, utilisées dans les simulations. Même si les différentes méthodes sont comparées en les appliquant sur des séries déjà désaisonnalisées, elles peuvent être intégrées dans l’algorithme de désaisonnalisation X-11 grâce à la fonction `rjdfilters::x11()`. L’ensemble des filtres utilisés sont résumés dans l’annexe B.

2 Quelques propriétés sur les moyennes mobiles

Cette section présente les définitions et les propriétés des moyennes mobiles utiles pour comprendre les méthodes présentées dans les prochaines sections. Pour plus de détails sur les moyennes mobiles, voir par exemple LADIRAY (2018).

Soient deux entiers p et f . Une *moyenne mobile* M_θ ou M est un opérateur linéaire défini par un

ensemble de coefficients $\theta = (\theta_{-p}, \dots, \theta_f)'$ qui transforme toute série temporelle X_t en :

$$M_\theta(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k}.$$

On a les définitions suivantes :

- La quantité $p + f + 1$ est appelée *ordre de la moyenne mobile*.
- Lorsque $p = f$ la moyenne mobile est dite *centrée*. Si de plus on a $\forall k : \theta_{-k} = \theta_k$, la moyenne mobile M_θ est dite *symétrique*. Dans ce cas, la quantité $h = p = f$ est appelée *fenêtre (bandwidth)*.

2.1 Gain et fonction de déphasage

Soit $X_t = e^{-i\omega t}$ avec $\omega \in [0, \pi]$. La moyenne mobile M_θ transforme X_t en :

$$Y_t = M_\theta X_t = \sum_{k=-p}^{+f} \theta_k e^{-i\omega(t+k)} = \left(\sum_{k=-p}^{+f} \theta_k e^{-i\omega k} \right) \cdot X_t.$$

La fonction $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{-i\omega k}$ est appelée *fonction de transfert* ou *fonction de réponse en fréquence (frequency response function)*⁸. Elle peut être réécrite en :

$$\Gamma_\theta(\omega) = \rho_\theta(\omega) e^{-i\varphi_\theta(\omega)},$$

où $\rho_\theta(\omega) = G_\theta(\omega) = |\Gamma_\theta(\omega)|$ est la fonction de *gain* ou *d'amplitude* et $\varphi_\theta(\omega)$ est le *déphasage (phase shift* ou *time shift)*⁹. Pour tous les filtres symétriques on a $\Phi_\theta(\omega) \equiv 0$ (*modulo* π).

En somme, appliquer une moyenne mobile à une série harmonique la modifie de deux façons :

- en la multipliant par un coefficient égal à $\rho_\theta(\omega)$ (gain) ;
- en la « décalant » dans le temps de $\varphi_\theta(\omega)/\omega$, ce qui a un impact sur la détection des points de retournement (déphasage)¹⁰.

2.2 Propriétés souhaitables d'une moyenne mobile

Pour décomposer une série temporelle en une composante saisonnière, une tendance-cycle et l'irrégulier, l'algorithme de décomposition X-11 (utilisé dans X-13ARIMA) utilise une succession de moyennes mobiles ayant toutes des contraintes spécifiques. Dans cette sous-section nous décrivons deux types de contraintes :

- la préservation de certaines tendances ;
- la réduction du bruit.

8. La fonction de transfert peut être définie de manière équivalente par $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{i\omega k}$ ou $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{2\pi i \omega k}$.

9. Cette fonction est parfois définie comme $\varphi_\theta(\omega) = \frac{\varphi_\theta(\omega)}{\omega}$ pour mesurer le déphasage en termes de période.

10. Lorsque $\varphi_\theta(\omega)/\omega > 0$ le déphasage est positif : le point de retournement est détecté avec retard.

2.2.1 Préservation de tendances

Il est souvent souhaitable qu'une moyenne mobile conserve certaines tendances. Une moyenne mobile M_θ conserve une fonction du temps $f(t)$ si $\forall t : M_\theta f(t) = f(t)$.

Nous avons les propriétés suivantes pour la moyenne mobile M_θ :

- Pour conserver les constantes $X_t = a$ il faut que

$$\forall t : M_\theta(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k} = \sum_{k=-p}^{+f} \theta_k a = a \sum_{k=-p}^{+f} \theta_k = a.$$

C'est-à-dire qu'il faut que la somme des coefficients $\sum_{k=-p}^{+f} \theta_k$ soit égale à 1.

- Pour conserver les tendances linéaires $X_t = at + b$ il faut que :

$$\forall t : M_\theta(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k} = \sum_{k=-p}^{+f} \theta_k [a(t+k) + b] = a \sum_{k=-p}^{+f} k \theta_k + (at + b) \sum_{k=-p}^{+f} \theta_k = at + b.$$

Ce qui est équivalent à :

$$\sum_{k=-p}^{+f} \theta_k = 1 \quad \text{et} \quad \sum_{k=-p}^{+f} k \theta_k = 0.$$

- De manière générale, M_θ conserve les tendances de degré d si et seulement si :

$$\sum_{k=-p}^{+f} \theta_k = 1 \quad \text{et} \quad \forall j \in \llbracket 1, d \rrbracket : \sum_{k=-p}^{+f} k^j \theta_k = 0.$$

- Si M_θ est symétrique ($p = f$ et $\theta_{-k} = \theta_k$) et conserve les tendances de degré $2d$ alors elle conserve aussi les tendances de degré $2d + 1$.

2.2.2 Réduction du bruit

Toutes les séries temporelles sont affectées par du bruit qui peut brouiller l'extraction du signal. C'est pourquoi on cherche à réduire ce bruit (en réduisant sa variance) tout en conservant le signal (en utilisant les propriétés vues dans les sections précédentes). La somme des carrés des coefficients $\sum_{k=-p}^{+f} \theta_k^2$ est le rapport de *réduction de la variance*.

En effet, soit $\{\varepsilon_t\}$ une suite de variables aléatoires indépendantes avec $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{V}[\varepsilon_t] = \sigma^2$. On a :

$$\mathbb{V}[M_\theta \varepsilon_t] = \mathbb{V}\left[\sum_{k=-p}^{+f} \theta_k \varepsilon_{t+k}\right] = \sum_{k=-p}^{+f} \theta_k^2 \mathbb{V}[\varepsilon_{t+k}] = \sigma^2 \sum_{k=-p}^{+f} \theta_k^2.$$

2.3 Estimation en temps réel et moyennes mobiles asymétriques

Pour les filtres symétriques, la fonction de déphasage est égale à zéro (modulo π). Il n'y a donc aucun retard dans la détection de points de retournement. Ils ne peuvent toutefois pas être utilisés au début et à la fin de la série car aucune valeur passée/future n'est connue.

2.3.1 Moyennes mobiles asymétriques et prévision

En début et en fin de série, les moyennes mobiles asymétriques ne peuvent être utilisées du fait du manque de données disponibles. Une solution est de prolonger la série par prévision pour ensuite appliquer le filtre symétrique. Cette méthode semble remonter à DE FOREST (1877) qui suggère également de modéliser en fin de période une tendance polynomiale de degré au plus trois :

« As the first m and last m terms of the series cannot be reached directly by the formula, the series should be graphically extended by m terms at both ends, first plotting the observations on paper as ordinates, and then extending the curve along what seems to be its probable course, and measuring the ordinates of the extended portions. It is not necessary that this extension should coincide with what would be the true course of the curve in those parts. The important point is that the m terms thus added, taken together with the $m + 1$ adjacent given terms, should follow a curve whose form is approximately algebraic and of a degree not higher than the third. »

C'est également l'approche utilisée dans les méthodes de désaisonnalisation TRAMO-SEATS et X-13ARIMA qui prolongent la série sur 1 an par un modèle ARIMA. In fine, cela revient à utiliser des moyennes mobiles asymétriques puisque les prévisions sont des combinaisons linéaires du passé !

Inversement, à partir d'une moyenne mobile symétrique de référence, on peut déduire les prévisions implicites d'une moyenne mobile asymétrique. Notons $v = (v_{-h}, \dots, v_h)$ la moyenne mobile symétrique de référence et w^0, \dots, w^{h-1} une suite de moyennes mobiles asymétriques, d'ordre $h + 1$ à $2h$ utilisée pour l'estimation des h derniers points avec, pour convention, $w_t^q = 0$ pour $t > q$. C'est-à-dire que $w^0 = (w_{-h}^0, \dots, w_0^0)$ est utilisée pour l'estimation en temps réel (lorsque l'on ne connaît aucun point dans le futur), $w^1 = (w_{-h}^1, \dots, w_1^1)$ pour l'estimation de l'avant-dernier point (lorsque l'on ne connaît qu'un point dans le futur), etc. Notons également y_{-h}, \dots, y_0 la série étudiée observée et y_1^*, \dots, y_h^* la prévision implicite induite par w^0, \dots, w^{h-1} . Cela signifie, que pour tout q on a :

$$\forall q, \quad \underbrace{\sum_{i=-h}^0 v_i y_i + \sum_{i=1}^h v_i y_i^*}_{\text{lissage par } v \text{ de la série prolongée}} = \underbrace{\sum_{i=-h}^0 w_i^q y_i + \sum_{i=1}^h w_i^q y_i^*}_{\text{lissage par } w^q \text{ de la série prolongée}} .$$

Ce qui est équivalent à :

$$\forall q, \quad \sum_{i=1}^h (v_i - w_i^q) y_i^* = \sum_{i=-h}^0 (w_i^q - v_i) y_i .$$

En somme, matriciellement, cela revient donc à résoudre :

$$\begin{pmatrix} v_1 & v_2 & \dots & v_h \\ v_1 - w_1^1 & v_2 & \dots & v_h \\ \vdots & \vdots & \dots & \vdots \\ v_1 - w_1^{h-1} & v_2 - w_2^{h-1} & \dots & v_h \end{pmatrix} \begin{pmatrix} y_1^* \\ \vdots \\ y_h^* \end{pmatrix} = \begin{pmatrix} w_{-h}^0 - v_{-h} & w_{-(h-1)}^0 - v_{-(h-1)} & \dots & w_0^0 - v_0 \\ w_{-h}^1 - v_{-h} & w_{-(h-1)}^1 - v_{-(h-1)} & \dots & w_0^1 - v_0 \\ \vdots & \vdots & \dots & \vdots \\ w_{-h}^{h-1} - v_{-h} & w_{-(h-1)}^{h-1} - v_{-(h-1)} & \dots & w_0^{h-1} - v_0 \end{pmatrix} \begin{pmatrix} y_{-h} \\ \vdots \\ y_0 \end{pmatrix} .$$

C'est ce qui implémenté dans la fonction `rjdfilters::implicit_forecast`.

Comme notamment souligné par WILDI et SCHIPS (2004), étendre la série par prévision d'un modèle ARIMA revient à calculer des filtres asymétriques dont les coefficients sont optimisés par rapport à la prévision avec une longueur d'avance — *one-step ahead forecasting*. Autrement dit, on cherche à minimiser les révisions entre la première et la dernière estimation (avec le filtre symétrique). Cependant, puisque les coefficients du filtre symétrique décroissent lentement, il faudrait également s'intéresser à la performance des prévisions avec plusieurs longueurs d'avance — *multi-step ahead forecasting*. Par ailleurs, le déphasage induit par les filtres asymétriques n'est pas contrôlé : on pourrait préférer avoir une détection plus rapide des points de retournement et une révision plus grande plutôt que de juste minimiser les révisions entre la première et la dernière estimation. C'est pourquoi il peut être nécessaire de définir des critères alternatifs pour juger la qualité des moyennes mobiles asymétriques.

2.3.2 Indicateurs de qualité des moyennes mobiles asymétriques

Pour les filtres asymétriques, la majorité des critères proviennent de ceux définis par GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) et WILDI et MCELROY (2019) pour construire les filtres asymétriques. Ils sont résumés dans le tableau 1 et calculables avec la fonction `rjdfilters::diagnostic_matrix`.

GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) proposent une approche générale pour dériver des filtres linéaires, basée sur un problème d'optimisation de trois critères : *Fidelity* (F_g , réduction de la variance), *Smoothness* (S_g , lissage) et *Timeliness* (T_g , déphasage). Voir section 3.2 pour plus de détails.

WILDI et MCELROY (2019) proposent une approche basée sur la décomposition de l'erreur quadratique moyenne entre le filtre symétrique et le filtre asymétrique en quatre quantités : *Accuracy* (A_w , précision), *Timeliness* (T_w , déphasage), *Smoothness* (S_w , lissage) et *Residual* (R_w , résidus). Voir section 4 pour plus de détails.

3 D'une théorie générale sur la construction des filtres asymétriques à l'approche FST

3.1 Théorie générale de construction des filtres asymétriques

Pour établir une théorie générale englobant les principaux filtres linéaires, GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) définissent deux critères. En changeant légèrement les notations utilisées par les auteurs afin d'avoir une formulation plus générale, ces deux critères peuvent s'écrire :

$$I(\theta, q, y_t, u_t) = \mathbb{E} [(\Delta^q M_\theta y_t - u_t)^2] \quad (1)$$

$$J(\theta, f, \omega_1, \omega_2) = \int_{\omega_1}^{\omega_2} f[\rho_\theta(\omega), \varphi_\theta(\omega), \omega] d\omega \quad (2)$$

où y_t est la série étudiée, u_t une série de référence et Δ est l'opérateur différence ($\Delta y_t = y_t - y_{t-1}$ et $\Delta^q = \underbrace{\Delta \circ \dots \circ \Delta}_{q \text{ fois}}$ pour $q \in \mathbb{N}$). Dans la majorité des cas, la fonction f ne dépendra que de la

fonction de gain, ρ_θ , et de la fonction de déphasage, φ_θ . Dans ce cas, par simplification on écrira $f[\rho_\theta(\omega), \varphi_\theta(\omega), \omega] = f[\rho_\theta(\omega), \varphi_\theta(\omega)]$.

TABLE 1 – Critères de qualité d’une moyenne mobile $\theta = (\theta_k)_{-p \leq k \leq f}$ définie par une fonction de gain ρ_θ et une fonction de déphasage φ_θ .

Sigle	Description	Formule
b_c	Biais constant	$\sum_{k=-p}^{+f} \theta_k - 1$
b_l	Biais linéaire	$\sum_{k=-p}^{+f} k \theta_k$
b_q	Biais quadratique	$\sum_{k=-p}^{+f} k^2 \theta_k$
F_g	Réduction de la variance / Fidelity (Guggemos)	$\sum_{k=-p}^{+f} \theta_k^2$
S_g	Smoothness (Guggemos)	$\sum_j (\nabla^3 \theta_j)^2$
T_g	Timeliness (Guggemos)	$\int_0^{\omega_1} \rho_\theta(\omega) \sin(\varphi_\theta(\omega))^2 d\omega$
A_w	Accuracy (Wildi)	$2 \int_0^{\omega_1} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) d\omega$
T_w	Timeliness (Wildi)	$8 \int_0^{\omega_1} \rho_s(\omega) \rho_\theta(\omega) \sin^2\left(\frac{\varphi_s(\omega) - \varphi_\theta(\omega)}{2}\right) h(\omega) d\omega$
S_w	Smoothness (Wildi)	$2 \int_{\omega_1}^{\pi} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) d\omega$
R_w	Residual (Wildi)	$8 \int_{\omega_1}^{\pi} \rho_s(\omega) \rho_\theta(\omega) \sin^2\left(\frac{\varphi_s(\omega) - \varphi_\theta(\omega)}{2}\right) h(\omega) d\omega$

Notes : X_g critères provenant de GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) et X_w critères provenant de WILDI et MCELROY (2019).

ρ_s et φ_s représentent le gain et la fonction de déphasage du filtre symétrique d’Henderson.

h est la densité spectrale de la série en entrée, fixée celle d’un bruit blanc, $h_{WN}(x) = 1$, ou d’une marche aléatoire,

$$h_{RW}(x) = \frac{1}{2(1-\cos(x))}.$$

Comme montré dans cette étude, la majorité des filtres linéaires peut s’obtenir par une minimisation d’une somme pondérée de ces critères, sous contrainte linéaire sur les coefficients :

$$\begin{cases} \min_{\theta} & \sum \alpha_i I(\theta, q_i, y_t, u_t^{(i)}) + \beta_i J(\theta, f_i, \omega_{1,i}, \omega_{2,i}) \\ \text{s.t.} & C\theta = a \end{cases}$$

C’est en particulier le cas de l’approche *Fidelity-Smoothness-Timeliness* (FST) développée par les mêmes auteurs.

3.2 Approche *Fidelity-Smoothness-Timeliness* (FST)

Pour construire les moyennes mobiles symétriques, GRUN-REHOMME et LADIRAY (1994) et GRAY et THOMSON (1996) proposent un programme de minimisation sous contrainte qui fait un compromis entre réduction de la variance et « lissage » de la tendance. GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) étendent ces approches en les appliquant à la construction des filtres asymétriques et en ajoutant un critère permettant de contrôler le déphasage. Il s’agit de l’approche *Fidelity-Smoothness-Timeliness* — Fidélité-Lissage-Temporalité — (FST). Pour la construction des filtres asymétriques, un quatrième critère pourrait également être rajouté qui prendrait en compte les révisions par rapport à l’utilisation d’un filtre symétrique de référence (cette méthode pourrait alors être appelée la méthode FRST — *Fidelity-Revisions-Smoothness-Timeliness*). Cependant, le *package* `rjdfilters` n’ayant implémenté que l’approche FST, nous nous restreignons dans cette étude qu’à la description de l’approche sans critère de révision.

Les trois critères utilisés sont les suivants :

— *Fidelity* (fidélité), F_g : c’est le rapport de réduction de la variance. Plus il est petit et plus

le signal de sortie (tendance-cycle estimée) est un bon estimateur du signal à estimer (tendance-cycle).

$$F_g(\theta) = \sum_{k=-p}^{+f} \theta_k^2.$$

F_g peut également être écrite comme une forme quadratique positive : $F_g(\theta) = \theta' F \theta$ avec F la matrice identité d'ordre $p + f + 1$.

— *Smoothness* (lissage), S_g :

$$S_g(\theta) = \sum_j (\nabla^d \theta_j)^2.$$

Ce critère mesure la proximité du signal de sortie à une tendance polynomiale de degré $d - 1$. Avec $d = 3$, Henderson utilise ce critère pour construire des moyennes mobiles conservant des polynômes de degré 2. S_g peut également s'écrire sous une forme quadratique positive $S_g(\theta) = \theta' S \theta$ avec S une matrice symétrique d'ordre $p + f + 1$ (voir section 5.5).

— *Timeliness* (temporalité), T_g : il mesure le déphasage entre le signal d'entrée et le signal de sortie à des fréquences spécifiques. Lorsqu'un filtre linéaire est appliqué, le niveau du signal d'entrée est également modifié par la fonction de gain : il est donc intuitif de considérer que plus le gain est élevé, plus l'impact du déphasage le sera.

C'est pourquoi le critère de déphasage dépend des fonctions de gain et de déphasage (ρ_θ et φ_θ), le lien entre les deux fonctions étant fait à partir d'une fonction de pénalité f ¹¹ :

$$\int_{\omega_1}^{\omega_2} f(\rho_\theta(\omega), \varphi_\theta(\omega)) d\omega.$$

Comme fonction de pénalité, les auteurs suggèrent de prendre $f : (\rho, \varphi) \mapsto \rho^2 \sin(\varphi)^2$. Cela permet notamment d'avoir une *timeliness* qui peut s'écrire comme une forme quadratique positive ($T_g(\theta) = \theta' T \theta$ avec T une matrice carré symétrique d'ordre $p + f + 1$, voir GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) pour la démonstration). Dans cet article nous utilisons $\omega_1 = 0$ et $\omega_2 = 2\pi/12$: on ne s'intéresse qu'à des séries mensuelles et au déphasage qui impactent les cycles d'au minimum 12 mois.

En somme, l'approche FST consiste à minimiser une somme pondérée de ces trois critères sous certaines contraintes (généralement préservation polynomiale).

$$\begin{cases} \min_{\theta} & \alpha F_g(\theta) + \beta S_g(\theta) + \gamma T_g(\theta) = \theta' (\alpha F + \beta S + \gamma T) \theta \\ \text{s.t.} & C \theta = a \end{cases} . \quad (3)$$

Les conditions $0 \leq \alpha, \beta, \gamma \leq 1$ et $\alpha + \beta \neq 0$ garantissent que $\alpha F_g(\theta) + \beta S_g(\theta) + \gamma T_g(\theta)$ soit strictement convexe et donc l'unicité de la solution. Dans ce cas, la solution s'écrit $\hat{\theta} = [\alpha F + \beta S + \gamma T]^{-1} C' (C[\alpha F + \beta S + \gamma T]^{-1} C')^{-1} a$.

Un autre avantage de cette approche est que les filtres asymétriques construits sont totalement indépendants des données, de la date d'estimation et du filtre symétrique choisis.

11. GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) suggèrent 6 conditions de régularité à la fonction de pénalité afin qu'elle soit adaptée au problème de déphasage. Dans leur article, la fonction f ne dépend que du gain et du déphasage de θ et les 6 conditions sont : $f \geq 0$, $f(\rho, 0) = 0$, $f(0, \varphi) = 0$, $f(\rho, \varphi) = f(\rho, -\varphi)$, $\frac{\partial f}{\partial \rho} \geq 0$ et $\varphi \cdot \frac{\partial f}{\partial \varphi} \geq 0$.

On obtient par exemple le filtre d'Henderson avec les paramètres suivants :

$$C = \begin{pmatrix} 1 & \cdots & 1 \\ -h & \cdots & h \\ (-h)^2 & \cdots & h^2 \end{pmatrix}, \quad a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \alpha = \gamma = 0, \quad \beta = 1, \quad d = 3.$$

Un des inconvénients de cette approche est que les différents critères ne sont pas normalisés : leurs valeurs ne peuvent pas être comparées et n'ont donc pas de sens. Il n'y a, par exemple, pas d'interprétation à donner un poids deux fois plus important à la *timeliness* qu'à la *fidelity*.

Les trois critères utilisés dans le programme de minimisation (3) sont des cas particuliers de ceux définis dans la section 3.1. En effet, en notant $y_t = TC_t + \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ avec TC_t une tendance déterministe, on a :

$$\begin{aligned} F_g(\theta) &= I(\theta, 0, y_t, \mathbb{E}[M_\theta y_t]) \\ S_g(\theta) &= I(\theta, q, y_t, \mathbb{E}[M_\theta y_t]) \\ T_g(\theta) &= J(f: (\rho, \varphi) \mapsto \rho^2 \sin(\varphi)^2, \omega_1, \omega_2). \end{aligned}$$

Encadré 1 — Filtres FST — Grun-Rehomme et alii (2018)

Avantages :

- Le filtre asymétrique est indépendant du filtre symétrique, des données et de la date d'estimation.
- Le problème d'optimisation admet une solution unique.

Inconvénients :

- Les différents critères ne sont pas normalisés : les poids accordés aux différents critères ne peuvent être comparés.

Fonction  : `rjdfilters::fst_filter()`.

4 Filtres dépendant des données : trilemme ATS

WILDI et MCELROY (2019) proposent une approche dépendante des données (*data-dependent*) pour calculer des filtres linéaires. Ils décomposent l'erreur quadratique moyenne de révision en un trilemme de trois quantités appelées *accuracy* (précision), *timeliness* (temporalité) et *smoothness* (lissage), d'où son nom *ATS-trilemma*.

Soient :

- $\{y_t\}$ notre série temporelle en entrée¹².
- $\{TC_t\}$ le signal cible, i.e. la série lissée avec un filtre symétrique fini ou non, et soient respectivement Γ_s , ρ_s et φ_s les fonctions de transfert, de gain et de déphasage associées à ce filtre symétrique.

12. Par rapport à l'article originel, les notations ont été modifiées afin de garder une cohérence entre les différentes sections.

- $\{\widehat{TC}_t\}$ une estimation de $\{TC_t\}$, i.e. le résultat d'un filtre asymétrique (lorsque toutes les observations ne sont pas disponibles), et soient respectivement Γ_θ , ρ_θ et φ_θ les fonctions de transfert, de gain et de déphasage associées à ce filtre asymétrique.

Une approche directe¹³, *Direct Filter Approach* (DFA), consiste à approcher directement le signal cible par minimisation de l'erreur quadratique moyenne :

$$\min_{\widehat{TC}_t} \mathbb{E} \left[(TC_t - \widehat{TC}_t)^2 \right].$$

Cette approche peut être approfondie en décomposant l'erreur de quadratique moyenne en plusieurs éléments d'intérêt.

Si l'on suppose que la série $\{y_t\}$ est faiblement stationnaire avec une densité spectrale continue h , l'erreur quadratique moyenne de révision, $\mathbb{E} \left[(TC_t - \widehat{TC}_t)^2 \right]$, peut s'écrire dans le domaine spectral comme:

$$\mathbb{E} \left[(TC_t - \widehat{TC}_t)^2 \right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 h(\omega) d\omega = \frac{1}{2\pi} \times 2 \times \int_0^{\pi} |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 h(\omega) d\omega \quad (4)$$

Cette égalité peut également se généraliser aux processus intégrés non-stationnaires (par exemple en imposant une cointégration entre les deux signaux et en utilisant le pseudo-spectre, voir WILDI et MCELROY (2013)).

On a :

$$\begin{aligned} |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 &= \rho_s(\omega)^2 + \rho_\theta(\omega)^2 + 2\rho_s(\omega)\rho_\theta(\omega) (1 - \cos(\varphi_s(\omega) - \varphi_\theta(\omega))) \\ &= (\rho_s(\omega) - \rho_\theta(\omega))^2 + 4\rho_s(\omega)\rho_\theta(\omega) \sin^2 \left(\frac{\varphi_s(\omega) - \varphi_\theta(\omega)}{2} \right) \end{aligned} \quad (5)$$

L'intervalle $[0, \pi]$ peut être coupé en deux : une partie dite *pass-band* $[0, \omega_1]$ (l'intervalle de fréquences qui contient le signal cible) et une partie dite *stop-band* $[\omega_1, \pi]$ (l'intervalle de fréquences associé aux résidus).

L'erreur de l'équation (4) peut être décomposée en 4 quantités :

$$\begin{aligned} Accuracy &= A_w(\theta) = 2 \int_0^{\omega_1} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) d\omega \\ Timeliness &= T_w(\theta) = 8 \int_0^{\omega_1} \rho_s(\omega)\rho_\theta(\omega) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) h(\omega) d\omega \\ Smoothness &= S_w(\theta) = 2 \int_{\omega_1}^{\pi} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) d\omega \\ Residual &= R_w(\theta) = 8 \int_{\omega_1}^{\pi} \rho_s(\omega)\rho_\theta(\omega) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) h(\omega) d\omega \end{aligned}$$

Remarque. Afin d'avoir des définitions cohérentes entre les différentes sections, les formules des quatre critères ont été légèrement modifiées par rapport à ceux définis par WILDI et MCELROY (2019) :

13. Par opposition aux approches indirectes par exemple utilisées dans X-13-ARIMA où le signal cible est approché en faisant une prévision sur la série initiale.

- dans cette étude, l'intervalle d'intégration est $[0, \pi]$ plutôt que $[-\pi, \pi]$ (toutes les fonctions étant paires cela revient à multiplier par 2 toutes les intégrales) ;
- dans l'article originel, l'intervalle pass-band dépend de la fonction de gain du filtre symétrique (pass-band = $\{\omega | \rho_s(\omega) \geq 0,5\}$) : cela correspond donc à l'intervalle contenant les fréquences conservées sans trop de distorsion par le filtre symétrique. Dans le cas du filtre symétrique d'Henderson de 13 termes, cela correspond à l'intervalle $[0, 2\pi/8]$, c'est-à-dire aux cycles de plus de 8 mois.

En général, le résidu R_w est petit puisque $\rho_s(\omega)\rho_\theta(\omega)$ est proche de 0 dans l'intervalle *stop-band*¹⁴. Il est de plus rare que les utilisateurs s'intéressent aux propriétés de déphasage dans les fréquences *stop-band*. C'est pourquoi, pour la construction de filtres linéaires les auteurs suggèrent de faire une minimisation d'une somme pondérée des trois premiers critères :

$$\mathcal{M}(\vartheta_1, \vartheta_2) = \vartheta_1 T_w(\theta) + \vartheta_2 S_w(\theta) + (1 - \vartheta_1 - \vartheta_2) A_w(\theta).$$

Comme le montrent MCELROY et WILDI (2020), cette méthode peut également être étendue au cas multivarié, ce qui permet de prendre en compte les corrélations entre les composantes de différentes séries.

Un des inconvénients de cette méthode est qu'il n'y a pas de garantie d'unicité de la solution. En revanche, son avantage par rapport à la méthode FST (section 3.2) est que la décomposition de l'erreur quadratique moyenne permet de normaliser les différents indicateurs, et les coefficients ϑ_1 , ϑ_2 et $1 - \vartheta_1 - \vartheta_2$ peuvent être comparés entre eux.

Les quatre critères A_w , T_w , S_w et R_w étant des cas particuliers du critère J défini dans l'équation (1), cette méthode s'inscrit dans le cadre de la théorie générale définie dans la section 3.1. En effet, en notant :

$$\begin{cases} f_1: & (\rho, \varphi, \omega) \mapsto 2(\rho_s(\omega) - \rho)^2 h(\omega) \\ f_2: & (\rho, \varphi, \omega) \mapsto 8\rho_s(\omega)\rho \sin^2\left(\frac{\varphi}{2}\right) h(\omega) \end{cases}$$

on a :

$$\begin{aligned} A_w(\theta) &= J(\theta, f_1, 0, \omega_1) \\ T_w(\theta) &= J(\theta, f_2, 0, \omega_1) \\ S_w(\theta) &= J(\theta, f_1, \omega_1, \pi) \\ R_w(\theta) &= J(\theta, f_2, \omega_1, \pi). \end{aligned}$$

Cette méthode étant totalement dépendante des données, son intégration dans des algorithmes non-paramétriques tels que X-11 serait compliquée. C'est pourquoi elle n'est pour l'instant pas comparée aux autres. Pour avoir des critères qui ne dépendent pas des données, une idée serait de fixer la densité spectrale, par exemple à celle d'un bruit blanc ($h_{WN}(x) = 1$) ou d'une marche aléatoire ($h_{RW}(x) = \frac{1}{2(1-\cos(x))}$). C'est ce qui est implémenté dans la fonction `rjdfilters::dfa_filter()`.

14. En pratique ce n'est pas toujours le cas comme montré dans la section 5.1.4.

Encadré 2 — ATS-trilemma – Wildi et McElroy (2019)

Avantages :

- Les valeurs des différents critères peuvent être comparées et les poids associés peuvent être interprétés.
- Méthode généralisable aux cas multivariés.

Inconvénients :

- Les filtres dépendent des données, de la date d'estimation et du filtre symétrique utilisé.
- Il peut y avoir des problèmes d'optimisation (plusieurs minimums, etc.).

📦 : *package* MDFA <https://github.com/wiaidp/MDFA> ou `rjdfilters::dfa_filter()` (version simplifiée).

5 Régression non paramétrique et régression polynomiale locale

Comme notamment montré par LOADER (1999), la régression locale est un cas particulier de la régression non paramétrique. Supposons que l'on ait un ensemble de points $(x_i, y_i)_{1 \leq i \leq n}$. La régression non paramétrique consiste à supposer qu'il existe une fonction μ , à estimer, telle que $y_i = \mu(x_i) + \varepsilon_i$ avec ε_i un terme d'erreur. D'après le théorème de Taylor, pour tout point x_0 , si μ est différentiable d fois, alors :

$$\forall x : \mu(x) = \mu(x_0) + \mu'(x_0)(x - x_0) + \dots + \frac{\mu^{(d)}(x_0)}{d!}(x - x_0)^d + R_d(x),$$

où R_d est un terme résiduel négligeable au voisinage de x_0 . Dans un voisinage $h(x_0)$ autour de x_0 , μ peut être approchée par un polynôme de degré d . La quantité $h(x_0)$ est appelée *fenêtre* (*bandwidth*). Si ε_i est un bruit blanc, on peut donc estimer par les moindres carrés $\mu(x_0)$ en utilisant les observations qui sont dans $[x_0 - h(x_0), x_0 + h(x_0)]$.

5.1 Régression polynomiale : approche de Proietti et Luati

5.1.1 Filtres symétriques

Reprenons maintenant les notations de PROIETTI et LUATI (2008) : supposons que notre série temporelle y_t peut être décomposée en

$$y_t = \mu_t + \varepsilon_t,$$

où μ_t est la tendance et $\varepsilon_t \stackrel{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$ est le bruit¹⁵. La tendance μ_t est localement approchée par un polynôme de degré d , de sorte que dans un voisinage h de t $\mu_t \simeq m_t$ avec :

$$\forall j \in \llbracket -h, h \rrbracket : y_{t+j} = m_{t+j} + \varepsilon_{t+j}, \quad m_{t+j} = \sum_{i=0}^d \beta_i j^i.$$

15. La série est donc désaisonnalisée.

Le problème d'extraction de la tendance est équivalent à l'estimation de $m_t = \beta_0$. En notation matricielle :

$$\underbrace{\begin{pmatrix} y_{t-h} \\ y_{t-(h-1)} \\ \vdots \\ y_t \\ \vdots \\ y_{t+(h-1)} \\ y_{t+h} \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & -h & h^2 & \cdots & (-h)^d \\ 1 & -(h-1) & (h-1)^2 & \cdots & -(h-1)^d \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & h-1 & (h-1)^2 & \cdots & (h-1)^d \\ 1 & h & h^2 & \cdots & h^d \end{pmatrix}}_X \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}}_\beta + \underbrace{\begin{pmatrix} \varepsilon_{t-h} \\ \varepsilon_{t-(h-1)} \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_{t+(h-1)} \\ \varepsilon_{t+h} \end{pmatrix}}_\varepsilon$$

Pour estimer β il faut $H \geq d+1$ et l'estimation est faite par moindres carrés pondérés — *weighted least squares* (WLS) —, ce qui revient à minimiser la fonction objectif suivante :

$$S(\hat{\beta}_0, \dots, \hat{\beta}_d) = \sum_{j=-h}^h \kappa_j (y_{t+j} - \hat{\beta}_0 - \hat{\beta}_1 j - \dots - \hat{\beta}_d j^d)^2$$

où κ_j est un ensemble de poids appelés *noyaux* (*kernel*). On a $\kappa_j \geq 0$: $\kappa_{-j} = \kappa_j$, et en notant $K = \text{diag}(\kappa_{-h}, \dots, \kappa_h)$, l'estimateur β peut s'écrire $\hat{\beta} = (X'KX)^{-1}X'Ky$. Avec $e_1 = (1 \ 0 \ \dots \ 0)'$, l'estimateur de la tendance peut donc s'écrire :

$$\hat{m}_t = e_1 \hat{\beta} = \theta' y = \sum_{j=-h}^h \theta_j y_{t-j} \text{ avec } \theta = KX(X'KX)^{-1}e_1$$

En somme, l'estimation de la tendance \hat{m}_t est obtenue en appliquant une moyenne mobile symétrique θ à y_t ¹⁶. De plus, $X'\theta = e_1$ donc :

$$\sum_{j=-h}^h \theta_j = 1, \quad \forall r \in \llbracket 1, d \rrbracket : \sum_{j=-h}^h j^r \theta_j = 0.$$

Ainsi, la moyenne mobile θ préserve les polynômes de degré d .

Concernant le choix des paramètres, l'idée générale qui prévaut est que le choix entre ces différents noyaux est secondaire¹⁷ et qu'il vaut mieux se concentrer sur deux autres paramètres :

- le degré du polynôme d : s'il est trop petit on risque d'avoir des estimations biaisées de la tendance-cycle et s'il est trop grand on risque d'avoir une trop grande variance dans les estimations (du fait d'un sur-ajustement) ;
- le nombre de voisins $H = 2h + 1$ (ou la fenêtre h) : s'il est trop petit alors trop peu de données seront utilisées pour les estimations (ce qui conduira à une grande variance dans les estimations) et s'il est trop grand alors l'approximation polynomiale sera vraisemblablement fautive ce qui conduira à avoir des estimations biaisées.

16. θ est symétrique du fait de la symétrie des noyaux κ_j .

17. Voir par exemple CLEVELAND et LOADER (1996) ou LOADER (1999). Les seules contraintes souhaitées sur le noyau est qu'il accorde un poids plus important à l'estimation centrale (κ_0) et qu'il décroît vers 0 lorsque l'on s'éloigne de l'estimation centrale. Le noyau uniforme est donc à éviter.

5.1.2 Les différents noyaux

Dans les problèmes d'extraction du signal, les observations sont généralement pondérées par rapport à leur distance à la date t : pour estimer la tendance-cycle à la date t , on accorde généralement plus d'importance aux observations qui sont proches de t .

Dans le cas continu, un noyau K est une fonction positive, paire et intégrable telle que $\int_{-\infty}^{+\infty} \kappa(u) du = 1$ et $\kappa(u) = \kappa(-u)$. Dans le cas discret, un noyau est un ensemble de poids κ_j , $j = 0, \pm 1, \dots, \pm h$ avec $\kappa_j \geq 0$ et $\kappa_j = \kappa_{-j}$.

Une classe importante de noyaux est celle des noyaux Beta. Dans le cas discret, à un facteur multiplicatif près (de sorte que $\sum_{j=-h}^h \kappa_j = 1$) :

$$\kappa_j = \left(1 - \left|\frac{j}{h+1}\right|^r\right)^s, \quad \text{avec } r > 0, s \geq 0$$

Cette classe englobe la majorité des noyaux présentés dans cette étude, à l'exception des noyaux d'Henderson, trapézoïdal et gaussien. Les principaux noyaux (qui sont également implémentés dans `rjdfilters`) sont :

— $r = 1, s = 0$ noyau uniforme :

$$\kappa_j^U = 1$$

— $r = s = 1$ noyau triangulaire :

$$\kappa_j^T = \left(1 - \left|\frac{j}{h+1}\right|\right)$$

— $r = 2, s = 1$ noyau d'Epanechnikov (ou parabolique) :

$$\kappa_j^E = \left(1 - \left|\frac{j}{h+1}\right|^2\right)$$

— $r = s = 2$ noyau quadratique (*bi-weight*) :

$$\kappa_j^{BW} = \left(1 - \left|\frac{j}{h+1}\right|^2\right)^2$$

— $r = 2, s = 3$ noyau cubique (*triweight*) :

$$\kappa_j^{TW} = \left(1 - \left|\frac{j}{h+1}\right|^2\right)^3$$

— $r = s = 3$ noyau tricube :

$$\kappa_j^{TC} = \left(1 - \left|\frac{j}{h+1}\right|^3\right)^3$$

— noyau d'Henderson (voir partie 5.1.3 pour plus de détails) :

$$\kappa_j = \left[1 - \frac{j^2}{(h+1)^2}\right] \left[1 - \frac{j^2}{(h+2)^2}\right] \left[1 - \frac{j^2}{(h+3)^2}\right]$$

— noyau trapézoïdal :

$$\kappa_j^{TP} = \begin{cases} \frac{1}{3(2h-1)} & \text{if } j = \pm h \\ \frac{2}{3(2h-1)} & \text{if } j = \pm(h-1) \\ \frac{1}{2h-1} & \text{otherwise} \end{cases}$$

— noyau gaussien¹⁸ :

$$\kappa_j^G = \exp\left(-\frac{j^2}{2\sigma^2 h^2}\right)$$

Les noyaux d'Henderson, trapézoïdal et gaussien sont particuliers :

— Les fonctions noyau d'Henderson et trapézoïdal changent avec la fenêtre (les autres dépendent uniquement du rapport $j/h + 1$).

18. Dans `rjdfilters` σ^2 est fixé arbitrairement à $\sigma^2 = 0,25$.

- Pour les noyaux trapézoïdal et gaussien, d'autres définitions pourraient être utilisées et sont donc définis arbitrairement.

Le noyau trapézoïdal est implémenté dans `rjdfilters` car il permet de calculer les moyennes mobiles utilisées dans l'algorithme X-13ARIMA pour l'extraction des composantes saisonnières. Il n'est pas adapté dans le cas de l'extraction de la tendance-cycle.

5.1.3 Quelques filtres symétriques particuliers

Lorsque $p = 0$ (ajustement local par une constante) on obtient l'estimateur de **Nadaraya-Watson** (ou l'estimateur par noyaux).

Avec le noyau uniforme on obtient le filtre de MACAULAY et alii (1931). Lorsque $p = 0$ ou $p = 1$, on retrouve la moyenne arithmétique : $w_j = w = \frac{1}{2h+1}$.

Le noyau d'**Epanechnikov** est souvent recommandé comme le noyau optimal car il minimise l'erreur quadratique moyenne de l'estimation par polynômes locaux.

Le **Loess**, *locally estimated scatterplot smoothing* (utilisé dans la méthode STL), est une régression locale pondérée qui utilise le noyau tricube.

Le **filtre d'Henderson** est un cas particulier de l'approximation locale cubique ($p = 3$), couramment utilisée pour l'extraction de la tendance-cycle (c'est par exemple le filtre utilisé dans le logiciel de désaisonnalisation X-13ARIMA). Pour une fenêtre fixée, Henderson a trouvé le noyau qui donnait l'estimation la plus lisse de la tendance. Il montre l'équivalence entre les trois problèmes suivants :

1. minimiser la variance de la différence d'ordre trois de la série lissée par l'application d'une moyenne mobile ;
2. minimiser la somme du carré de la différence d'ordre trois des coefficients du filtre, c'est le critère de lissage (*smoothness*) : $S = \sum_j (\nabla^3 \theta_j)^2$;
3. estimer une tendance localement cubique par les moindres carrés pondérés, où les poids sont choisis de sorte à minimiser la *smoothness* (cela conduit au noyau présenté dans la section 5.1.2).

Le filtre d'Henderson étant couramment utilisé pour l'extraction de la tendance-cycle, nous nous intéresserons uniquement aux filtres issus du noyau d'Henderson.

5.1.4 Filtres asymétriques

Comme mentionné dans la partie 2.3.1, pour l'estimation en temps réel, plusieurs approches peuvent être utilisées :

1. Construire un filtre asymétrique par approximation polynomiale locale sur les observations disponibles (y_t pour $t \in \llbracket n-h, n \rrbracket$).
2. Appliquer les filtres symétriques sur les séries prolongées par prévision $\hat{y}_{n+l|n}, l \in \llbracket 1, h \rrbracket$.
3. Construire des filtres asymétriques qui minimisent l'erreur quadratique moyenne de révision sous des contraintes de reproduction de tendances polynomiales.

PROIETTI et LUATI (2008) montrent que les deux premières approches sont équivalentes lorsque les prévisions sont faites par extrapolation polynomiale de degré d . Elles sont également équivalentes à la troisième approche sous les mêmes contraintes que celles du filtre symétrique. Cette méthode est appelée *direct asymmetric filter* (DAF). C'est cette méthode qui est utilisée pour l'estimation en temps réel dans la méthode de désaisonnalisation STL (*Seasonal-Trend decomposition based on Loess*, voir CLEVELAND et alii (1990)). Même si les estimations des filtres DAF sont sans biais, c'est au coût d'une plus grande variance dans les estimations.

Pour résoudre le problème de la variance des estimations des filtres temps réel, PROIETTI et LUATI (2008) proposent une méthode générale pour construire les filtres asymétriques qui permet de faire un compromis biais-variance. Il s'agit d'une généralisation des filtres asymétriques de MUSGRAVE (1964) (utilisés dans l'algorithme de désaisonnalisation X-13ARIMA).

On modélise ici la série en entrée par :

$$y = U\gamma + Z\delta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, D) \quad (6)$$

où $[U, Z]$ est de rang plein et forme un sous-ensemble des colonnes de X . L'objectif est de trouver un filtre v qui minimise l'erreur quadratique moyenne de révision (au filtre symétrique θ) sous certaines contraintes. Ces contraintes sont représentées par la matrice $U = (U_p' \ U_f')'$: $U_p'v = U'\theta$ (avec U_p la matrice $(h+q+1) \times (d+1)$ qui contient les observations de la matrice U connues lors de l'estimation par le filtre asymétrique). Le problème est équivalent à trouver v qui minimise :

$$\varphi(v) = \underbrace{(v - \theta_p)' D_p (v - \theta_p)}_{\text{variance de l'erreur de révision}} + \underbrace{\theta_f' D_f \theta_f + [\delta'(Z_p'v - Z'\theta)]^2}_{\text{biais}^2} + \underbrace{2l'(U_p'v - U'\theta)}_{\text{contraintes}} \quad (7)$$

Erreur quadratique moyenne de révision

où l est le vecteur des multiplicateurs de Lagrange.

Lorsque $U = X$, la contrainte équivaut à préserver les polynômes de degré d : on retrouve les filtres directs asymétriques (DAF) lorsque $D = K^{-1}$.

Lorsque $U = (1 \ \dots \ 1)'$, $Z = (-h \ \dots \ +h)'$, $\delta = \delta_1$, $D = \sigma^2 I$ et lorsque le filtre symétrique est le filtre d'Henderson, on retrouve les filtres asymétriques de Musgrave. Ce filtre suppose, que pour l'estimation en temps réel, les données sont générées par un processus linéaire et que les filtres asymétriques préservent les constantes ($\sum v_i = \sum \theta_i = 1$). Ces filtres asymétriques dépendent du rapport δ_1/σ , qui est lié à l'I-C ratio $R = \frac{\bar{I}}{\bar{C}} = \frac{\sum |I_t - I_{t-1}|}{\sum |C_t - C_{t-1}|}$ ($\delta_1/\sigma = 2/(R\sqrt{\pi})$), qui est notamment utilisé dans X-13ARIMA pour déterminer la longueur du filtre d'Henderson¹⁹.

Lorsque U correspond aux $d^* + 1$ premières colonnes de X , $d^* < d$, la contrainte consiste à reproduire des tendances polynomiales de degré d^* . Cela introduit du biais mais réduit la variance. Ainsi, PROIETTI et LUATI (2008) proposent trois classes de filtres asymétriques :

1. *Linear-Constant* (LC) : y_t linéaire ($d = 1$) et v préserve les constantes ($d^* = 0$). On obtient le filtre de Musgrave avec le filtre d'Henderson comme filtre symétrique.
2. *Quadratic-Linear* (QL) : y_t quadratique ($d = 2$) et v préserve les tendances linéaires ($d^* = 1$).

19. Dans la majorité des cas un filtre de 13 termes est utilisé. Si le ratio est grand alors un filtre de 23 termes est utilisé (pour supprimer d'avantage de bruit) et si le ratio est petit un filtre de 9 termes est utilisé.

3. *Cubic-Quadratic* (CQ) : y_t cubic ($d = 3$) et v préserve les tendances quadratiques ($d^* = 2$).

Le tableau 2 compare les critères de qualité des différentes méthodes en utilisant le filtre d'Henderson et $h = 6$ (filtre symétrique de 13 termes). Pour les filtres en temps réel ($q = 0$), plus le filtre asymétrique est complexe (en termes de préservation polynomiale), moins la *timeliness* est élevée et plus la *fidelity/smoothness* est grande : la réduction du déphasage se fait au détriment d'une augmentation de la variance. Ce résultat varie lorsque q augmente : pour $q = 2$ le filtre QL a une plus grande *timeliness* que le filtre LC. Ce résultat étonnant souligne le fait que le déphasage n'est pas contrôlé par l'approche de PROIETTI et LUATI (2008).

En termes de révision, ($A_w + S_w + T_w + R_w$), les filtres LC et QL donnent toujours de meilleurs résultats que les filtres CQ et DAF.

TABLE 2 – Critères de qualité des filtres asymétriques ($q = 0, 1, 2$) calculés par polynômes locaux en utilisant le noyau d'Henderson avec $h = 6$ et $R = 3, 5$.

Method	b_c	b_l	b_q	F_g	S_g	$T_g \times 10^{-3}$	A_w	S_w	T_w	R_w
$q = 0$										
LC	0	-0,41	-2,16	0,39	1,27	30,34	0,10	0,49	0,41	0,55
QL	0	0,00	-0,47	0,71	5,15	0,05	0,07	1,89	0,00	0,11
CQ	0	0,00	0,00	0,91	11,94	0,01	0,02	2,23	0,00	0,10
DAF	0	0,00	0,00	0,94	14,20	0,00	0,01	2,18	0,00	0,10
$q = 1$										
LC	0	-0,12	-0,52	0,27	0,43	4,80	0,01	0,12	0,06	0,11
QL	0	0,00	-0,06	0,29	0,71	0,69	0,00	0,19	0,01	0,04
CQ	0	0,00	0,00	0,37	0,57	0,16	0,02	0,58	0,00	0,06
DAF	0	0,00	0,00	0,41	0,37	0,06	0,02	0,76	0,00	0,06
$q = 2$										
LC	0	0,00	1,08	0,20	0,08	0,35	0,01	0,01	0,00	0,01
QL	0	0,00	0,03	0,22	0,05	2,08	0,00	0,01	0,02	0,07
CQ	0	0,00	0,00	0,37	0,66	0,13	0,02	0,56	0,00	0,06
DAF	0	0,00	0,00	0,40	0,77	0,02	0,02	0,68	0,00	0,05

Une application en ligne, disponible à l'adresse <https://aqlt.shinyapps.io/FilterProperties/>, permet de comparer les coefficients, les fonctions de gain et de déphasage entre les différentes méthodes et les différents noyaux.

Encadré 3 — Filtres locaux polynomiaux — Proietti et Luati (2008)

Avantages :

- Modèles avec une interprétation simple.
- Le filtre asymétrique est indépendant de la date d'estimation. Toutefois, il dépend indirectement des données si le filtre est calibré sur l'I-C ratio.

Inconvénients :

- La *timeliness* n'est pas contrôlée (mais peut être introduite dans le programme de minimisation).

5.2 Extension avec le critère de *timeliness*

Un inconvénient de la méthode précédente est que le déphasage n'est pas contrôlé. Il est en revanche possible de généraliser d'avantage la modélisation en ajoutant le critère de *timeliness* défini par GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) dans l'équation (7). C'est ce qui a été proposé par Jean Palate, puis codé en Java²⁰ et intégré dans `rjdfilters`.

En utilisant les mêmes notations que dans 5.1.4, θ le filtre symétrique et v le filtre asymétrique. Notons également $\theta = \begin{pmatrix} \theta_p \\ \theta_f \end{pmatrix}$ avec θ_p de même longueur que v , et $g = v - \theta_p$. Le critère de *timeliness* s'écrit :

$$T_g(v) = v'Tv = g'Tg + 2\theta_p'Tg + \theta_p'T\theta_p \quad (T \text{ étant symétrique}).$$

De plus, la fonction objectif φ de l'équation (7) peut se réécrire :

$$\begin{aligned} \varphi(v) &= (v - \theta_p)'D_p(v - \theta_p) + \theta_f'D_f\theta_f + [\delta'(Z_p v - Z'\theta)]^2 + 2l'(U_p v - U'\theta) \\ &= g'Qg - 2Pg + 2l'(U_p v - U'\theta) + c \quad \text{avec} \begin{cases} Q = D_p + Z_p\delta\delta'Z_p' \\ P = \theta_f Z_f \delta\delta' Z_p' \\ c \text{ une constante indépendante de } v \end{cases} \end{aligned}$$

En ajoutant le critère de *timeliness*, on obtient :

$$\tilde{\varphi}(v) = g'\tilde{Q}g - 2\tilde{P}g + 2l'(U_p v - U'\theta) + \tilde{c} \quad \text{avec} \begin{cases} \tilde{Q} = D_p + Z_p\delta\delta'Z_p' + \alpha_T T \\ \tilde{P} = \theta_f Z_f \delta\delta' Z_p' - \alpha_T \theta_p T \\ \tilde{c} \text{ une constante indépendante de } v \end{cases}$$

où α_T est le poids associé au critère de *timeliness*. Avec $\alpha_T = 0$ on retrouve $\varphi(v)$. Cette extension permet donc de retrouver tous les filtres symétriques et asymétriques présentés dans la section précédente mais généralise également l'approche de GRAY et THOMSON (1996) présentée dans la section 5.3. Elle est proche de l'approche FRST.

Cette extension s'inscrit dans le cadre de la théorie générale définie dans la section 3.1. Cela revient en effet à minimiser une somme pondérée de l'erreur quadratique de révision :

$$\mathbb{E} \left[\left(\sum_{i=-h}^h \theta_i^s y_{t+s} - \sum_{i=-h}^q v_i y_{t+s} \right)^2 \right] = I(v, 0, y_t, M_{\theta^s} y_t)$$

20. <https://github.com/palatej/jdemetra-core>.

et du critère de *timeliness* :

$$T_g(\theta) = J(f : (\rho, \varphi) \mapsto \rho^2 \sin(\varphi)^2, \omega_1, \omega_2)$$

sous une contrainte linéaire.

5.3 Régression polynomiale : Gray et Thomson

5.3.1 Filtres symétriques

L'approche de GRAY et THOMSON (1996) est proche de celles de PROIETTI et LUATI (2008) et de GRUN-REHOMME, GUGGEMOS et LADIRAY (2018). De la même façon que pour les autres méthodes, ils considèrent que la série initiale y_t peut se décomposer entre une somme entre la tendance-cycle g_t et d'un bruit blanc ε_t de variance σ^2 :

$$y_t = g_t + \varepsilon_t.$$

Toutefois, plutôt que de directement remplacer g_t par un polynôme local de degré d , ils prennent en compte l'erreur d'approximation de la tendance :

$$g_t = \sum_{j=0}^d \beta_j t^j + \xi_t,$$

où ξ_t est un processus stochastique de moyenne nulle, autocorrélé mais non corrélé à ε_t .

La tendance g_t est estimée par une moyenne mobile :

$$\hat{g}_t = \sum_{s=-r}^r \theta_s y_{t+s}.$$

Pour le filtre central, les auteurs cherchent à avoir un estimateur \hat{g}_t qui soit sans biais (ce qui implique que θ conserve les tendances de degré d) et qui minimise une somme pondérée d'un critère de *fidelity* et d'un critère de *smoothness* :

$$Q = \alpha \underbrace{\mathbb{E} [(\hat{g}_t - g_t)^2]}_{=F_{GT}} + (1 - \alpha) \underbrace{\mathbb{E} [(\Delta^{p+1} \hat{g}_t)^2]}_{=S_{GT}} \quad (8)$$

La solution est un filtre symétrique qui peut s'écrire sous la forme

$$\theta = E_\alpha^{-1} X [X' E_\alpha^{-1} X]^{-1} e_1 \text{ avec } E_\alpha = \alpha (\sigma^2 I + \Omega) + (1 - \alpha) (\sigma^2 B_{p+1} + \Gamma)$$

où :

$$\begin{cases} \Omega_{jk} & = \text{cov}(\xi_{t+j} - \xi_t, \xi_{t+k} - \xi_t) \\ \Gamma_{jk} & = \text{cov}(\Delta^{p+1} \xi_{t+j}, \Delta^{p+1} \xi_{t+k}) \\ \sigma^2 (B_{p+1})_{jk} & = \text{cov}(\Delta^{p+1} \varepsilon_{t+j}, \Delta^{p+1} \varepsilon_{t+k}) \end{cases}$$

Les deux critères utilisés dans le programme de minimisation (8) sont des cas particuliers du critère I défini dans l'équation (1) :

$$\begin{aligned} F_{GT}(\theta) &= I(\theta, 0, y_t, M_\theta y_t) \\ S_{GT}(\theta) &= I(\theta, d+1, y_t, 0). \end{aligned}$$

La théorie générale définie dans la section 3.1 permet donc de retrouver les filtres de GRAY et THOMSON (1996).

En ne minimisant que la *smoothness* et avec $\xi_t = 0$ on retrouve le filtre d'Henderson. En ne minimisant que la *fidelity*, cette méthode est équivalente à l'estimation de polynômes locaux par moindres carrés généralisés : on retrouve donc les filtres de PROIETTI et LUATI (2008) avec $\sigma^2 = 0$ et $\Omega = K^{-1}$, ainsi que le filtre de Macaulay.

L'avantage de la modélisation de Gray et Thomson est que le paramètre ξ_t permet une spécification plus précise du modèle en prenant notamment en compte la corrélation entre les observations. Par exemple, MCLAREN et STEEL (2001) ont étudié le lien entre le plan de sondage et l'estimation de la composante tendance-cycle et de la composante saisonnière. Cette modélisation leur permet de prendre en compte, dans l'estimation de la tendance-cycle, la structure de corrélation induite par le plan de sondage de l'enquête emploi mensuelle de l'Australie (groupe de rotations avec une période de recouvrement). Cependant, les auteurs avertissent que dans leur simulations (et dans la modélisation de Gray et Thomson) la structure d'autocorrélation de la variable aléatoire ξ_t est supposée connue. Ce n'est généralement pas le cas en pratique, où cette structure doit être estimée, ce qui rajoute de l'incertitude dans les estimations.

5.3.2 Filtres asymétriques

L'approche retenue par GRAY et THOMSON (1996) est une approche de minimisation des révisions sous contraintes. Étant donné un filtre symétrique θ^s utilisé pour estimer la tendance au centre de la série, l'objectif est de chercher un filtre asymétrique $v = (v_{-h}, \dots, v_q)$ de sorte à minimiser l'erreur quadratique moyenne de révision :

$$\mathbb{E} \left[(Y - \hat{Y})^2 \right] = \mathbb{E} \left[\left(\sum_{i=-h}^h \theta_i^s y_{t+s} - \sum_{i=-h}^q v_i y_{t+s} \right)^2 \right].$$

Les auteurs étudient deux cas :

1. Dans le premier cas, ils cherchent un estimateur sans biais : cela implique que v conserve les mêmes tendances polynomiales que θ^s . \hat{Y} est alors le meilleur prédicteur linéaire sans biais — *best linear unbiased predictor* (BLUP) — de Y .
2. Dans le second cas, ils autorisent l'estimateur à être biaisé mais imposent que ce biais soit constant dans le temps : si l'on modélise localement la tendance par un polynôme de degré d , cela implique que v conserve les tendances polynomiales de degré $d - 1$. \hat{Y} est alors le meilleur prédicteur linéaire à biais constant — *best linear time invariant predictor* (BLIP) — de Y . Cela permet notamment de reproduire les filtres asymétriques de Musgrave.

La méthode utilisée est donc très proche de celle de PROIETTI et LUATI (2008) : on retrouve d'ailleurs le filtre DAF avec $\sigma^2 = 0$ et $\Omega = K^{-1}$ et en utilisant la première méthode (estimation du BLUP) et les méthodes LC (filtre de Musgrave), QL et CQ avec la seconde méthode en utilisant respectivement $d = 1$, $d = 2$ et $d = 3$.

La théorie générale définie dans la section 3.1 permet également de retrouver les filtres asymétriques puisqu'ils sont construits en minimisant l'erreur quadratique moyenne des révisions sous contraintes linéaires (préservation d'un polynôme de degré p).

Remarque. Pour la construction des filtres asymétriques, une approche alternative pourrait être d'utiliser la même méthode que celle utilisée pour construire les filtres symétriques. C'est-à-dire minimiser Q (équation (8)) sous contrainte que le filtre asymétrique fournisse un estimateur dans biais de la tendance. Comme discuté dans GRAY et THOMSON (1996), les auteurs ne retiennent pas cette méthode pour deux raisons :

- Il n'est pas évident qu'il faudrait chercher à maintenir le même équilibre entre smoothness et fidelity en fin de série et au centre de la série. Le problème rencontré en fin de série est transitoire et disparaît au fur et à mesure que l'on a de nouvelles observations. Minimiser des critères de révision serait donc préférable puisque cela reviendrait à minimiser le coût de la transition (mais dans le cas où l'on ne minimise que la fidelity les deux méthodes sont équivalentes).
- Les valeurs de la fidelity et de la smoothness ne dépendent pas du temps au centre de la série mais en dépendent en fin de série. Ainsi, même si au centre de la série le choix des poids entre les deux critères contrôle indirectement le niveau des indicateurs, ce n'est plus le cas en fin de série. De plus, en fin de série, cela pourrait introduire des déphasages plus importants car S_{GT} dépend du temps et des valeurs passées (du fait de l'utilisation de l'opérateur différence).

Inversement, GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) justifie de ne pas intégrer le critère de révision dans leur problème car ce critère est fortement corrélé à une combinaison fixée, donc non ajustable par l'utilisateur, des critères fidelity et timeliness.

Encadré 4 — Filtres locaux polynomiaux — Gray et Thomson (1996)

Avantages :

- Modèles généraux qui permettent de prendre en compte l'autocorrélation entre les observations.
- Interprétation statistique des différentes méthodes.
- Le filtre asymétrique est indépendant de la date d'estimation. Toutefois, il dépend indirectement des données si le filtre est calibré sur l'I-C ratio.

Inconvénients :

- La *timeliness* n'est pas contrôlée.
- La spécification du modèle (i.e., du paramètre ξ_t) peut être compliquée : si la structure d'autocorrélation est estimée à partir des données, cela rajoute de l'incertitude dans les estimations, ce qui peut avoir des effets indésirables.

5.4 Reproducing Kernel Hilbert Space (RKHS) : approche de Dagum et Bianconcini

La théorie des *Reproducing Kernel Hilbert Space* (RKHS) — espaces de Hilbert à noyau reproduisant — est une théorie générale dans l'apprentissage statistique non-paramétrique qui permet d'englober un grand nombre de méthodes. C'est par exemple le cas des méthodes de régression par moindres carrés pénalisés, des Support Vector Machine (SVM), du filtre d'Hodrick-Prescott (utilisé pour décomposer tendance et cycle) ou encore des moyennes mobiles telles que celle d'Henderson. Ainsi, DAGUM et BIANCONCINI (2008) utilise la théorie des RKHS pour

approcher le filtre d'Henderson et en dériver des filtres asymétriques associés.

Un RKHS $\mathbb{L}^2(f_0)$ est un espace de Hilbert caractérisé par un noyau qui permet de reproduire toutes les fonctions de cet espace. Il est caractérisé par une fonction de densité f_0 et un produit scalaire $\langle \cdot, \cdot \rangle$ défini par :

$$\langle U(t), V(t) \rangle = \mathbb{E}[U(t)V(t)] = \int_{\mathbb{R}} U(t)V(t)f_0(t) dt \quad \forall U, V \in \mathbb{L}^2(f_0).$$

La fonction f_0 pondère donc chaque valeur en fonction de sa position temporelle : il s'agit de la version continue des noyaux définis dans la partie 5.1.2.

Dans notre cas, on suppose que notre série initiale y_t est désaisonnalisée et peut s'écrire comme la somme d'une tendance-cycle, TC_t , et d'une composante irrégulière, I_t (qui peut être un bruit blanc ou suivre un modèle ARIMA) : $y_t = TC_t + I_t$. La tendance-cycle peut être déterministe ou stochastique. On suppose que c'est une fonction régulière du temps, elle peut être localement approchée par un polynôme de degré d :

$$TC_{t+j} = TC_t(j) = a_0 + a_1j + \dots + a_dj^d + \varepsilon_{t+j}, \quad j \in \llbracket -h, h \rrbracket,$$

où ε_t est un bruit blanc non corrélé à I_t .

Les coefficients a_0, \dots, a_d peuvent être estimés par projection des observations au voisinage de y_t sur le sous-espace \mathbb{P}_d des polynômes de degré d , ou, de manière équivalente, par minimisation de la distance entre y_t et $TC_t(j)$:

$$\min_{TC \in \mathbb{P}_d} \|y - TC\|^2 = \min_{TC \in \mathbb{P}_d} \int_{\mathbb{R}} (y(t+s) - TC_t(s))^2 f_0(s) ds. \quad (9)$$

L'espace \mathbb{P}_d étant un espace de Hilbert à dimension finie, il admet un noyau reproduisant (voir, par exemple, BERLINET et THOMAS-AGNAN (2004)). Il existe ainsi une fonction $R_d(\cdot, \cdot)$ telle que :

$$\forall P \in \mathbb{P}_d : \forall t : R_d(t, \cdot) \in \mathbb{P}_d \quad \text{et} \quad P(t) = \langle R_d(t, \cdot), P(\cdot) \rangle.$$

Le problème (9) admet une solution unique qui dépend d'une fonction K_{d+1} , appelée *fonction de noyau (kernel function)*. Cette fonction est dite d'ordre $d+1$ car elle conserve les polynômes de degré d ²¹. Cette solution s'écrit :

$$\widehat{TC}(t) = \int_{\mathbb{R}} y(t-s)K_{d+1}(s) ds. \quad (10)$$

Généralement $f_0(t) = 0$ pour $|t| > 1$. Cette solution s'écrit alors :

$$\widehat{TC}(t) = \int_{[-1,1]} y(t-s)K_{d+1}(s) ds. \quad (11)$$

On peut, par ailleurs, montrer que K_{d+1} s'écrit en fonction f_0 et du noyau reproduisant $K(\cdot, \cdot)$ et que ce dernier peut s'écrire en fonction de polynômes $(P_i)_{i \in \llbracket 0, d-1 \rrbracket}$ qui forme une base orthonormée de $\mathbb{L}^2(f_0)$ (voir par exemple BERLINET (1993)) :

$$K_{d+1}(t) = R_d(t, 0)f_0(t) = \sum_{i=0}^d P_i(t)P_i(0)f_0(t).$$

21. C'est-à-dire $\int_{\mathbb{R}} K_{d+1}(s) ds = 1$ et $\int_{\mathbb{R}} K_{d+1}(s)s^i ds = 1$ pour $i \in \llbracket 1, d \rrbracket$.

De plus, dans le cas discret, la solution (11) s'écrit comme une somme pondérée au voisinage de y_t :

$$\widehat{TC}_t = \sum_{j=-h}^h w_j y_{t+j} \quad \text{où} \quad w_j = \frac{K_{d+1}(j/b)}{\sum_{i=-h}^h K_{d+1}(i/b)}. \quad (12)$$

Le paramètre b est choisi de sorte que les $2h + 1$ points autour de y_t soient utilisés avec un poids non nul.

Pour les filtres asymétriques, la formule (12) est simplement adaptée au nombre d'observations connues :

$$\forall j \in \llbracket -h, q \rrbracket : w_{a,j} = \frac{K_{d+1}(j/b)}{\sum_{i=-h}^q K_{d+1}(i/b)}. \quad (13)$$

En utilisant $b = h + 1$ on retrouve les filtres symétriques obtenues par polynômes locaux.

Comme notamment montré par DAGUM et BIANCONCINI (2016), K_{d+1} peut s'exprimer simplement à partir des moments de f_0 ²². Ainsi, notons H_{d+1} la matrice de Hankel associée aux moments de f_0 :

$$\forall i, j \in \llbracket 0, d \rrbracket : (H_{d+1})_{i,j} = \langle X^i, X^j \rangle = \int s^{i+j} f_0(s) ds.$$

Notons également $H_{d+1}[1, t]$ la matrice obtenue en remplaçant la première ligne de H_{d+1} par $(1 \ t \ t^2 \ \dots \ t^d)$. On a :

$$K_{d+1}(t) = \frac{\det H_{d+1}[1, t]}{\det H_{d+1}} f_0(t). \quad (14)$$

C'est cette formule qui est utilisée dans le *package* `rjdfilters` pour calculer les différentes moyennes mobiles.

Comme discuté dans la partie 5.1, le noyau d'Henderson dépend de la fenêtre utilisée. Ainsi, tous les moments de l'équation (14) doivent être recalculés pour chaque valeur de h . Pour éviter cela, DAGUM et BIANCONCINI (2008) suggèrent d'utiliser le noyau quadratique (*biweight*) pour approcher le noyau d'Henderson lorsque h est petit ($h < 24$) et le noyau cubique (*triweight*) lorsque h est grand $h \geq 24$.

Dans DAGUM et BIANCONCINI (2015), les auteures suggèrent de faire une sélection optimale du paramètre b , par exemple en minimisant l'erreur quadratique moyenne (option "frequencyresponse" dans `rjdfilters::rkhs_filter`) :

$$b_{q,\Gamma} = \min_{b_q \in [h; 3h]} 2 \int_0^\pi |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 d\omega.$$

Cela suppose en fait que la série entrée y_t est un bruit blanc. En supposant y_t stationnaire, les critères définis dans l'article originel peuvent donc être étendus en multipliant les quantités sous les intégrales par la densité spectrale de y_t notée h :

$$b_{q,\Gamma} = \min_{b_q \in [h; 3h]} 2 \int_0^\pi |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 h(\omega) d\omega.$$

Cette erreur quadratique moyenne peut également se décomposer en plusieurs termes (voir équation (4) de la section 4) qui peuvent également être minimisés :

22. Cela vient en fait du procédé d'orthonormalisation de Gram-Schmidt.

- l'*accuracy* qui correspond à la part de la révision liée aux différences de fonction de gain dans les fréquences liées à la tendance-cycle

$$b_{q,G} = \min_{b_q \in [h; 3h]} 2 \int_0^{\omega_1} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) d\omega$$

- la *smoothness* qui correspond à la part de la révision liée aux différences de fonction de gain dans les fréquences liées aux résidus

$$b_{q,s} = \min_{b_q \in [h; 3h]} 2 \int_{\omega_1}^{\pi} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) d\omega$$

- la *timeliness* qui correspond à la part de la révision liée au déphasage

$$b_{q,\varphi} = \min_{b_q \in [h; 3h]} 8 \int_0^{\omega_1} \rho_s(\lambda) \rho_\theta(\lambda) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) h(\omega) d\omega$$

Dans `rjdfilters` h peut être fixée à la densité spectrale d'un bruit blanc ($h_{WN}(x) = 1$, comme c'est le cas dans DAGUM et BIANCONCINI (2015)) ou d'une marche aléatoire ($h_{RW}(x) = \frac{1}{2(1-\cos(x))}$).

Remarque. Pour assurer une cohérence dans les définitions entre les différentes sections, les définitions de $b_{q,G}$, $b_{q,\Gamma}$ et $b_{q,\varphi}$ ont été légèrement modifiées par rapport à celles définies dans DAGUM et BIANCONCINI (2015) où :

- dans $b_{q,G}$ le terme à minimiser est sous une racine carrée (sans impact sur le minimum) :

$$b_{q,G} = \min_{b_q} \sqrt{2 \int_0^{\pi} (\rho_s(\omega) - \rho_\theta(\omega))^2 d\omega}$$

- $b_{q,s}$ n'est pas considéré, $b_{q,\Gamma}$ et $b_{q,G}$ sont définis avec $\omega_1 = \pi$
- $b_{q,\varphi}$ est défini par :

$$b_{q,\varphi} = \min_{b_q} \sqrt{2 \int_{\Omega_S} \rho_s(\lambda) \rho_\theta(\lambda) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) d\omega}$$

où $\Omega_S = [0, 2\pi/32]$ est l'intervalle de fréquences associées aux cycles d'au moins 16 mois.

- une formule différente est utilisée pour la fonction de réponse ($\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{2\pi i \omega k}$), ce qui conduit à des bornes d'intégrales légèrement différentes, sans effet sur le résultat.

Un des inconvénients de cette méthode est qu'il n'y a pas unicité de la solution et donc qu'il y a parfois plusieurs extremum (uniquement pour le calcul de $b_{q,\varphi}$. Ainsi, la valeur optimal retenue par défaut par `rjdfilters` produit des discontinuités dans l'estimation de la tendance-cycle. Par ailleurs, les valeurs de $b_{q,G}$ varient fortement en fonction de si l'on retient $\omega_1 = 2\pi/12$ ou $\omega_1 = \pi$ (tableau 3). Par cohérence, simplicité, nous utiliserons dans cet article les valeurs optimales présentées dans DAGUM et BIANCONCINI (2015).

TABLE 3 – Fenêtres optimales pour les filtres asymétriques associés à un filtre symétrique de 13 termes ($h = 6$) avec le noyau biweight.

	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
$b_{q,\Gamma}$						
$\omega_1 = \pi$	9,54	7,88	7,07	6,88	6,87	6,94
$\omega_1 = 2\pi/12$	9,54	7,88	7,07	6,88	6,87	6,94
$b_{q,G}$						
$\omega_1 = \pi$	11,78	9,24	7,34	6,85	6,84	6,95
$\omega_1 = 2\pi/12$	8,61	7,64	6,01	6,01	6,01	6,59
$b_{q,\varphi}$						
Valeurs de l'article (avec $\omega_1 = 2\pi/36$)	6,01	6,01	7,12	8,44	9,46	10,39
$\omega_1 = 2\pi/36$	6,01	6,01	7,21	8,47	9,46	6,01
$\omega_1 = 2\pi/12$	6,01	6,01	6,38	8,15	9,35	6,01

Encadré 5 — RKHS filters — Dagum et Bianconcini (2008)

Avantages :

- Le filtre asymétrique est indépendant des données et de la date d'estimation.
- La méthode est généralisable à des séries avec des fréquences irrégulières (par exemple avec beaucoup de valeurs manquantes).

Inconvénient :

- Il peut y avoir des problèmes de minimisation (notamment en minimisant la *timeliness*).

Tous les critères utilisés de la sélection optimale du paramètre b pouvant s'écrire comme cas particuliers du critère J défini dans l'équation (1) (voir section 4), cette méthode s'inscrit dans le cadre de la théorie générale définie dans la section 3.1 en imposant comme contrainte linéaire que les coefficients soient sous la forme $w_j = \frac{K_{d+1}(j/b)}{\sum_{i=-h}^p K_{d+1}(i/b)}$.

5.5 Liens entre les différentes méthodes

5.5.1 Critères de Gray et Thomson et ceux de Grun-Rehomme *et alii*

Les critères F_g et S_g peuvent se déduire de F_{GT} et S_{GT} . Les approches de GRAY et THOMSON (1996) et GRUN-REHOMME, GUGGEMOS et LADIRAY (2018) sont donc équivalentes pour la construction de filtres symétriques.

Notons $x_t = (1 \ t \ t^2 \ \dots \ t^d)$, $\beta_t = (\beta_0 \ \dots \ \beta^d)'$.

Pour le critère de *fidelity* :

$$\hat{g}_t - g_t = \left(\sum_{j=-h}^{+h} \theta_j x_{t+j} - x_t \right) \beta + \sum_{j=-h}^{+h} \theta_j \varepsilon_{t+j} + \sum_{j=-h}^{+h} \theta_j (\xi_{t+j} - \xi_t),$$

Si θ préserve les polynômes de degré d alors $\sum_{j=-h}^{+h} \theta_j x_{t+j} = x_t$. où $x_t = (1 \ t \ t^2 \ \dots \ t^d)$. Puis, comme ξ_t et ε_t sont de moyenne nulle et sont non corrélés :

$$F_{GT}(\theta) = \mathbb{E} [(\hat{g}_t - g_t)^2] = \theta' (\sigma^2 I + \Omega) \theta.$$

Si $\xi_t = 0$ alors $\Omega = 0$ et $F_{GT}(\theta) = F_g(\theta)$.

Pour la *smoothness* on a :

$$\nabla^q \hat{g}_t = \sum_{j=-h}^h \theta_j \underbrace{\nabla^q ((x_j - x_0) \beta)}_{=0 \text{ si } q \geq d+1} + \sum_{j=-h}^h \theta_j \nabla^q \varepsilon_{t+j} + \sum_{j=-h}^h \theta_j \nabla^q \xi_{t+j}.$$

D'où pour $q = d + 1$:

$$S_{GT}(\theta) = \mathbb{E} [(\nabla^q \hat{g}_t)^2] = \theta' (\sigma^2 B_q + \Gamma_q) \theta.$$

On peut par ailleurs montrer que pour toute série temporelle X_t ,

$$\nabla^q (M_\theta X_t) = (-1)^q \sum_{k \in \mathbb{Z}} (\nabla^q \theta_k) X_{t+k-q}$$

avec $\theta_k = 0$ pour $|k| \geq h + 1$. Avec $\xi_t = 0$ on trouve donc que $S_{GT}(\theta) = \sigma^2 S_g(\theta)$.

5.5.2 Équivalence avec les moindres carrés pondérés

Du fait de la forme des filtres obtenus par la méthode de GRUN-REHOMME, GUGGEMOS et LADIRAY (2018), lorsque les contraintes imposées sont la préservation des tendances de degré d , celle-ci est équivalente à une estimation locale d'une tendance polynomiale de degré d par moindres carrés généralisés. En effet, dans ce cas, la solution est $\hat{\theta} = \Sigma^{-1} X_p' (X_p \Sigma^{-1} X_p')^{-1} e_1$ avec $\Sigma = \alpha F + \beta S + \gamma T$, et c 'est l'estimation de la constante obtenue par moindres carrés généralisés lorsque la variance des résidus est Σ . L'équivalence entre les deux méthodes peut donc se voir comme un cas particulier de l'équivalence entre les moindres carrés pondérés et les moindres carrés généralisés. C'est par exemple le cas des filtres symétriques d'Henderson qui peuvent s'obtenir par les deux méthodes.

Dans ce sens, HENDERSON (1916) a montré que les poids $w = (w_{-p}, \dots, w_f)$ associés à une moyenne mobile issue de la régression polynomiale locale par moindres carrés pondérés pouvaient s'écrire sous la forme :

$$w_i = \kappa_i P \left(\frac{i}{p+f+1} \right) \text{ où } P \text{ est un polynôme de degré } d.$$

Il a également montré l'inverse : toute moyenne mobile $\theta = (\theta_{-p}, \dots, \theta_f)$ qui préserve les tendances de degré d et dont le diagramme des coefficients change au plus d fois de signes peut être obtenue par une régression polynomiale locale de degré p estimée par moindres carrés pondérés. Pour cela il suffit de trouver un polynôme $P \left(\frac{X}{p+f+1} \right)$ de degré inférieur ou égal à d et dont les changements de signes coïncident avec les changements de signes de θ . Le noyau associé est alors $\kappa_i = \frac{\theta_i}{P \left(\frac{i}{p+f+1} \right)}$. C'est le cas de tous les filtres symétriques issues de l'approche FST et de la majorité des filtres asymétriques. L'annexe C présente les quelques poids pour lesquels il n'y a pas équivalence.

Plus récemment, LUATI et PROIETTI (2011) se sont intéressés aux cas d'équivalences entre les moindres carrés pondérés et les moindres carrés généralisés pour déterminer des noyaux optimaux (au sens de Gauss-Markov). Ils montrent que le noyau d'Epanechnikov est le noyau optimal associé à la régression polynomiale locale où le résidu, ε_t , est un processus moyenne mobile (MA) non inversible d'ordre 1 (i.e., $\varepsilon_t = (1 - B)\xi_t$, avec ξ_t un bruit blanc). Dans ce cas, la matrice Σ de variance-covariance correspond à la matrice obtenue par le critère de *smoothness* avec le paramètre $q = 2$ ($\sum_j (\nabla^2 \theta_j)^2 = \theta' \Sigma \theta$) : il y a donc équivalence avec l'approche FST. De même, le noyau d'Henderson est le noyau optimal associé à la régression polynomiale locale où le résidu est un processus moyenne mobile (MA) non inversible d'ordre 2 (i.e., $\varepsilon_t = (1 - B)^2 \xi_t$, avec ξ_t un bruit blanc).

5.5.3 RKHS et polynômes locaux

Comme montré dans la section précédente, la théorie des espaces de Hilbert à noyau reproduisant permet de reproduire les filtres symétriques par approximation polynomiale locale. Comme le montrent LUATI et PROIETTI (2011), cette théorie permet donc également de reproduire les filtres directs asymétriques (DAF), qui sont équivalents à l'approximation polynomiale locale mais en utilisant une fenêtre d'estimation asymétrique. Cependant, ils ne peuvent pas être obtenus par la formalisation de DAGUM et BIANCONCINI (2008) mais par une discrétisation différente de la formule (14) :

$$K_{d+1}(t) = \frac{\det H_{d+1}[1, t]}{\det H_{d+1}} f_0(t).$$

Dans le cas discret, $f_0(t)$ est remplacé par κ_j et en remplaçant les moments théoriques par les moments empiriques H_{d+1} devient $X_p' K_p X_p$ et les coefficients du filtre asymétrique sont obtenus en utilisant la formule :

$$w_{a,j} = \frac{\det X_p' K_p X_p[1, j]}{\det X_p' K_p X_p} \kappa_j.$$

En effet, la règle de Cramer permet de trouver une solution explicite à l'équation des moindres carrés ($X_p' K_p X_p$) $\hat{\beta} = X_p' K_p y_p$ où $\hat{\beta}_0 = \hat{m}_t$:

$$\hat{\beta}_0 = \frac{\det X_p' K_p X_p[1, b]}{\det X_p' K_p X_p} f_0(t) \quad \text{où} \quad b = X_p' K_p y_p.$$

Comme $b = \sum_{j=-h}^q x_j \kappa_j y_{t+j}$ il vient :

$$\det X_p' K_p X_p[1, b] = \sum_{j=-h}^q \det X_p' K_p X_p[1, x_j] \kappa_j y_{t+j}.$$

Et enfin :

$$\hat{\beta}_0 = \hat{m}_t = \sum_{j=-h}^q \frac{\det X_p' K_p X_p[1, j]}{\det X_p' K_p X_p} \kappa_j y_{t+j}.$$

6 Comparaison des différentes méthodes

Les différentes méthodes de construction de moyennes mobiles asymétriques sont comparées sur des données simulées et des données réelles. Pour toutes les séries, un filtre symétrique de 13 termes est utilisé. Ces méthodes sont également comparées aux estimations obtenues

en prolongeant la série par un modèle ARIMA²³ puis en appliquant un filtre symétrique de Henderson de 13 termes.

6.1 Méthodologie

En suivant une méthodologie proche de celle de DARNE et DAGUM (2009), neuf séries mensuelles sont simulées entre janvier 1960 et décembre 2020 avec différents niveaux de variabilité. Chaque série simulée $y_t = C_t + T_t + I_t$ peut s'écrire comme la somme de trois composantes :

- le cycle $C_t = \rho[\cos(2\pi t/\lambda) + \sin(2\pi t/\lambda)]$, λ est fixé à 72 (cycles de 6 ans, il y a donc 19 points de retournement détectables) ;
- la tendance $T_t = T_{t-1} + v_t$ avec $v_t \sim \mathcal{N}(0, \sigma_v^2)$, σ_v étant fixé à 0,08 ;
- et l'irrégulier $I_t = e_t$ avec $e_t \sim \mathcal{N}(0, \sigma_e^2)$.

Pour les différentes simulations, nous faisons varier les paramètres ρ et σ_e^2 afin d'avoir des séries avec différents rapports signal sur bruit :

- Fort rapport signal sur bruit (c'est-à-dire un I-C ratio faible et une faible variabilité) : $\sigma_e^2 = 0,2$ et $\rho = 3,0, 3,5$ ou $4,0$ (I-C ratio compris entre 0,9 et 0,7) ;
- Rapport signal sur bruit moyen (c'est-à-dire un I-C ratio moyen et une variabilité moyenne) : $\sigma_e^2 = 0,3$ et $\rho = 1,5, 2,0$ ou $3,0$ (I-C ratio compris entre 2,3 et 1,4) ;
- Faible rapport signal sur bruit (c'est-à-dire un I-C ratio fort et une forte variabilité) : $\sigma_e^2 = 0,4$ et $\rho = 0,5, 0,7$ ou $1,0$ (I-C ratio compris entre 8,9 et 5,2).

Pour chaque série et chaque date, la tendance-cycle est estimée en utilisant les différentes méthodes présentées dans ce rapport. Pour les régressions polynomiales locales, les filtres asymétriques sont calibrés en utilisant l'I-C ratio estimé à chaque date (en appliquant un filtre de Henderson de 13 termes) et pour la méthode FST, un quadrillage du plan est réalisé avec un pas de 0,05 et avec comme contraintes linéaires la préservation des polynômes de degrés 0 à 3. Trois critères de qualité sont également calculés :

1. Calcul du déphasage dans la détection des points de retournement. La définition ZELLNER, HONG et MIN (1991) est utilisée pour déterminer les points de retournement :

- on parle de redressement (*upturn*) lorsque l'on passe d'une phase de récession à une phase d'expansion de l'économie. C'est le cas à la date t lorsque $y_{t-3} \geq y_{t-2} \geq y_{t-1} < y_t \leq y_{t+1}$.

- on parle de ralentissement (*downturn*) lorsque l'on passe d'une phase d'expansion à une phase de récession. C'est le cas à la date t lorsque $y_{t-3} \leq y_{t-2} \leq y_{t-1} > y_t \geq y_{t+1}$.

Le déphasage est souvent défini comme le nombre de mois nécessaires pour détecter le bon point de retournement (i.e., le point de retournement sur la composante cyclique). Nous utilisons ici un critère légèrement modifié : le déphasage est défini comme le nombre de mois nécessaires pour détecter le bon point de retournement sans aucune révision future. Il peut en effet arriver que le bon point de retournement soit détecté par des filtres asymétriques mais ne le soit pas avec l'estimation finale avec un filtre symétrique (c'est le cas de 41 points de retournements sur l'ensemble des 9 séries avec les filtres asymétriques de Musgrave) ou qu'il y ait des révisions dans les estimations successives

23. Le modèle ARIMA est déterminé automatiquement en n'utilisant pas de retard saisonnier (les séries étant désaisonnalisées) et en n'utilisant aucune variable extérieure (comme des régresseurs de correction des points atypiques).

(c'est le cas de 7 points de retournements sur l'ensemble des 9 séries avec les filtres asymétriques de Musgrave). Finalement, relativement peu de points de retournement sont détectés à la bonne date avec l'estimation finale. Avec le filtre de Henderson de 13 termes, 18 sont correctement détectés sur les séries avec une faible variabilité (sur les 57 possibles), 11 sur les séries à variabilité moyenne et 12 sur les séries à forte variabilité.

2. Calcul de deux critères de révisions : la moyenne des écarts relatifs entre la q^e estimation et la dernière estimation $MAE_{fe}(q)$ et la moyenne des écarts relatifs entre la q^e et la $q + 1^e$ estimation $MAE_{qe}(q)$

$$MAE_{fe}(q) = \mathbb{E} \left[\left| \frac{y_{t|t+q} - y_{t|last}}{y_{t|last}} \right| \right] \quad \text{et} \quad MAE_{qe}(q) = \mathbb{E} \left[\left| \frac{y_{t|t+q} - y_{t|t+q+1}}{y_{t|t+q+1}} \right| \right]$$

6.2 Séries simulées

6.2.1 Comparaison des filtres polynomiaux locaux et des filtres RKHS

Du fait du fort degré de liberté de l'approche FST (dans le choix des différents paramètres), dans cette section, on ne compare les méthodes issues l'approche polynomiale locale utilisant le noyau de Henderson²⁴ et les filtres RKHS $b_{q,\Gamma}$, $b_{q,G}$ et $b_{q,\varphi}$.

En termes de déphasage dans la détection des points de retournement, c'est le filtre LC qui semble donner les meilleurs résultats lorsque la variabilité de la série est moyenne ou forte (figure 1). Les performances sont relativement proches de celles obtenues en prolongeant la série grâce à un modèle ARIMA. Toutefois, lorsque la variabilité est faible, ils semblent donner de moins bons résultats. Étonnement, c'est le filtre $b_{q,\varphi}$ qui minimise le déphasage qui donne les moins bons résultats. Les autres filtres RKHS ont également une grande variabilité en termes de déphasage. Cela peut s'expliquer par le fait que la courbe des coefficients des moyennes mobiles asymétriques sont assez éloignées des coefficients du filtre symétrique : il y a donc potentiellement beaucoup de révisions dans la détection des points de retournement. En effet, lorsque le déphasage est défini comme la première date à laquelle le bon point de retournement est détecté, c'est le filtre $b_{q,G}$ qui donne les meilleurs résultats.

Concernant les révisions, la variabilité de la série a peu d'impact sur les performances respectives des différentes méthodes mais joue sur les ordres de grandeurs. Globalement, les filtres LC minimisent toujours les révisions (voir tableau 4) et les révisions sont plus importantes avec les filtres CQ, DAF et les filtres RKHS autres que $b_{q,\varphi}$. Pour les filtres QL, il y a une forte révision entre la deuxième et la troisième estimation : cela peut venir du fait que pour la deuxième estimation (lorsque l'on connaît un point dans le futur), le filtre QL associe un poids plus important à l'estimation en $t + 1$ qu'à l'estimation en t , ce qui crée une discontinuité. Pour les filtres polynomiaux autres que le filtre LC, les révisions importantes à la première estimation étaient prévisibles au vu de la courbe des coefficients : un poids très important est associé à l'observation courante et il y a une forte discontinuité entre la moyenne mobile utilisée pour l'estimation en temps réelle (lorsqu'aucun point dans le futur n'est connu) et les autres moyennes mobiles.

24. Il est en effet difficile de comparer proprement les résultats entre les différents noyaux car le filtre symétrique n'est pas le même. Cela a pour conséquence que des points de retournement peuvent être détectés. Par exemple, pour le filtre LC, sur les trois séries ayant une variabilité moyenne, seul 1 point de retournement est correctement détecté par l'ensemble des noyaux. Toutefois, une première analyse des résultats montrent que les différents noyaux ont des performances proches en termes de déphasage et de révisions, sauf le noyau uniforme qui produit de moins bons résultats.

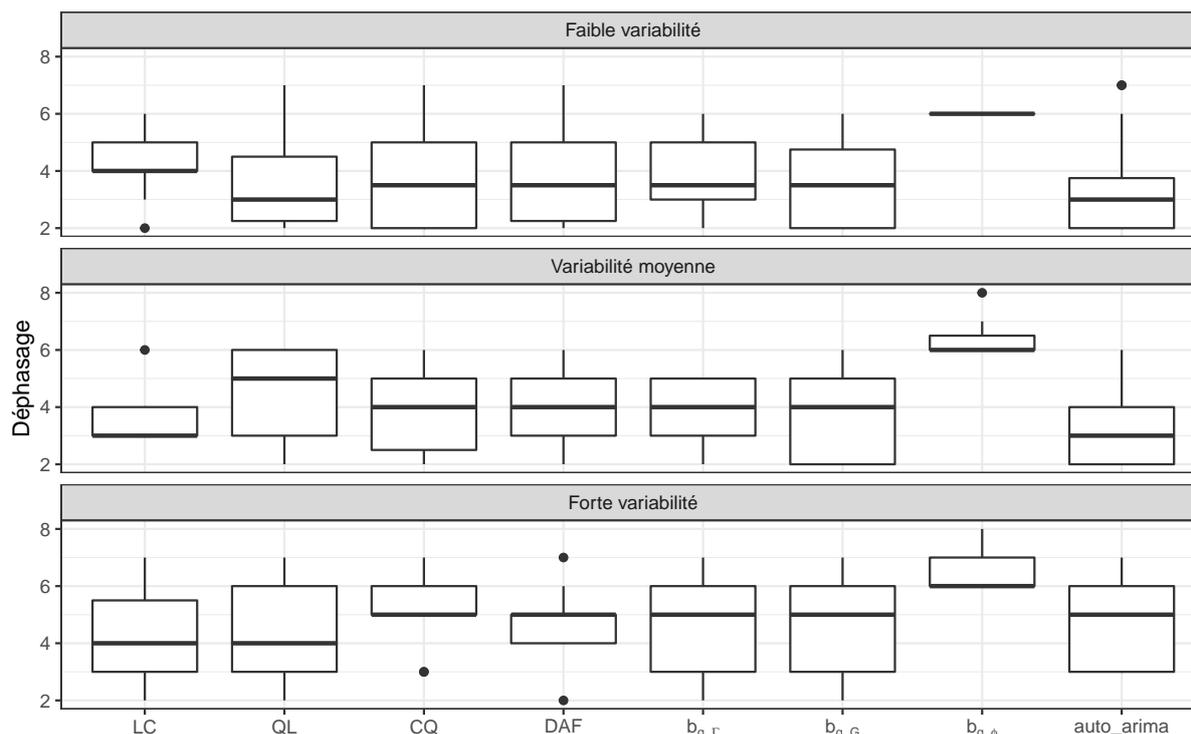


FIGURE 1 – Distribution des déphasages sur les séries simulées.

TABLE 4 – Moyenne des écarts relatifs des révisions pour les différents filtres sur les séries à variabilité moyenne.

Méthode	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
$MAE_{fe}(q) = \mathbb{E} [(y_{t t+q} - y_{t last}) / y_{t last}]$						
LC	0,21	0,10	0,03	0,03	0,03	0,01
QL	0,33	0,10	0,04	0,04	0,03	0,01
CQ	0,45	0,13	0,13	0,09	0,06	0,02
DAF	0,47	0,15	0,15	0,09	0,06	0,02
$b_{q,\Gamma}$	0,63	0,21	0,03	0,09	0,09	0,04
$b_{q,G}$	0,83	0,37	0,03	0,09	0,09	0,04
$b_{q,\phi}$	0,31	0,11	0,03	0,05	0,07	0,09
ARIMA	0,22	0,10	0,03	0,03	0,03	0,01
$MAE_{ce}(q) = \mathbb{E} [(y_{t t+q} - y_{t t+q+1}) / y_{t t+q+1}]$						
LC	0,19	0,10	0,02	0,01	0,07	0,01
QL	0,29	3,46	0,00	0,03	0,04	0,01
CQ	0,43	0,02	0,10	0,07	0,05	0,02
DAF	0,66	0,24	0,11	0,14	0,06	0,02
$b_{q,\Gamma}$	0,38	0,32	0,09	0,00	0,41	0,04
$b_{q,G}$	0,70	0,46	0,10	0,00	0,43	0,04
$b_{q,\phi}$	0,22	0,16	0,08	0,05	0,03	0,09
ARIMA	0,21	0,13	0,02	0,02	0,25	0,01

Le prolongement de la série par un modèle ARIMA donnent révisions avec les dernières estimations du même ordre de grandeur que le filtre LC mais des révisions plus importantes entre les estimations consécutives (on pouvait s’y attendre comme souligné dans la section 2.3.1).

En somme, par rapport au filtre LC, la réduction du déphasage du filtre $b_{q,G}$ se fait au coût de révisions 4 fois plus importantes lorsque la variabilité de la série est moyenne. Pour les séries à forte variabilité, les révisions sont du même ordre de grandeur mais l’écart est bien plus important pour les séries à faible variabilité.

6.2.2 Comparaison avec l’approche FST

Pour le choix des poids dans l’approche FST, l’idée retenue dans cette étude est de faire un quadrillage du plan $[0, 1]^3$ avec un pas de 0,05 et en imposant $\alpha + \beta + \gamma = 1$ ²⁵. Pour chaque combinaison de poids, quatre ensembles de moyennes mobiles sont construits en forçant dans la minimisation la préservation de polynômes de degré 0 à 3. Le filtre symétrique utilisé est toujours celui de Henderson. Ces différentes moyennes mobiles sont ensuite comparées relativement aux performances des filtres LC et, par simplification, uniquement sur les séries simulées à variabilité moyenne.

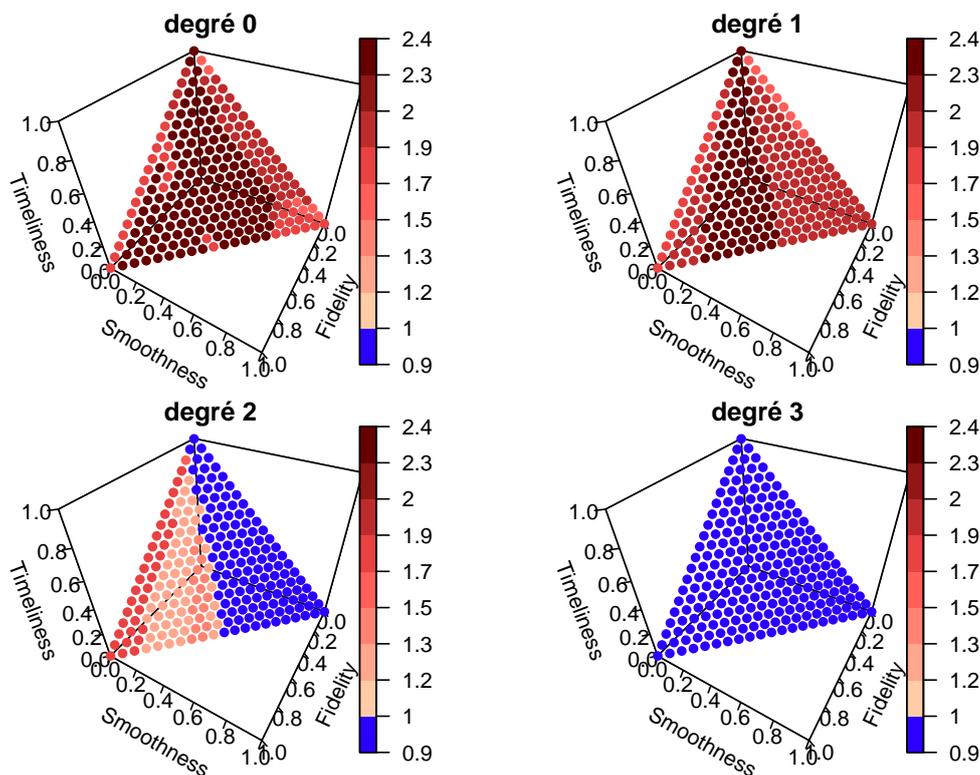


FIGURE 2 – Médiane du déphasage relatif des filtres FST par rapport au filtre LC selon les poids sur les séries simulées à variabilité moyenne.

En termes de déphasage, en médiane, les filtres qui sont plus performants que les filtres LC sont ceux qui préservent le polynômes de degré 2 et ayant un poids associé à la *fidelity* (β) inférieur à 0,5 et ceux qui préservent les polynômes de degré 3 (figure 2). En revanche, une analyse plus fine des résultats montre qu’en moyenne le déphasage est plus élevé qu’avec la méthode LC

25. Comme il n’est pas possible d’avoir un poids associé à la *timeliness* (γ) égale à 1 (sinon la fonction objectif n’est pas strictement convexe), on construit également un filtre avec un poids très proche de 1 ($1 - 1/1000$).

pour tous les filtres FST mais les résultats sont quasiment équivalents (entre 1,0 et 1,1) pour les filtres qui préservent les polynômes de degré 2 avec $\alpha = \beta = 0,05$ et $\alpha = 0,05, \beta = 0$ et ceux qui préservent les polynômes de degré 3 avec $\beta = 0$.

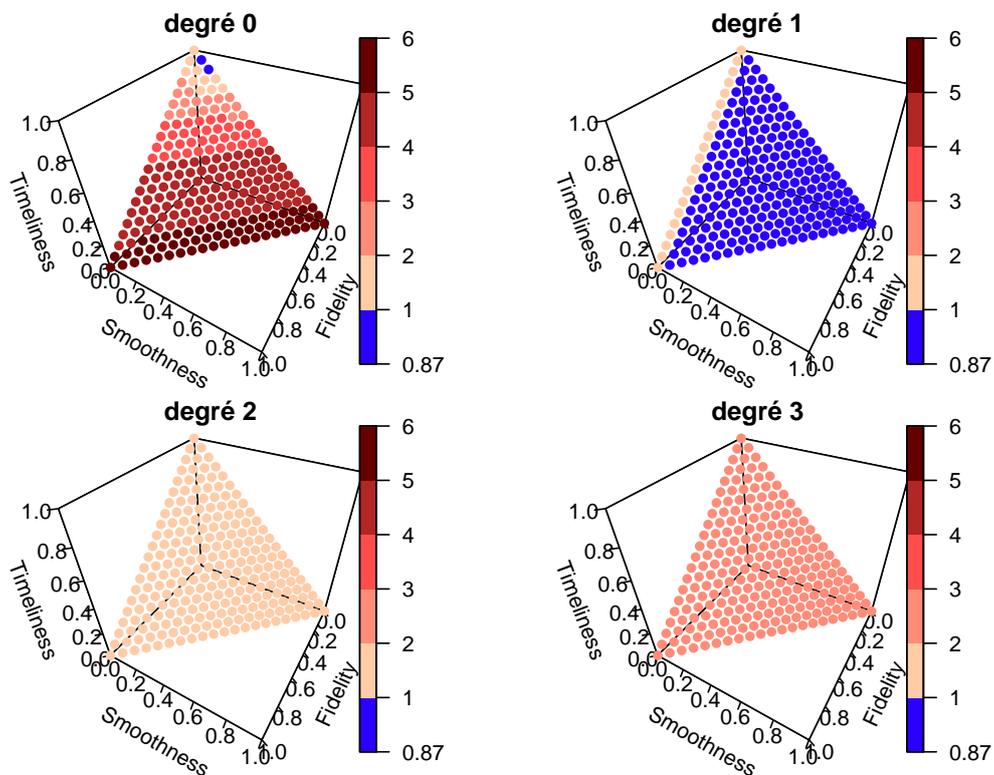


FIGURE 3 – Moyenne des écarts relatifs des révisions entre la première et la dernière estimation ($MAE_{fe}(0)$), comparativement aux révisions du filtre LC sur les séries à variabilité moyenne.

En termes de révisions (figures 3 et 4), les révisions entre la première et dernière estimation et entre la première et deuxième estimation sont inférieures à celles du filtre LC lorsque les filtres préservent les polynômes de degré 1. En revanche, avec les filtres qui préservent les polynômes de degré 3 les révisions entre la première et dernière estimation sont en moyenne plus de deux fois plus importantes qu'avec le filtre LC et sont modérément plus élevées (rapport entre 1 et 2).

En somme, même si une étude plus approfondie devrait être menée, pour l'étude de la méthode FST il paraît opportun de se concentrer sur les filtres qui préservent les polynômes de degré 2 avec un poids associé au critère *fidelity* élevé et un poids faible associé au critère *smoothness*.

6.3 Série réelle

Les différentes méthodes sont également comparées sur le point de retournement d'avril 2020 sur les ventes au détail des États-Unis (série RETAILx de la base FRED-MD MCCracken et NG (2016), utilisée en logarithme). C'est une série avec une variabilité moyenne. Les résultats des différentes estimations sont tracés dans les figures 5 à 8. Pour l'approche FST, on ne retient que les filtres obtenus avec les poids $(\alpha \beta \gamma) = (0,05 \ 0,00 \ 0,95)$, $(0,00 \ 0,05 \ 0,95)$, $(0,05 \ 0,05 \ 0,90)$ ou $(0 \ 0 \ 1)$ et en préservant les tendances de degré 2 ou 3.

Même si presque toutes les méthodes ont un déphasage d'au plus 3 mois sur ce point de retournement, les estimations intermédiaires diffèrent grandement. Il y a par exemple nettement

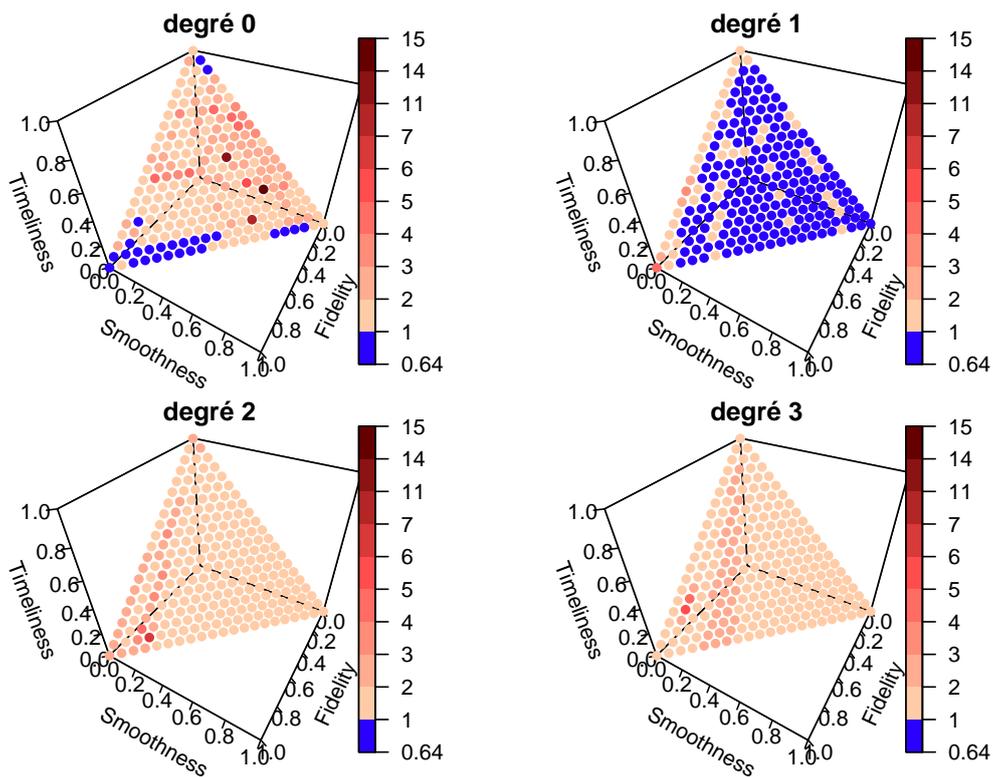


FIGURE 4 – Moyenne des écarts relatifs des révisions entre la première et la deuxième estimation ($MAE_{ce}(0)$), comparativement aux révisions du filtre LC sur les séries à variabilité moyenne.

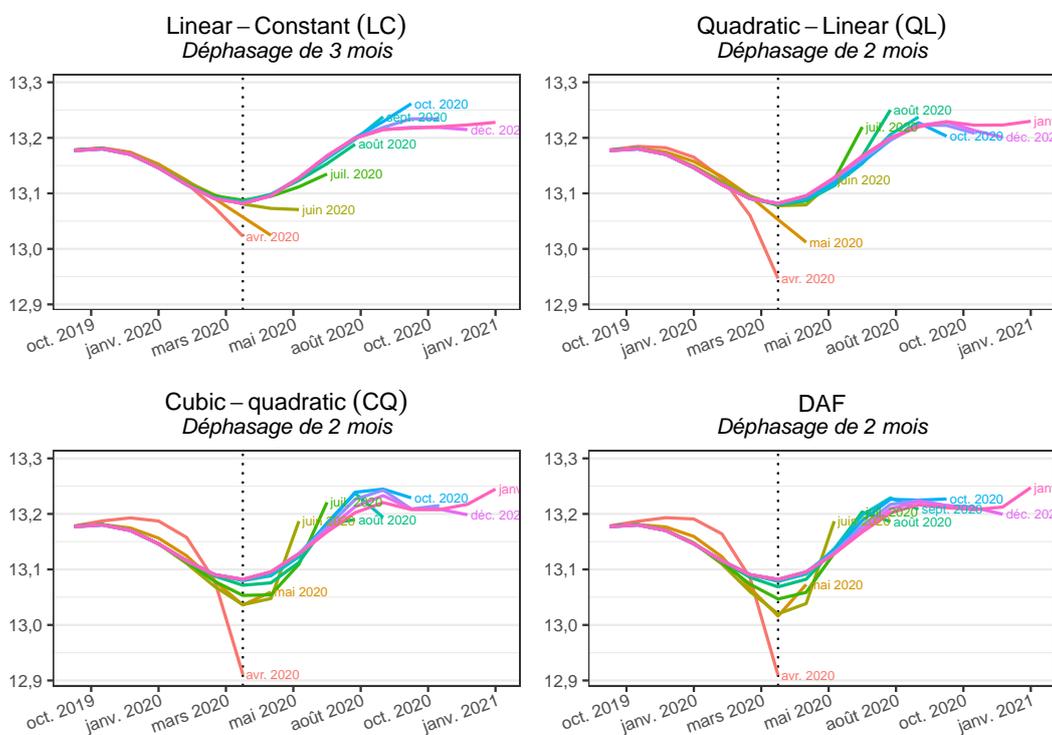


FIGURE 5 – Estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec les méthodes polynomiales locales.

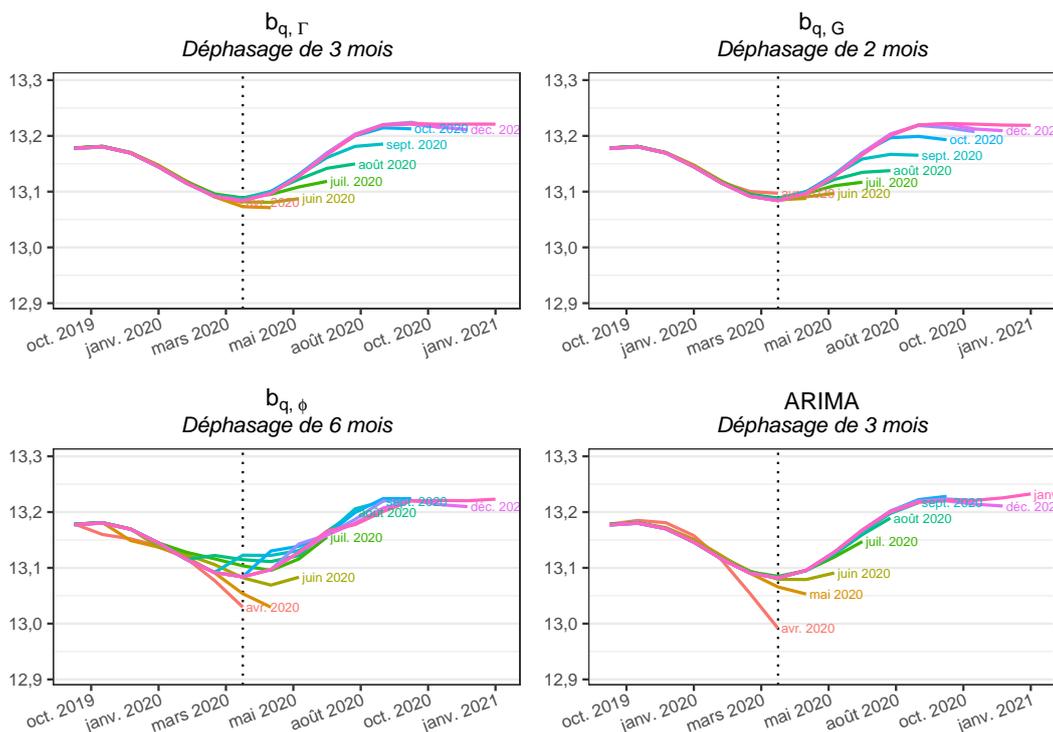


FIGURE 6 – Estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec les RKHS et en prolongeant la série par modèle ARIMA.

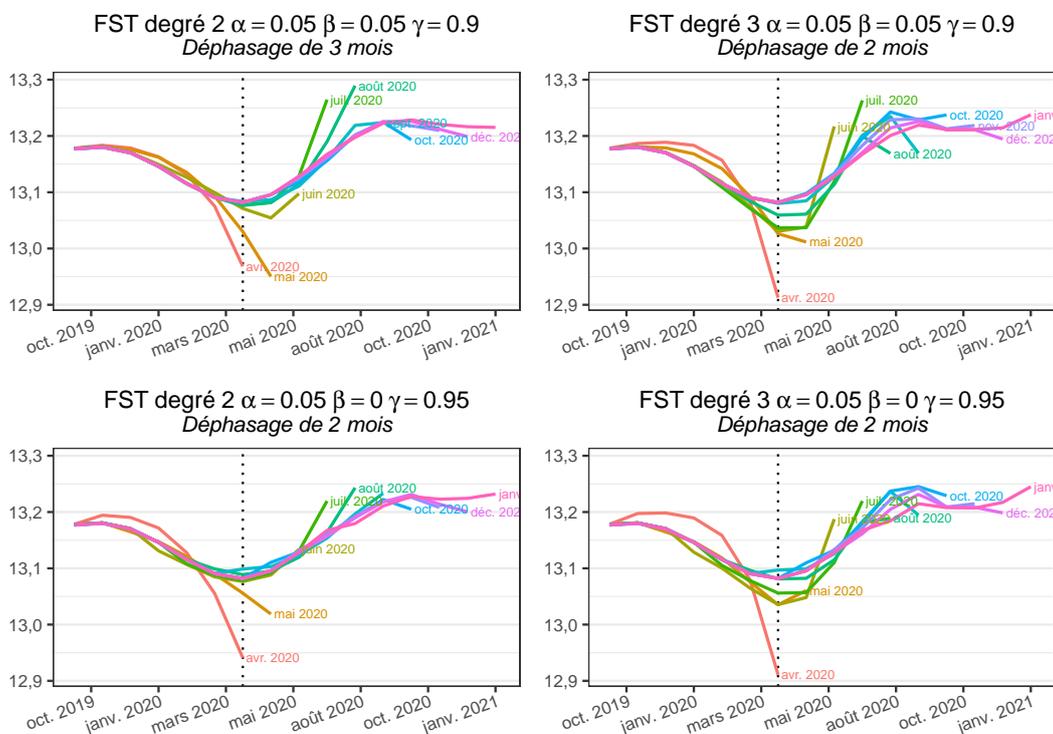


FIGURE 7 – Estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec la méthode FST et $(\alpha \ \beta \ \gamma) = (0,05 \ 0,00 \ 0,95)$ ou $(0,05 \ 0,05 \ 0,90)$.

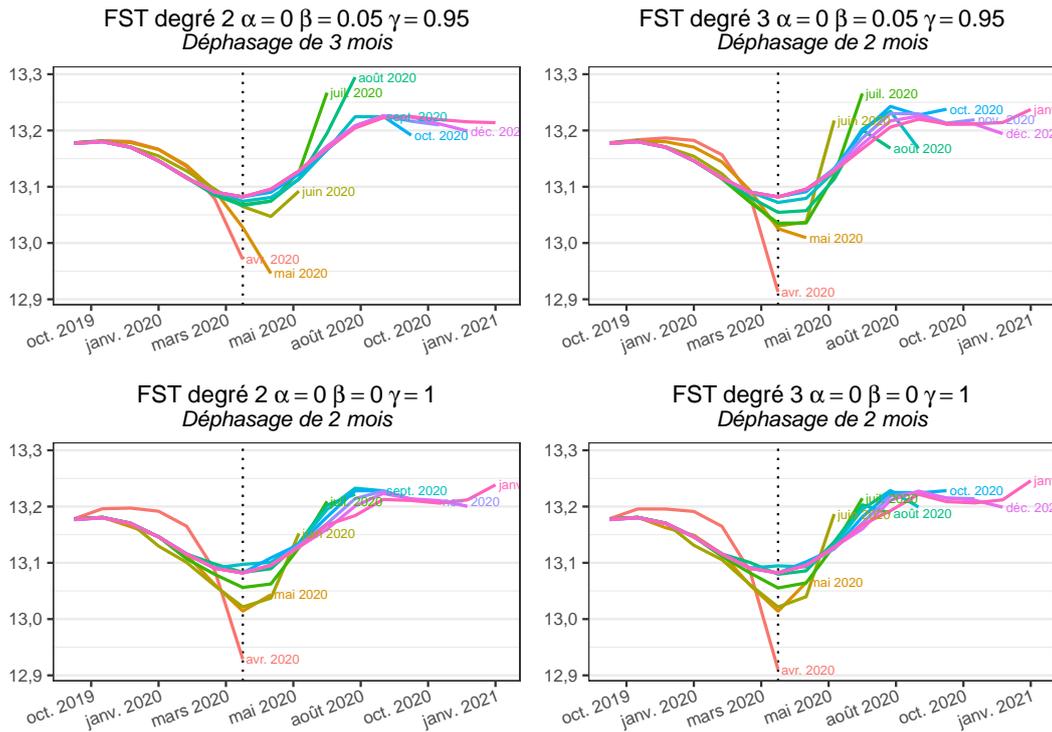


FIGURE 8 – Estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec la méthode FST et $(\alpha \ \beta \ \gamma) = (0,00 \ 0,05 \ 0,95)$ ou $(0 \ 0 \ 1)$.

plus de variabilités dans les estimations en temps réel pour les méthodes QL, CQ et DAF et les filtres FST (figures 5, 7 et 8), avec des estimations intermédiaires qui semblent peu plausibles (notamment pour la première estimation du mois d’avril 2020). Les premières estimations avec le filtre LC ne capture pas la remontée en mai et juin 2020 : cela peut provenir du fait qu’en fin de série, la tendance modélisée est linéaire. Concernant les filtres RKHS, les estimations intermédiaires du filtre $b_{q,\varphi}$ semblent très erratiques, ce qui s’explique, encore une fois, par le fait que les moyennes mobiles asymétriques utilisées lorsque l’on se rapproche du cas symétrique sont éloignées de la moyenne mobile symétrique d’Henderson. En revanche, les filtres $b_{q,\Gamma}$ et $b_{q,G}$ donnent des estimations intermédiaires proches de l’estimation finale, tout en minimisant le déphasage.

La qualité des estimations intermédiaires peut également être analysée grâce aux prévisions implicites des différentes méthodes. Pour rappel, il s’agit des prévisions de la série brute qui, en appliquant le filtre symétrique de Henderson sur la série prolongée, donne les mêmes estimations que les moyennes mobiles asymétriques. L’annexe D rassemble les différentes prévisions implicites des filtres étudiés dans cette section. Les prévisions implicites des filtres polynomiaux autres que LC, des filtres FST et du filtre $b_{q,\varphi}$ sont très peu plausibles et très éloignés des valeurs futures. Le contrecoup en mai 2020 suite à la baisse d’avril 2020 ne sont pas prévus par le modèle ARIMA mais le sont par les filtres LC et RKHS, en revanche, les prévisions à l’horizon de 6 mois peuvent être assez éloignées des valeurs attendues.

Conclusion

Pour l'analyse conjoncturel, la majorité des statisticiens fait directement ou indirectement appel à des méthodes d'extraction de la tendance-cycle. Elles sont par exemple utilisées pour réduire le bruit d'un indicateur afin d'en améliorer son analyse, et les modèles utilisées (comme les modèles de prévision) utilisent généralement sur des séries désaisonnalisées qui s'appuient sur ces méthodes.

Cette étude fait une première revue de la littérature des méthodes de construction des filtres asymétriques pour l'extraction de la tendance-cycle, utilisées pour l'estimation en temps réel (i.e., l'estimation des derniers points connus). Toutes ces méthodes peuvent se voir comme des cas particuliers d'une théorie générale de construction des moyennes mobiles. Elles sont par ailleurs facilement mobilisables et comparables grâce au *package*  `rjdfilters`. Celui-ci permet d'utiliser plusieurs outils, comme la construction des prévisions implicites, qui peuvent aider les statisticiens à évaluer la qualité des estimations récentes des différents filtres.

La comparaison des différentes méthodes, bien que perfectible, permet de tirer quelques enseignements pour la construction de ces moyennes mobiles. Premièrement, en fin de période, chercher à conserver des tendances polynomiales de degré supérieur à un (filtres QL, CQ et DAF et certains filtres FST) semble introduire de la variance dans les estimations (et donc plus de révisions) sans gain significatif en termes de détection de point de retournement. La méthode de désaisonnalisation STL utilisant les filtres DAF en fin de période, elle pourrait facilement être améliorée. Il faut en revanche que la longueur du filtre utilisé soit adapté à la variabilité de la série : si le filtre utilisé est trop long (c'est-à-dire si la variabilité de la série est « moyenne »), conserver des tendances polynomiales de degré au plus 1 (méthode LC) produit de moins bons résultats en termes de détection des points de retournement.

Deuxièmement, plus d'investigations devraient être faites sur la théorie des RKHS qui pourrait donner un bon compromis entre minimisation du déphasage et révisions. En revanche, ils devraient uniquement être utilisées pour les premières estimations de la tendance-cycle (lorsque peu de points dans le futur sont connus), lorsqu'au moins 3 points dans le futur sont connus, un filtre simple (comme le filtre LC) peut être utilisé pour limiter les révisions avec le filtre final. Par ailleurs, leur calibration peut être sujette à des problèmes d'optimisation (filtre $b_{q,\varphi}$).

Enfin, la moins bonne performance apparente de certains filtres basés sur l'approche FST ou les RKHS pourrait provenir de l'utilisation de filtres sous-optimaux lorsque que l'on s'approche du cas d'utilisation du filtre symétrique. Sur l'approche FST par exemple, rien ne justifie que l'on devrait utiliser les mêmes poids entre fidélité, lissage et temporalité pour la construction de toutes les moyennes mobiles asymétriques. Chaque moyenne mobile asymétrique devraient ainsi être construite de manière indépendante

Cette étude pourrait être étendue de plusieurs manières.

Tout d'abord, elle n'est pas exhaustive et pourrait donc être complétée.

Parmi les approches étudiées, l'extension proposée aux méthodes polynomiales locales afin d'ajouter un critère sur le déphasage pourrait donner des résultats prometteurs. Sur les méthodes FST et DAF, des méthodes de réduction de dimension pourraient être mobilisées afin d'étudier les paramètres les plus déterminants sur les performances des filtres.

Parmi les approches récentes non étudiées, nous pouvons citer VASYECHKO et GRUN-REHOMME (2014) qui utilisent le noyau d'Epanechnikov pour construire des filtres asymétriques de 13 termes (i.e., en utilisant un nombre de points dans le passé différent du filtre symétrique), et

FENG et SCHÄFER (2021) qui proposent, en fin de période, l'utilisation de poids optimaux (au sens de l'erreur quadratique moyenne) dans les régressions polynomiales locales.

Parmi les pistes d'extension, on pourrait s'intéresser à l'impact de la longueur des filtres dans la détection des points de retournement. En effet, les filtres asymétriques sont calibrés avec des indicateurs calculés pour l'estimation des filtres symétriques (par exemple pour déterminer automatiquement sa longueur), alors qu'une estimation locale pourrait être préférée. Par ailleurs, nous nous sommes concentrés uniquement sur les séries mensuelles dont le filtre symétrique est de 13 termes, mais les résultats peuvent être différents si le filtre symétrique étudié est plus long/court et si l'on étudie des séries à d'autres fréquences (trimestrielles ou journalières par exemple).

Une autre piste pourrait être d'étudier l'impact des points atypiques : les moyennes mobiles, comme tout opérateur linéaire, sont très sensibles à la présence des points atypiques. Pour limiter leur impact, dans X-13ARIMA une forte correction des points atypiques est effectuée sur la composante irrégulière avant d'appliquer les filtres pour extraire la tendance-cycle. Cela amène donc à étudier l'impact de ces points sur l'estimation de la tendance-cycle et des points de retournement, mais aussi à explorer de nouveaux types de filtres asymétriques basés sur des méthodes robustes (comme les régressions locales robustes ou les médianes mobiles).

A Synthèse des liens entre les différentes méthodes de construction de moyennes mobiles

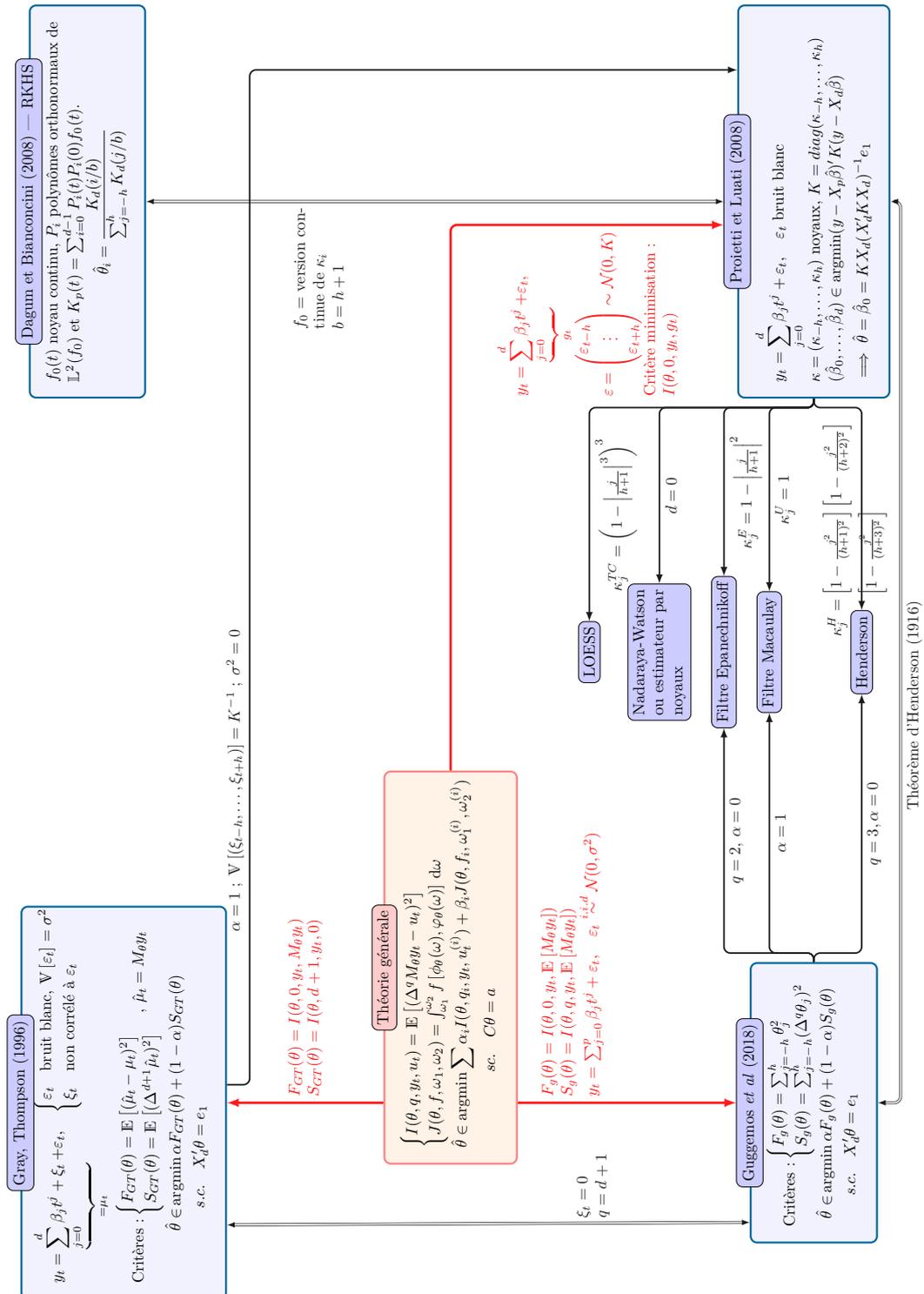


FIGURE 9 – Synthèse des méthodes de construction de moyennes mobiles symétriques $\theta = (\theta_{-h}, \dots, \theta_h)$ de $2h + 1$ termes.

Note de lecture : $X = X_d = (x_0 \ \dots \ x_d)$ avec $x_i' = ((-h)^i \ \dots \ (h)^i)$.

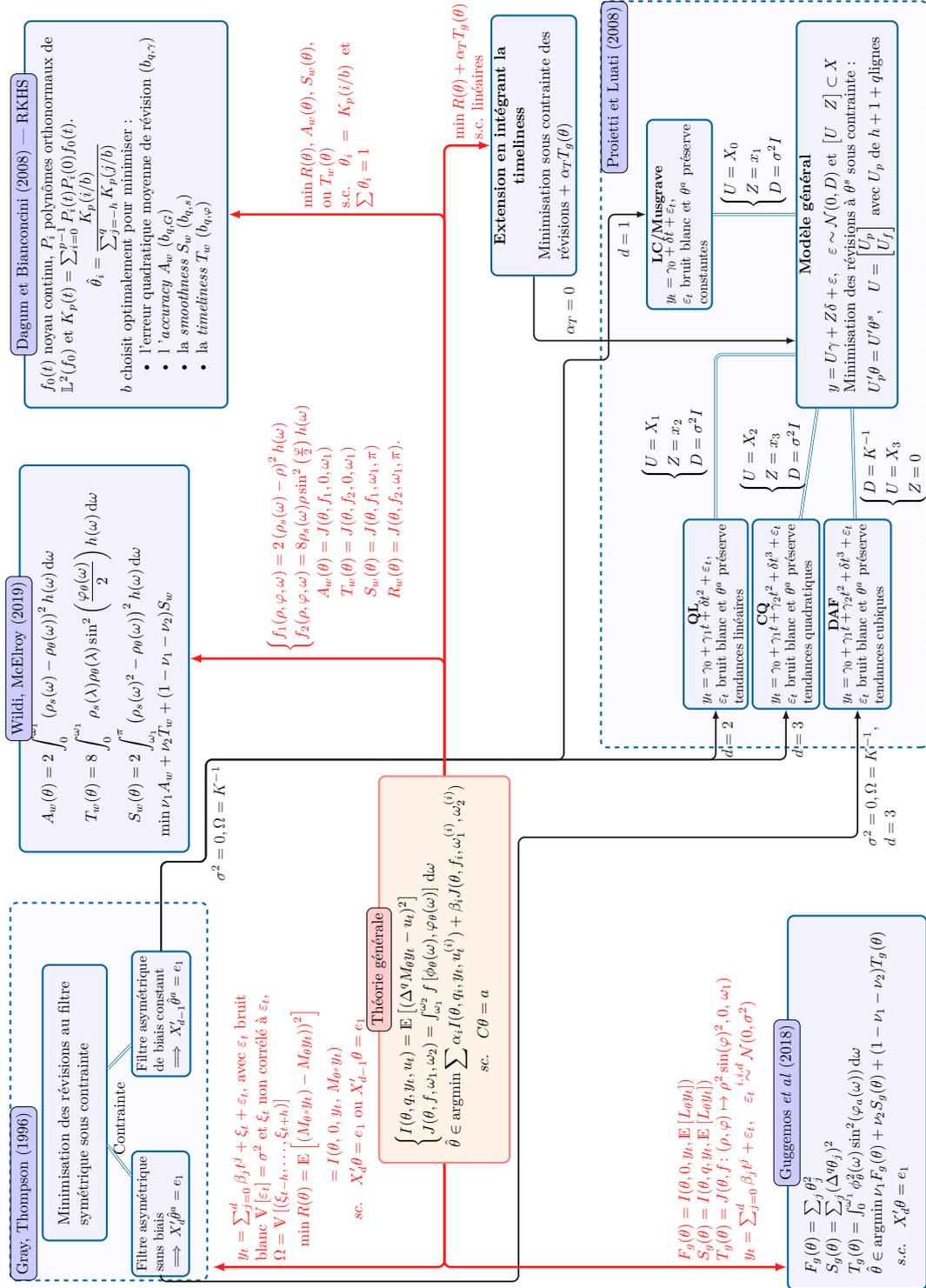


FIGURE 10 – Synthèse des méthodes de construction de moyennes mobiles asymétriques $\theta = (\theta_{-h}, \dots, \theta_q)$, $0 \leq q < h$ avec θ^s le filtre symétrique de référence de $2h + 1$ termes.
 Note de lecture : $X_d = (x_0 \ \dots \ x_d)$ avec $x_i^j = ((-h)^i \ \dots \ (q)^i)$ et $X = X_d$ avec $q = h$.

B Coefficients, fonctions de gain et de déphasage

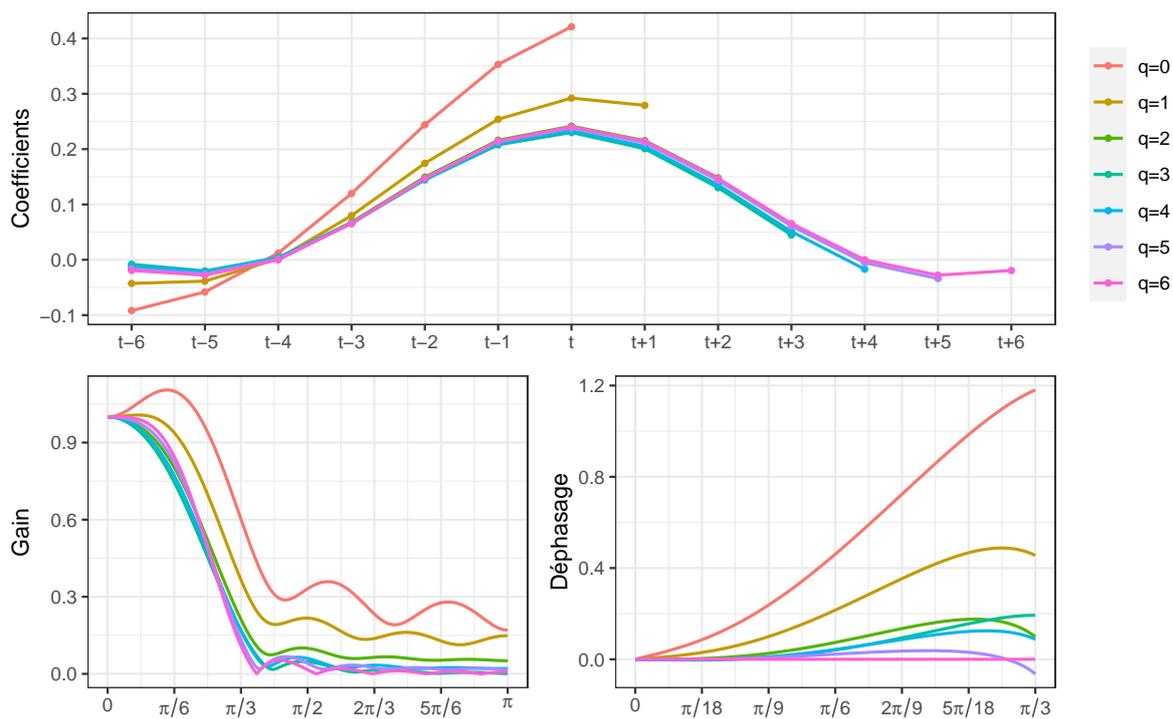


FIGURE 11 – Coefficients, fonctions de gain et de déphasage pour le filtre *Linear-Constant* (LC) avec $I/C = 3,5$.

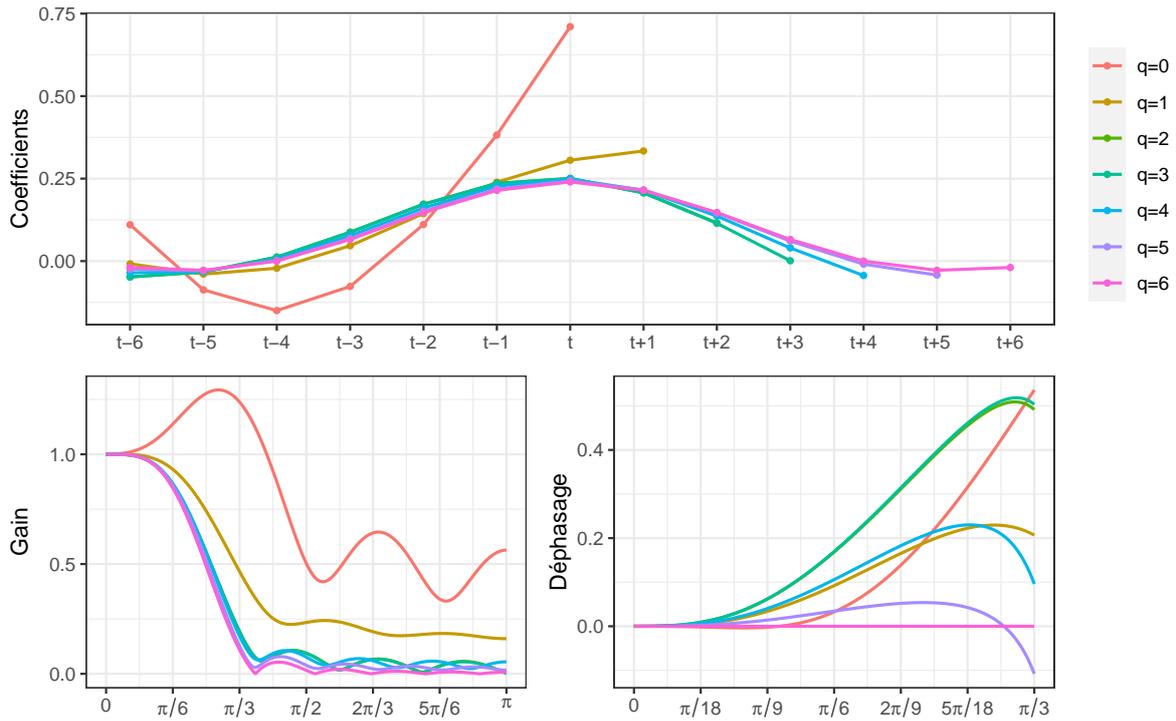


FIGURE 12 – Coefficients, fonctions de gain et de déphasage pour le filtre *Quadratic-Linear* (QL) avec $I/C = 3,5$.

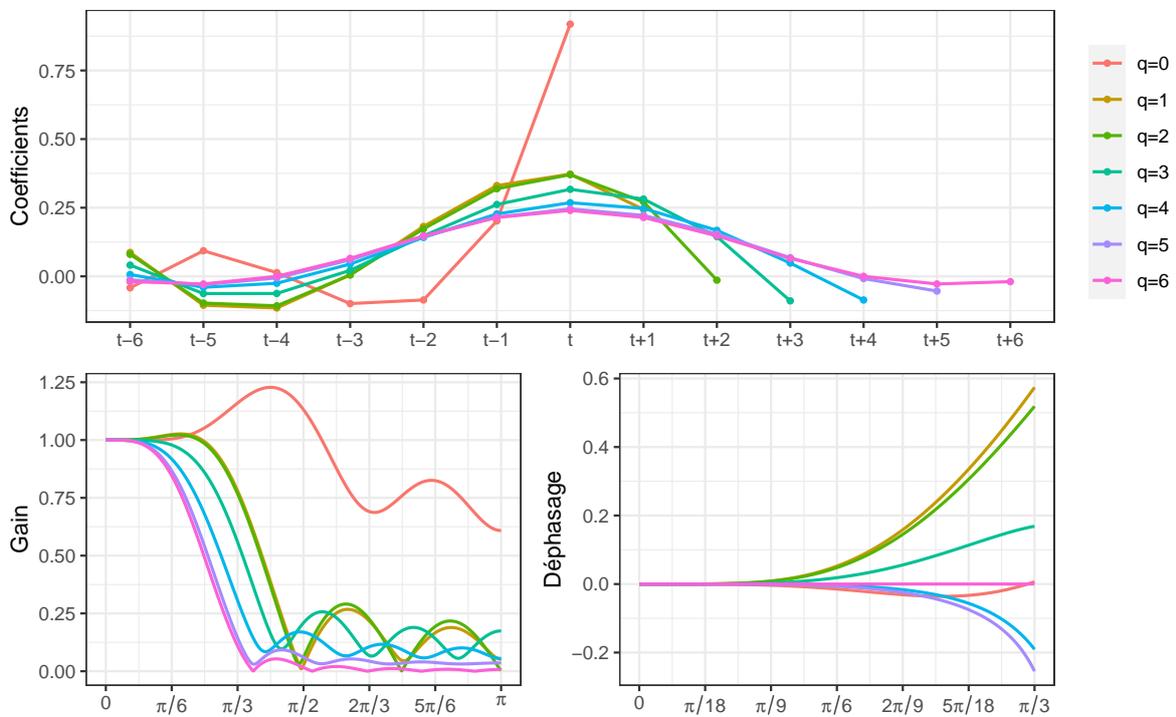


FIGURE 13 – Coefficients, fonctions de gain et de déphasage pour le filtre *Cubic-Quadratic* (QL) avec $I/C = 3,5$.

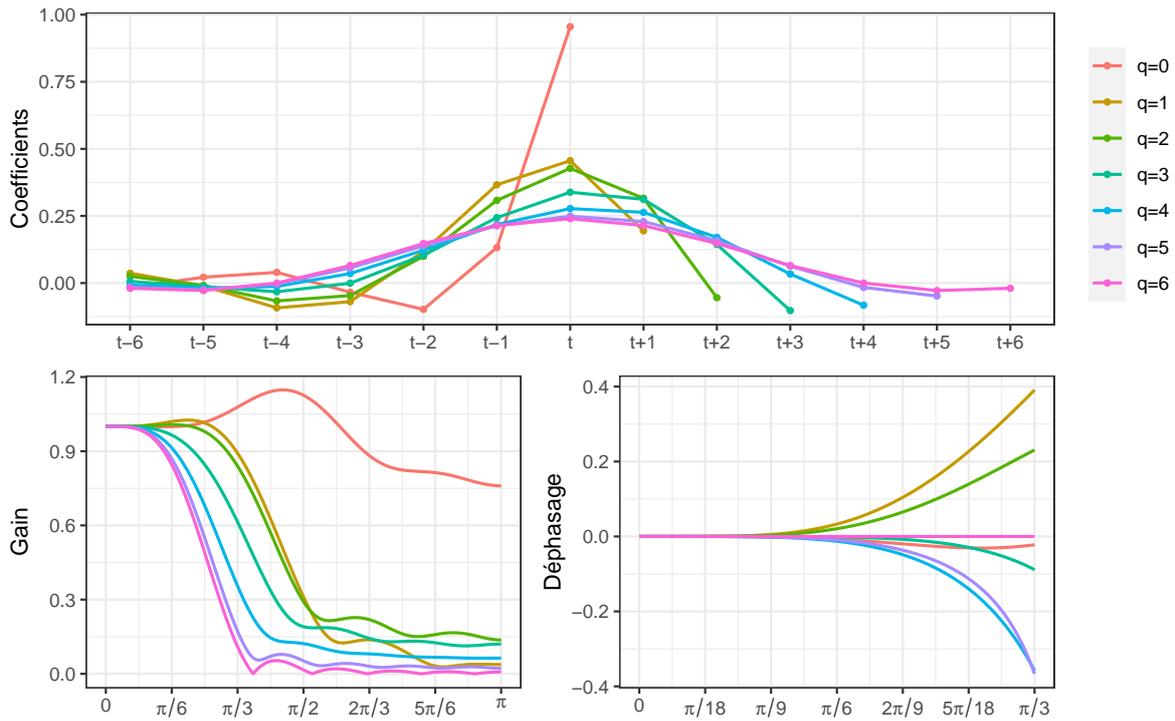


FIGURE 14 – Coefficients, fonctions de gain et de déphasage pour le filtre asymétrique direct (DAF) avec $I/C = 3, 5$.

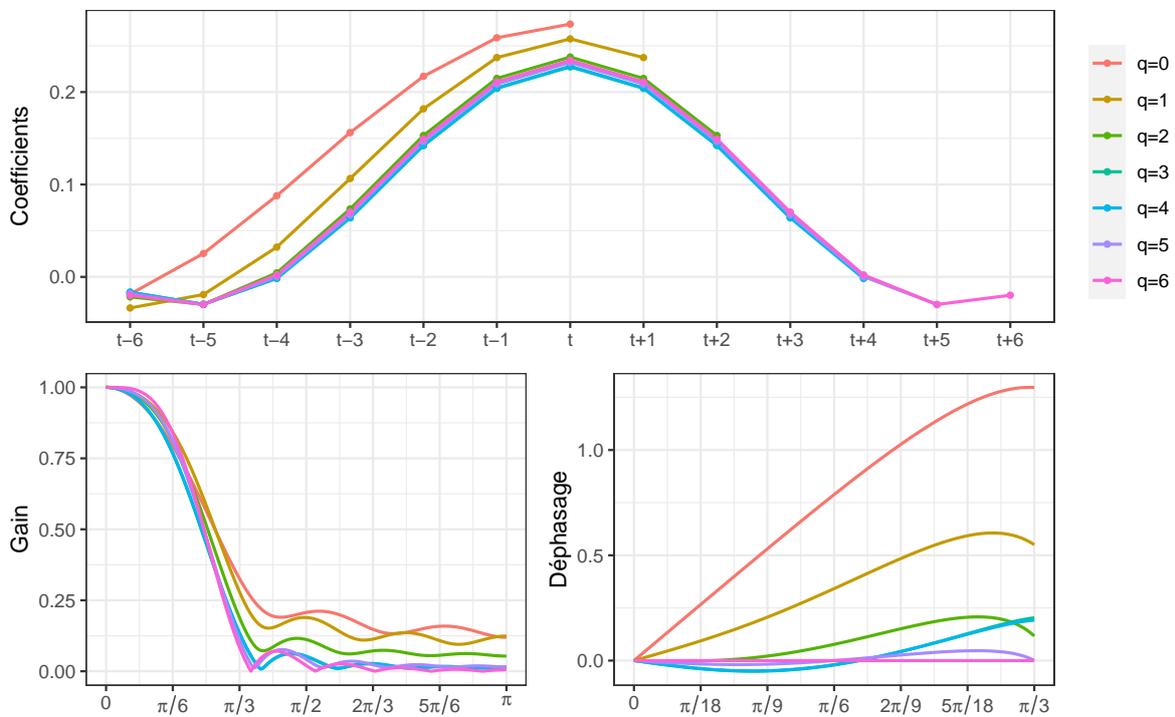


FIGURE 15 – Coefficients, fonctions de gain et de déphasage pour le filtre RKHS $b_{q,\Gamma}$.

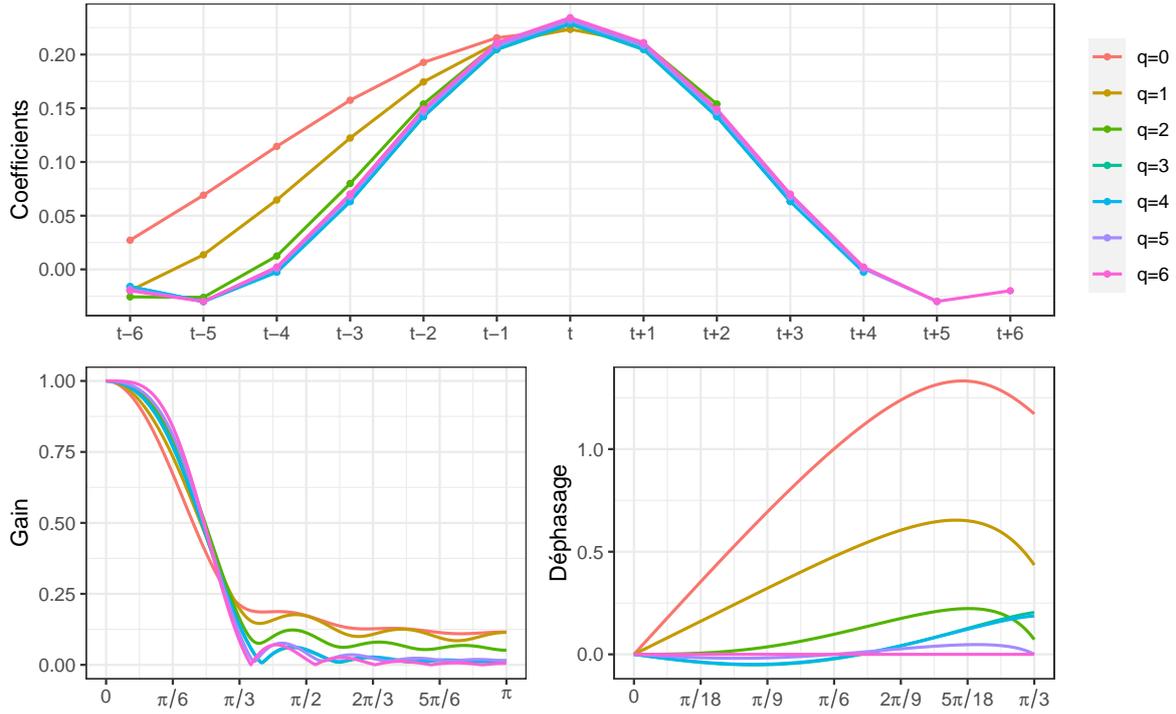


FIGURE 16 – Coefficients, fonctions de gain et de déphasage pour le filtre RKHS $b_{q,G}$.

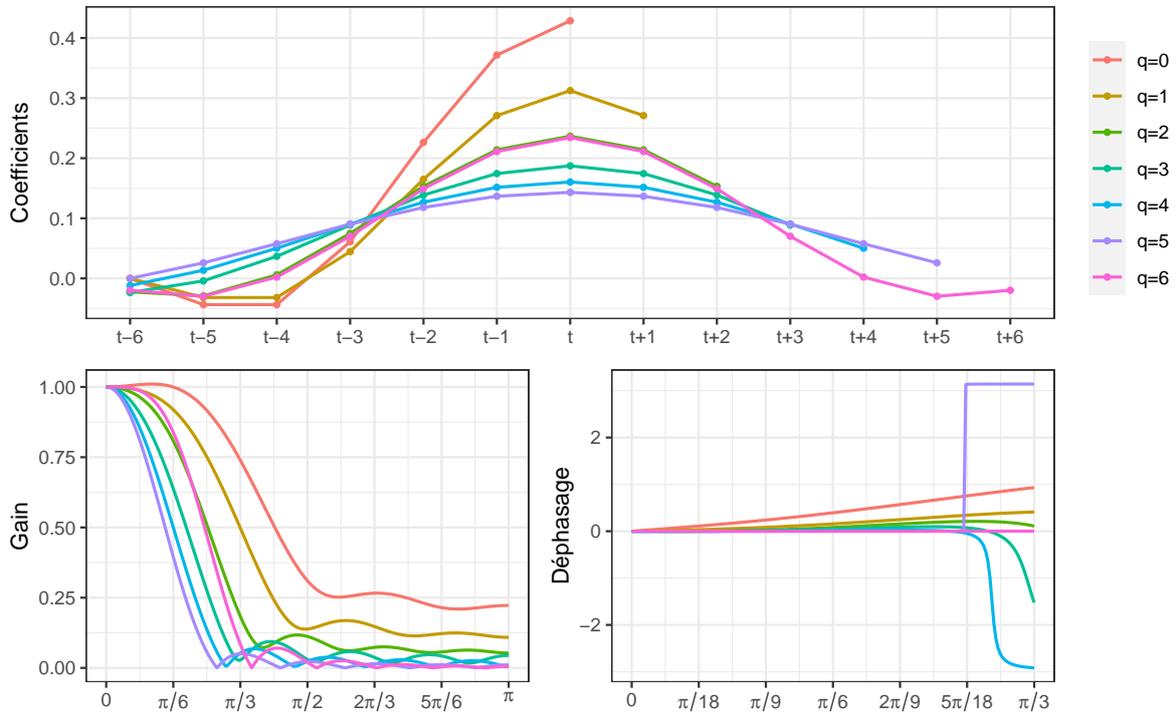


FIGURE 17 – Coefficients, fonctions de gain et de déphasage pour le filtre RKHS $b_{q,\varphi}$.

C Équivalence entre l'approche FST et les filtres polynomiaux locaux

Dans cette annexe sont tracés les rares poids pour lesquels l'approche FST **n'est pas** équivalente à l'approche polynomiale locale pour $h = 6$ (filtre symétrique de 13 termes). Lorsqu'un graphique n'est pas affiché c'est que tous les filtres FST sont équivalents une approche polynomiale locale par moindres carrés pondérés. Par exemple pour les filtres associés au filtre symétrique de 13 termes ($h = 6$, figure 18) il n'y a des graphiques que pour les filtres utilisés en temps réel ($q = 0$) et lorsque ce filtre conserve les constantes ($d = 0$), les droites ($d = 1$) et les polynômes de degré 2 ($d = 2$). Dans tous les autres cas (i.e., dès que l'on connaît au moins un point dans le futur, $q \geq 1$), il y a équivalence pour tous les poids testés²⁶.

La *smoothness* est calculée avec le paramètre $q = 3$ ($S_g(\theta) = \sum_j (\nabla^3 \theta_j)^2$), comme pour le filtre symétrique d'Henderson.

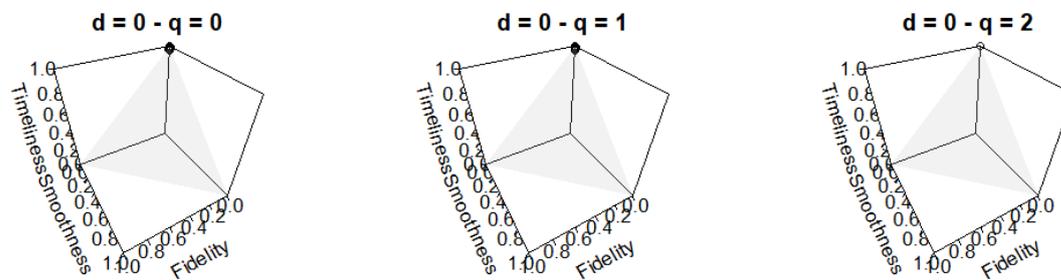


FIGURE 18 – Ensemble des poids pour lesquels la méthode FST n'est pas équivalente aux moindres carrés pondérés pour $h = 6$ (filtre symétrique de 13 termes), sous contrainte de préservation des polynômes de degré au plus 3 ($d = 0, 1, 2, 3$).

26. Un quadrillage de 200 points de l'intervalle $[0, 1]$ a été effectuée et on ne garde que l'ensemble des poids tels que leur somme fasse 1.

D Prévisions implicites pour séries RETAILx

Cette annexe montre les prévisions implicites associées aux différentes estimations de la tendance-cycle sur les ventes au détail des États-Unis autour du point de retournement d'avril 2020.

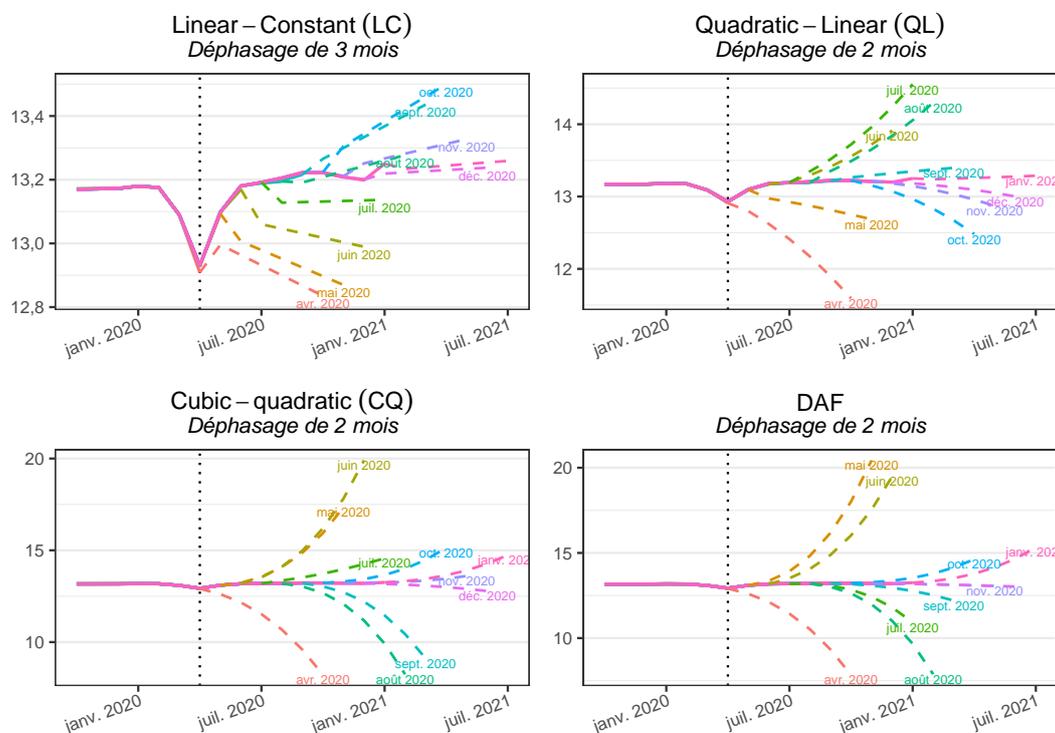


FIGURE 19 – Prévisions implicites associées aux estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec les méthodes polynomiales locales.

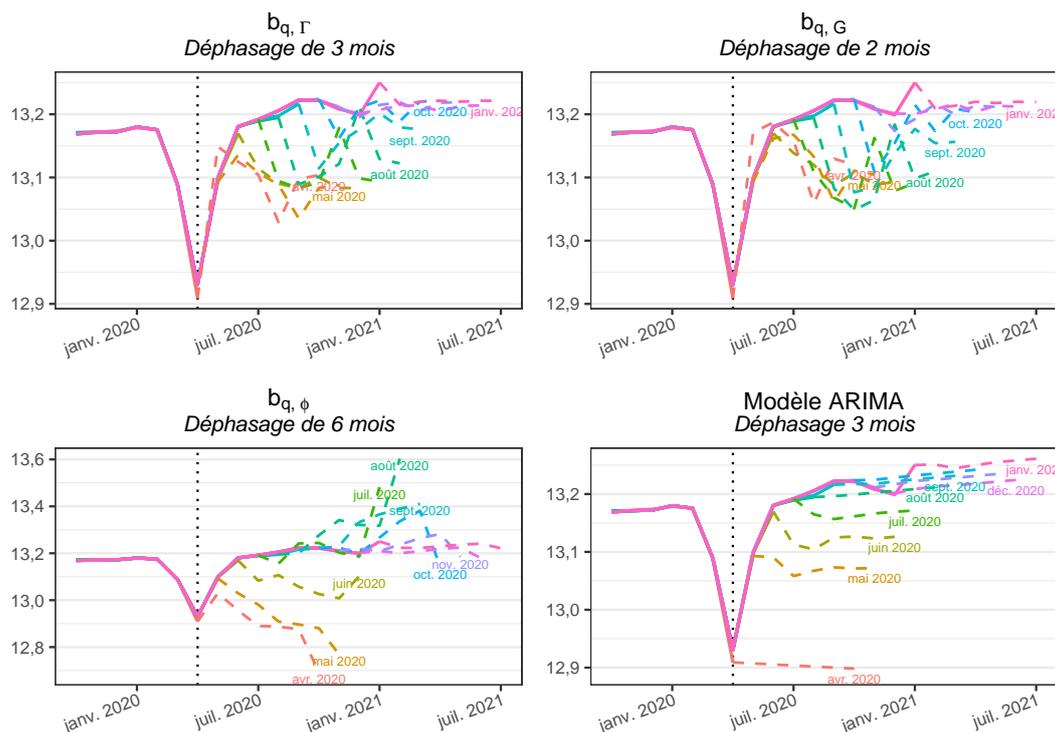


FIGURE 20 – Prévisions implicites associées aux estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec les RKHS et en prolongeant la série par modèle ARIMA.

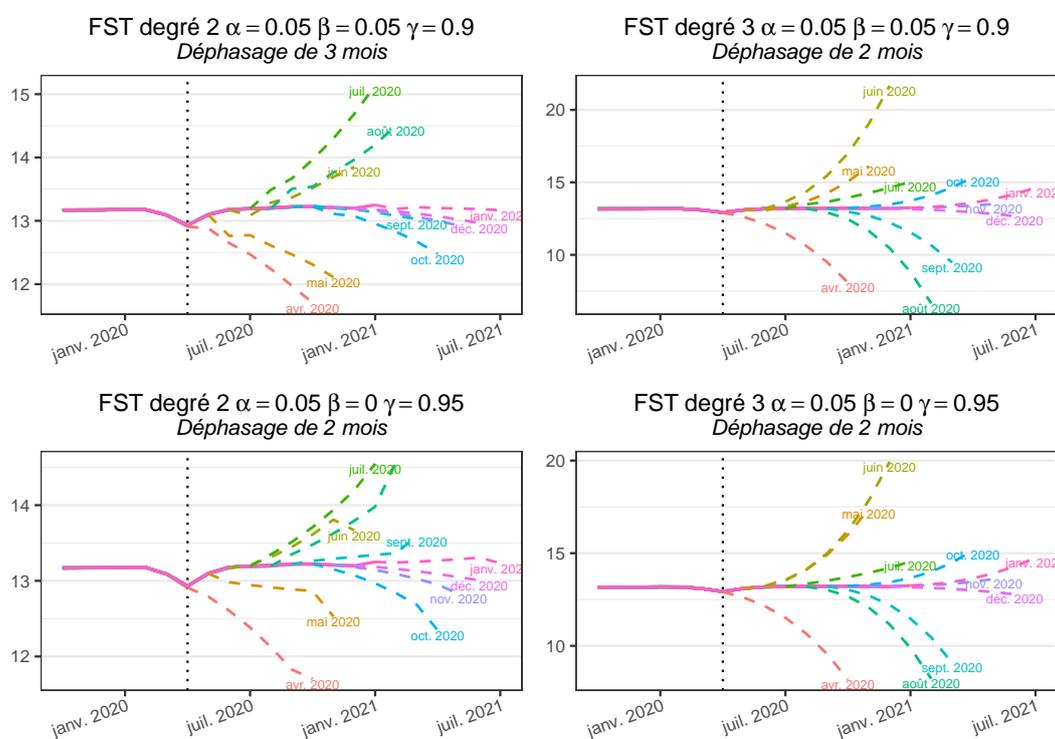


FIGURE 21 – Prévisions implicites associées aux estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec la méthode FST et $(\alpha \ \beta \ \gamma) = (0,05 \ 0,00 \ 0,95)$ ou $(0,05 \ 0,05 \ 0,90)$.

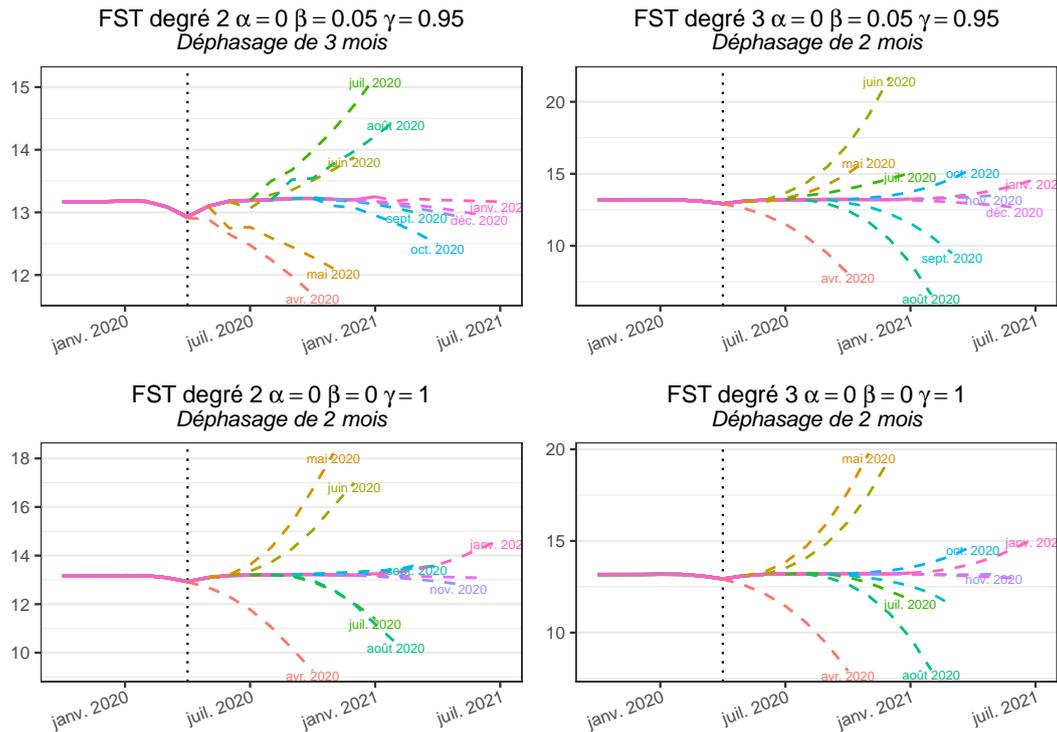


FIGURE 22 – Prévisions implicites associées aux estimations successives de la tendance-cycle des ventes au détail aux États-Unis avec la méthode FST et $(\alpha \ \beta \ \gamma) = (0,00 \ 0,05 \ 0,95)$ ou $(0 \ 0 \ 1)$.

Références

- ALEXANDROV, Theodore et alii (nov. 2012). “A Review of Some Modern Approaches to the Problem of Trend Extraction”. In : *Econometric Reviews* 31, p. 593-624. DOI : [10.1080/07474938.2011.608032](https://doi.org/10.1080/07474938.2011.608032).
- BERLINET, Alain (1993). “Hierarchies of higher order kernels”. In : *Probability Theory and Related Fields* 94, p. 489-504.
- BERLINET, Alain et Christine THOMAS-AGNAN (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer.
- CLEVELAND, Robert B. et alii (1990). “STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)”. In : *Journal of Official Statistics* 6, p. 3-73.
- CLEVELAND, William S. et Clive LOADER (1996). “Smoothing by local regression: Principles and methods”. In : *Statistical theory and computational aspects of smoothing*. Springer, p. 10-49.
- DAGUM, Estela Bee et Silvia BIANCONCINI (2008). “The Henderson Smoother in Reproducing Kernel Hilbert Space”. In : *Journal of Business & Economic Statistics* 26, p. 536-545. URL : <https://ideas.repec.org/a/bes/jnlbes/v26y2008p536-545.html>.
- (2015). “A new set of asymmetric filters for tracking the short-term trend in real-time”. In : *The Annals of Applied Statistics* 9.3, p. 1433-1458. DOI : "[10.1214/15-AOAS856](https://doi.org/10.1214/15-AOAS856)".
- (2016). *Seasonal adjustment methods and real time trend-cycle estimation*. Springer.
- DARNE, Olivier et Estelle Bee DAGUM (2009). “Performance of short-term trend predictors for current economic analysis”. In : *Economics Bulletin* 29.1, p. 79-89.
- DE FOREST, Erastus L (1877). “On adjustment formulas”. In : *The Analyst* 4.3, p. 79-86.

- FENG, Yuanhua et Bastian SCHÄFER (août 2021). *Boundary modification in local polynomial regression*. Working Papers CIE 144. Paderborn University, CIE Center for International Economics. URL : <https://ideas.repec.org/p/pdn/ciepap/144.html>.
- GRAY, Alistair et Peter THOMSON (1996). “Design of Moving-Average Trend Filters using Fidelity and Smoothness Criteria”. In : *Athens Conference on Applied Probability and Time Series Analysis*. Sous la dir. de P. M. ROBINSON et Murray ROSENBLATT. New York, NY : Springer New York, p. 205-219. ISBN : 978-1-4612-2412-9.
- GRUN-REHOMME, Michel, Fabien GUGGEMOS et Dominique LADIRAY (2018). “Asymmetric Moving Averages Minimizing Phase Shift”. In : *Handbook on Seasonal Adjustment*. URL : ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.
- GRUN-REHOMME, Michel et Dominique LADIRAY (1994). “Moyennes mobiles centrées et non-centrées. Construction et comparaison”. fr. In : *Revue de Statistique Appliquée* 42.3, p. 33-61. URL : http://www.numdam.org/item/RSA_1994__42_3_33_0/.
- HENDERSON, Robert (1916). “Note on graduation by adjusted average”. In : *Transactions of the actuarial society of America* 17, p. 43-48.
- LADIRAY, Dominique (2018). “Moving Average Based Seasonal Adjustment”. In : *Handbook on Seasonal Adjustment*. URL : ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.
- LADIRAY, Dominique et Benoît QUENNEVILLE (2011). *Seasonal Adjustment with the X-11 Method*. Springer. ISBN : 9781461301769.
- LOADER, Clive (1999). *Local regression and likelihood*. New York: Springer-Verlag, p. 290. ISBN : 0-387-9877.
- LUATI, Alessandra et Tommaso PROIETTI (2011). “On the equivalence of the weighted least squares and the generalised least squares estimators, with applications to kernel smoothing”. In : *Annals of the Institute of Statistical Mathematics* 63.4, p. 851-871. DOI : [10.1007/s10463-009-0267-8](https://doi.org/10.1007/s10463-009-0267-8). URL : <https://doi.org/10.1007/s10463-009-0267-8>.
- MACAULAY, Frederick R et alii (1931). “The smoothing of time series”. In : *NBER Books*.
- MARAVALL, A et G CAPORELLO (2004). “Program TSW: Revised reference manual”. In : URL : http://www.bde.es/f/webbde/SES/servicio/Programas_estadisticos_y_econometricos/Programas/ficheros/tswrm.pdf.
- MCCRACKEN, Michael W. et Serena NG (2016). “FRED-MD: A Monthly Database for Macroeconomic Research”. In : *Journal of Business & Economic Statistics* 34.4, p. 574-589. DOI : [10.1080/07350015.2015.1086655](https://doi.org/10.1080/07350015.2015.1086655).
- MCELROY, Tucker et Marc WILDI (2020). “The Multivariate Linear Prediction Problem: Model-Based and Direct Filtering Solutions”. In : *Econometrics and Statistics* 14.C, p. 112-130. DOI : [10.1016/j.ecosta.2019.12..](https://doi.org/10.1016/j.ecosta.2019.12..) URL : <https://ideas.repec.org/a/eee/ecosta/v14y2020icp112-130.html>.
- MCLAREN, Craig H et David G STEEL (2001). “Rotation patterns and trend estimation for repeated surveys using rotation group estimates”. In : *Statistica Neerlandica* 55.2, p. 221-238. URL : https://documents.uow.edu.au/~craigmc/sn_2001.pdf.
- MUSGRAVE, John C (1964). “A set of end weights to end all end weights”. In : *US Census Bureau [custodian]*. URL : <https://www.census.gov/ts/papers/Musgrave1964a.pdf>.
- PIERCE, David A. (sept. 1980). “Data revisions with moving average seasonal adjustment procedures”. In : *Journal of Econometrics* 14.1, p. 95-114. URL : <https://ideas.repec.org/a/eee/econom/v14y1980i1p95-114.html>.
- PROIETTI, Tommaso et Alessandra LUATI (déc. 2008). “Real time estimation in local polynomial regression, with application to trend-cycle analysis”. In : *Ann. Appl. Stat.* 2.4, p. 1523-1553. DOI : [10.1214/08-AOAS195](https://doi.org/10.1214/08-AOAS195). URL : <https://doi.org/10.1214/08-AOAS195>.

- VASYECHKO, Olga et Michel GRUN-REHOMME (2014). “A new smoothing technique for univariate time series: the endpoint problem”. In : *Economics Bulletin* 34.3, p. 1419-1430. URL : <https://EconPapers.repec.org/RePEc:ebl:ecbull:eb-13-00344>.
- WILDI, Marc et Tucker MCELROY (déc. 2013). “Optimal Real-Time Filters for Linear Prediction Problems”. In : *Journal of Time Series Econometrics* 8. DOI : [10.1515/jtse-2014-0019](https://doi.org/10.1515/jtse-2014-0019).
- (2019). “The trilemma between accuracy, timeliness and smoothness in real-time signal extraction”. In : *International Journal of Forecasting* 35.3, p. 1072-1084. URL : <https://EconPapers.repec.org/RePEc:eee:intfor:v:35:y:2019:i:3:p:1072-1084>.
- WILDI, Marc et Bernd SCHIPS (déc. 2004). *Signal Extraction: How (In)efficient Are Model-Based Approaches? An Empirical Study Based on TRAMO/SEATS and Census X-12-ARIMA*. KOF Working papers 04-96. KOF Swiss Economic Institute, ETH Zurich. DOI : [10.3929/ethz-a-004957347](https://doi.org/10.3929/ethz-a-004957347). URL : <https://ideas.repec.org/p/kof/wpskof/04-96.html>.
- ZELLNER, Arnold, Chansik HONG et Chung-ki MIN (1991). “Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques”. In : *Journal of Econometrics* 49.1-2, p. 275-304. URL : <https://EconPapers.repec.org/RePEc:eee:econom:v:49:y:1991:i:1-2:p:275-304>.