

Alain QUARTIER-LA-TENTE

ENSAE 3^{ème} année
Stage d'application
Année scolaire 2020 - 2021

Détection en temps réels des points de retournement : apport de l'utilisation des filtres asymétriques dans l'analyse conjoncturelle

Université de Nantes
Laboratoire d'économie et de management Nantes atlantique (LEMNA)

Maître de stage : Olivier DARNÉ
18/03/2021 - 09/07/2021

ENSAE Paris
TSA 26644

www.ensae.fr

Service des relations entreprises et des stages
5, avenue Henry Le Chatelier – 91764 PALAISEAU CEDEX – FRANCE – Tél : +33 (0)1 70 26 67 39 – Courriel : stage@ensae.fr

Table des matières

Résumés	1
Introduction	3
1 De la désaisonnalisation à l'estimation tendance-cycle	4
2 Quelques propriétés sur les moyennes mobiles	5
2.1 Gain et fonction de déphasage	6
2.2 Propriétés souhaitables d'une moyenne mobile	6
2.2.1 Préservation de tendances	7
2.2.2 Élimination de la saisonnalité	8
2.2.3 Réduction du bruit	8
2.3 Estimation en temps réel et moyennes mobiles asymétriques	8
2.3.1 Moyennes mobiles asymétriques et prévision	8
2.3.2 Indicateurs de qualité des moyennes mobiles asymétriques	9
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 polynomiale locale	16
5.1 Approche de Proietti et Luati	16
5.1.1 Filtres symétriques	16
5.1.2 Quelques filtres symétriques particuliers	19
5.1.3 Filtres asymétriques	19
5.1.4 Classe générale	20
5.2 Extension avec le critère de <i>timeliness</i>	22
5.3 Gray et Thomson	23
5.3.1 Filtres symétriques	23
5.3.2 Filtres asymétriques	25
5.4 Équivalence avec l'approche FST	26
5.4.1 Liens entre les critères de Gray et Thomson et ceux de Guggemos <i>et alii</i>	26
5.4.2 Équivalence avec les moindres carrés pondérés	27
6 Filtres et Reproducing Kernel Hilbert Space (RKHS)	28
6.1 Approche de Dagum et Bianconcini	28
6.2 RKHS et polynômes locaux	31
7 Comparaison des différentes méthodes	32
7.1 Méthodologie	32
7.2 Comparaison des filtres polynomiaux locaux et des filtres RKHS	33

7.3	Comparaison avec l'approche FST	35
7.4	Impact du noyau	38
7.5	Discussion	39
Conclusion		41
A	Synthèse des liens entre les différentes méthodes de construction de moyennes mobiles	42
B	Coefficients, fonctions de gain et de déphasage	45
C	Équivalence entre l'approche FST et les filtres polynomiaux locaux	50
D	Critères de minimisation dans les filtres RKHS	52
D.1	Filtre symétrique de 9 termes ($h = 4$)	52
D.2	Filtre symétrique de 13 termes ($h = 6$)	55
D.3	Filtre symétrique de 23 termes ($h = 11$)	58
E	Comparaison avec le filtre FST	62
E.1	Filtres polynomiaux locaux	62
E.2	Filtres RKHS	66
F	Décomposition avec X-13ARIMA	70
G	Exemples d'estimations de la tendance-cycle	74
Références		82


Résumés

Résumé

La crise du COVID-19 met en évidence que 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 du cycle économique. Au centre de la série, des filtres symétriques sont appliqués. 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 (déphasage).

La construction de moyennes mobiles asymétriques performantes en termes de fidélité (de préservation du signal), de révision, de lissage et de déphasage est un sujet de recherche toujours ouvert. Ce rapport décrit et compare des approches récentes pour construire des filtres asymétriques : filtres polynomiaux locaux, méthodes basées sur une optimisation des propriétés des filtres et filtres basés sur les espaces de Hilbert à noyau reproduisant (RKHS). Il décrit également comment les filtres polynomiaux locaux peuvent être étendus pour inclure un critère de temporalité afin de minimiser le déphasage. Toutes ces méthodes peuvent se voir comme un cas particulier d'une théorie générale de construction des filtres et ont été comparées en les intégrant dans la méthode de désaisonnalisation X-13ARIMA.

Ce rapport montre notamment que contraindre les moyennes mobiles asymétriques à ne conserver que les constantes (et pas nécessairement polynomiales) réduit à la fois l'erreur de révision et le déphasage dans la détection de points de retournement. Les futures études sur le sujet peuvent donc se concentrer sur ces filtres. Ce rapport met également en évidence le lien entre désaisonnalisation et extraction de tendance-cycle. Les deux ne peuvent s'étudier de manière indépendante et négliger la méthode de désaisonnalisation peut conduire à des estimations de la tendance-cycle biaisées par la présence de points atypiques mais aussi par l'introduction de faux points de retournement.

Toutes les méthodes décrites sont implémentées dans le package  `rjdfilters` (<https://github.com/palatej/rjdfilters>) et tous les résultats peuvent facilement être reproduits. Les codes utilisés, ainsi qu'une version web de ce rapport, sont disponibles sur https://github.com/AQLT/Stage_3A.


Abstract

The COVID-19 crisis highlights that business cycle analysis, and in particular the early detection of turning points, is a major topic in the analysis of economic outlook. In the business cycle analysis, estimates are usually derived from moving average (also called linear filters) techniques. In the center of the series, symmetric filters are applied. However, due to the lack of future observations, real-time estimates must rely on asymmetric moving averages. Classic asymmetric moving averages minimize revisions errors but introduce delays in the detecting turning points.

Construction of good asymmetric filters, in terms of fidelity, revisions, smoothness and timeliness, is still an open topic. 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, and have been compared by integrating them into the X-13ARIMA seasonal adjustment method.

This paper shows that 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. This report also highlights the link between seasonal adjustment method and

trend-cycle extraction methods. Both cannot be studied independently and neglecting the seasonal adjustment method can lead to trend-cycle estimates biased by the presence of outliers but also by the introduction of false turning points.

All the methods are implemented in the  package `rjdfilters` (<https://github.com/palatej/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/Stage_3A.

Introduction

La crise du COVID-19 a montré que l'analyse du cycle économique, et en particulier la détection rapide des points de retournement, est un sujet de première importance dans l'analyse de la conjoncture économique.

Les moyennes mobiles, ou les filtres linéaires, sont omniprésents dans les méthodes d'extraction du cycle économique et d'ajustement saisonnier ¹. En effet, les méthodes d'extraction de tendance-cycle sont, par construction, étroitement liées aux méthodes de désaisonnalisation puisqu'elles sont généralement appliquées sur des séries corrigées des variations saisonnières. 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 lissage ², elles induisent également des déphasages qui retardent en général la détection en temps réel des points de retournement.

L'objectif de ce rapport 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).

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 englobant les différentes méthodes développées par PROIETTI et LUATI (2008), GRAY et THOMSON (1996), GRUN-REHOMME, GUGGEMOS et Dominique LADIRAY (2018), WILDI et T. MCELROY (2019) et DAGUM et BIANCONCINI (2008) (sections 3 à 6). Enfin, dans la section 7, nous comparons les différentes méthodes en les intégrant dans X-13ARIMA-SEATS et en les appliquant sur des séries réelles.

1. 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.


2. Voir par exemple PIERCE (1980).

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 de jours ouvrables, la *désaisonnalisation* étant le processus consistant à supprimer les effets des fluctuations saisonnières et des jours ouvrables. Un effet saisonnier est un phénomène qui se répète chaque année autour de la période, avec une ampleur et une direction similaire d'une année à l'autre. Par exemple, la production automobile est généralement plus faible en été, en raison des vacances, et les ventes de chocolat sont généralement plus élevées en décembre, en raison de Noël. L'effet jours ouvrables apparaît lorsqu'une série temporelle est affectée par la composition journalière du mois. Par exemple, les ventes au détail sont généralement plus élevées le samedi que le dimanche, il est donc probable qu'elles soient plus élevées les mois contenant plus de samedis que de dimanches.

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 D. LADIRAY et B. 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). La tendance et le cycle n'étant pas observés et étant difficiles à séparer, ils sont estimés de manière conjointes dans la désaisonnalisation.
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+³, le logiciel de désaisonnalisation recommandé par Eurostat. Une interface , développée au cours de ce stage, est implémentée dans le package `rjdfilters`⁴.

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). 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).

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

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

Dans la littérature, différentes approches ont été envisagées pour l'extraction de tendance-cycle en temps réel⁵. Parmi les plus récentes, on peut citer :

- 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 T. 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 HENDERSON (1916) et 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). 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 6).

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 ce rapport, nous cherchons à proposer 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 Dominique LADIRAY (2018) (section 3.1) qui ont également proposé une procédure globale pour construire des moyennes mobiles asymétriques qui permet 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.

Dans ce rapport, nous nous concentrons sur les méthodes qui pourraient être implémentées dans X-12ARIMA. Pour maintenir la cohérence avec l'approche non paramétrique de X-12ARIMA, nous nous concentrons sur les méthodes d'extraction non paramétriques. C'est pourquoi ni les filtres de l'approche de WILDI et T. MCELROY (2019) (section 4), ni d'autres approches basées sur des modèles ne sont, pour l'instant, utilisées dans les simulations. 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 (voir par exemple Dominique LADIRAY (2018) pour plus de détails).

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

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

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 $X_t = e^{-i\omega t}$. 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)*⁶. Elle peut être réécrite en :

$$\Gamma_\theta(\omega) = G_\theta(\omega) e^{-i\Phi_\theta(\omega)},$$

où $G_\theta(\omega) = |\Gamma_\theta(\omega)|$ est la fonction de *gain* ou *d'amplitude* et $\Phi_\theta(\omega)$ est le *déphasage (phase shift ou time shift)*⁷. Pour tous les filtres symétriques on a $\Phi_\theta(\omega) \equiv 0 \pmod{\pi}$.

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

- en la multipliant par un coefficient égal à $G_\theta(\omega)$ (gain) ;
- en la “décalant” dans le temps de $\Phi_\theta(\omega)/\omega$, ce qui a un impact sur la détection des points de retournement (déphasage)⁸.

Par exemple, avec $M_{\theta_0} X_t = \frac{1}{2} X_{t-1} + \frac{1}{2} X_t$ on a :

$$\Gamma_{\theta_0}(\omega) = \frac{1}{2} + \frac{1}{2} e^{-i\omega} = |\cos(\omega/2)| e^{-i\frac{\omega}{2}}.$$

La figure 1 montre l'impact gain et du déphasage pour $\omega = \pi/2$ and $X_t = \sin(\omega t)$.

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 trois types de contraintes :

6. 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 $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{2\pi i \omega k}$.

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

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

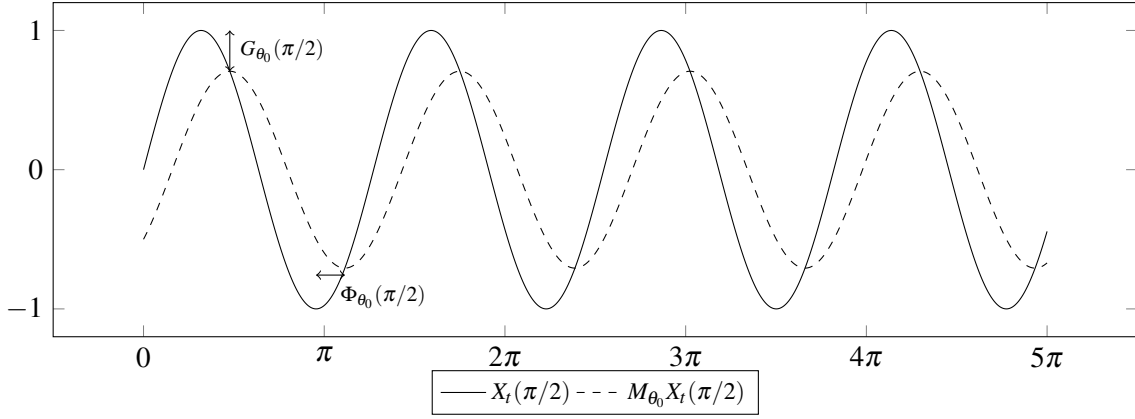


FIGURE 1 – Lissage de la série $X_t = \sin(\omega t)$ par la moyenne mobile $M_{\theta_0}X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}X_t$ pour $\omega = \pi/2$.

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

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 :

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

- 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 Élimination de la saisonnalité

Soit S_t une série à saisonnalité fixe de périodicité P ($P = 12$ pour une série mensuelle, $P = 4$ pour une série trimestrielle, etc.). On a donc $S_t = S_{t+P}$ et la somme de P termes consécutifs est constante dans le temps⁹. Ainsi, une moyenne mobile simple d'ordre P (dont tous les coefficients sont égaux à $1/P$) supprime les saisonnalités fixes.

Il est également possible de constituer des moyennes mobiles plus complexes supprimant la saisonnalité qui évoluerait polynomialement dans le temps (voir par exemple GRUN-REHOMME et Dominique LADIRAY (1994)).

2.2.3 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 : c'est notamment pourquoi ils sont préférés aux filtres asymétriques. Ils ne peuvent toutefois pas être utilisés au début et à la fin de la série car aucune valeur passée/future ne peut être utilisée.

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 suggèrent également de modéliser en fin de période une tendance polynomiale de degré au plus 3 :

« 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. »

9. Dans le cadre de la désaisonnalisation, cette constante est supposée égale à 0 car, dans le cas contraire, elle serait plutôt affectée à la composante tendance-cycle.

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*. C'est-à-dire que l'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

Dans ce rapport, nous nous concentrons sur ceux définis par GRUN-REHOMME, GUGGEMOS et Dominique LADIRAY (2018) et WILDI et T. MCELROY (2019) pour construire les filtres asymétriques.

GRUN-REHOMME, GUGGEMOS et Dominique LADIRAY (2018) proposent une approche générale pour

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	Constant biais	$\sum_{k=-p}^{+f} \theta_k - 1$
b_l	Linéaire biais	$\sum_{k=-p}^{+f} k \theta_k$
b_q	Quadratique biais	$\sum_{k=-p}^{+f} k^2 \theta_k$
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 Dominique LADIRAY (2018) et X_w critères provenant de WILDI et T. 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, $h_{RW}(x) = \frac{1}{2(1-\cos(x))}$.

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 T. 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.

Tous les indicateurs de qualité sont résumés dans le tableau 1 et sont calculables avec la fonction `rjdfilters::diagnostic_matrix`.

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 Dominique LADIRAY (2018) définissent deux critères :

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

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

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

Comme nous le montrerons dans ce rapport, 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 Dominique 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 Dominique 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).

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 5.4).

- *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 faite à partir d'une fonction de pénalité f ¹⁰ :

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

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

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

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 garantissent 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.

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 les différents critères ne sont pas normalisés : leurs valeurs ne peuvent pas être comparés 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

10. GRUN-REHOMME, GUGGEMOS et Dominique 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 $\varphi \cdot \frac{\partial f}{\partial \varphi} \geq 0$.

déterministe, on a :

$$F_g(\theta) = I(\theta, 0, y_t, \mathbb{E}[M_\theta y_t])$$

$$S_g(\theta) = I(\theta, q, y_t, \mathbb{E}[M_\theta y_t])$$

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

Encadré 1 — Filtres FST — Guggemos 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 \mathbb{R} : `rjdfilters::fst_filter()`.

4 Filtres dépendant des données : trilemme ATS

WILDI et T. 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 entrée¹¹.
- $\{TC_t\}$ 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 réponse en fréquence, de gain et de déphasage associées à ce filtre symétrique.
- $\{\widehat{TC}_t\}$ 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 réponse en fréquence, de gain et de déphasage associées à ce filtre asymétrique.

Une approche directe¹², *Direct Filter Approach* (DFA), consiste à approcher directement le signal cible par minimisation de l'erreur quadratique moyenne :

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

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

12. 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.

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-stationnaire (par exemple en imposant une cointégration entre les deux signaux et en utilisant le pseudo-spectre, voir WILDI et T. 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 :

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

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 T. MCELROY (2019) :

- dans ce rapport 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 original, 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*. 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 T. S. 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 de 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 :

$$A_w(\theta) = J(\theta, f_1, 0, \omega_1)$$

$$T_w(\theta) = J(\theta, f_2, 0, \omega_1)$$

$$S_w(\theta) = J(\theta, f_1, \omega_1, \pi)$$

$$R_w(\theta) = J(\theta, f_2, \omega_1, \pi).$$

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>.

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()`.

5 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 $[x_0 - h(x_0), x_0 + h(x_0)]$ 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 moindres carrés $\mu(x_0)$ en utilisant les observations qui sont dans $[x_0 - h(x_0), x_0 + h(x_0)]$.

5.1 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 bruit¹³. 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 :

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

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

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

13. La série est donc désaisonnalisée.

Pour estimer β il faut $H \geq d + 1$ et l'estimation est faite par moindres carrés pondérés — *weighted least squares* (WLS) —, ce qui 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 $\hat{\beta} = (X' K X)^{-1} X' K y$. Avec $e_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}'$, l'estimateur de la tendance peut donc s'écrire :

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

En somme, l'estimation de la tendance \hat{m}_t est obtenue en appliquant une moyenne mobile symétrique θ à y_t ¹⁴. 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 la forme du noyau est secondaire¹⁵ 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 fausse ce qui conduira à avoir des estimations biaisées.

5.1.1.1 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 , les observations celles 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 factor multiplicatif

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

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

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 ce rapport, à l'exception du noyau 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 (*biweight*) :

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

— $r = 2, s = 3$ 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 5.1.2 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 gaussien ¹⁶:

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

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

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

16. Dans `rjdfilters` σ^2 est fixé arbitrairement à $\sigma^2 = 0.25$.

5.1.2 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 al. (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** 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 5.1.1.1).

Par simplification, nous nous intéresserons uniquement aux filtres issus du noyau d'Henderson.

5.1.3 Filtres asymétriques

5.1.3.1 Direct asymmetric filters (DAF) 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 (prévisions générées en utilisant les mêmes contraintes polynomiales que les filtres symétriques). Elles sont également équivalentes à la troisième approche sous mêmes contraintes que le filtre symétrique. Cette méthode est appelée *direct asymmetric filter* (DAF).

Notons q le nombre d'observations futures disponibles : q varie entre 0 (filtre en temps réel) et h (filtre symétrique).

Réécrivons les matrices X , K et y :

$$X = \begin{pmatrix} X_p \\ X_f \end{pmatrix}, \quad y = \begin{pmatrix} y_p \\ y_f \end{pmatrix}, \quad K = \begin{pmatrix} K_p & 0 \\ 0 & K_f \end{pmatrix}$$

où y_p correspond aux données disponibles et y_f aux données manquantes. Le filtre DAF θ_a et les prévisions \hat{y}_f peuvent s'écrire :

$$\theta_a = K_p X_p (X_p' K_p X_p)^{-1} e_1, \quad \hat{y}_f = X_f (X_p' K_p X_p)^{-1} X_p' K_p y_p$$

De plus, on a les propriétés suivantes sur θ_a :

- il préserve les tendances polynomiales de degré d comme le filtre symétrique.
- θ_a minimise la distance pondérée (par le noyau) entre les coefficients du filtre asymétrique et ceux du filtre symétrique. Cela prouve donc l'équivalence entre les trois méthodes.

L'inconvénient de cette méthode est que les poids de $\theta_{a,0}$ sont fortement concentrés sur l'estimation courante, avec une révision importante entre $q = 0$ (filtre en temps réel) et $q = h$ (filtre symétrique, voir graphique 7). Par ailleurs, le filtre en temps réel n'a pas de fonction de gain satisfaisante : elle est proche de 1 pour toutes les fréquences et a donc un pouvoir de réduction du bruit très faible. Ainsi, même si les estimations sont sans biais, c'est au coût d'une plus grande variance dans les estimations.

5.1.4 Classe générale

Pour résoudre le problème de la variance des estimations des filtres temps réel, PROIETTI et LUATI (2008) propose 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 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 plein rang et forme un sous-ensemble des colonnes de X . L'objectif est de trouver un filtre v qui minimisent l'erreur quadratique moyenne de révision (au filtre symétrique θ) sous certaines contraintes. Ces contraintes sont représentées par la matrice $U = \begin{pmatrix} U_p' & U_f' \end{pmatrix}' : U_p' v = U' \theta$ (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 :

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

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 lorsque $D = K^{-1}$.

Lorsque $U = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}'$, $Z = \begin{pmatrix} -h & \dots & +h \end{pmatrix}'$, $\delta = \delta_1$, $D = \sigma^2 I$ 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 $R = \frac{\bar{I}}{\bar{C}} = \frac{\sum |I_t - I_{t-1}|}{\sum |C_t - C_{t-1}|}$ ($\delta_1/\sigma = 2/(R\sqrt{\pi})$). Dans l'algorithme X-13ARIMA, l'I-C ratio est utilisé pour déterminer la longueur du filtre d'Henderson¹⁷. Pour les séries mensuelles :

- si $R < 1$ un filtre d'Henderson de 9 termes est utilisé ($h = 4$) ;
- si $1 \leq R \leq 3,5$ un filtre d'Henderson de 13 termes est utilisé ($h = 6$) ;
- si $3,5 < R$ un filtre d'Henderson de 23 termes est utilisé ($h = 12$).

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, c'est l'idée suivie par PROIETTI et LUATI (2008) qui 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 ($d^* = 1$).
3. *Cubic-Quadratic* (CQ) : y_t cubique ($d = 3$) 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.

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.

17. Pour calculer l'I-C ratio, une première décomposition de la série désaisonnalisée est faite en utilisant un filtre d'Henderson de 13 termes.

TABLE 2 – Critères de qualité des filters 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	-0.407	-2.161	0.388	1.272	30.341	0.098	0.488	0.409	0.548
QL	0	0.000	-0.473	0.711	5.149	0.047	0.067	1.894	0.000	0.106
CQ	0	0.000	0.000	0.913	11.942	0.015	0.016	2.231	0.000	0.102
DAF	0	0.000	0.000	0.943	14.203	0.003	0.015	2.178	0.000	0.098
$q = 1$										
LC	0	-0.121	-0.525	0.268	0.433	4.797	0.009	0.119	0.063	0.112
QL	0	0.000	-0.061	0.287	0.707	0.694	0.005	0.192	0.007	0.042
CQ	0	0.000	0.000	0.372	0.571	0.158	0.022	0.575	0.001	0.061
DAF	0	0.000	0.000	0.409	0.366	0.061	0.020	0.760	0.000	0.059
$q = 2$										
LC	0	0.003	1.076	0.201	0.080	0.347	0.009	0.012	0.004	0.015
QL	0	0.000	0.033	0.215	0.052	2.083	0.000	0.011	0.023	0.067
CQ	0	0.000	0.000	0.370	0.658	0.131	0.021	0.558	0.001	0.055
DAF	0	0.000	0.000	0.398	0.768	0.023	0.017	0.677	0.000	0.048

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*

La méthode précédente peut être facilement étendue pour ajouter le critère de *timeliness* défini par GRUN-REHOMME, GUGGEMOS et Dominique LADIRAY (2018)¹⁸. En utilisant les mêmes notations que dans 5.1.4, θ le filtre symétrique et ν le filtre asymétrique. Notons également $\theta = \begin{pmatrix} \theta_p \\ \theta_f \end{pmatrix}$ avec θ_p de même longueur que ν , et $g = \nu - \theta_p$. Le critère de *timeliness* s’écrit :

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

18. Cela a été codé en Java par Jean Palate dans in <https://github.com/palatej/jdemetra-core>.

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

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

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

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

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.

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 :

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

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 Gray et Thomson

5.3.1 Filtres symétriques

L'approche de GRAY et THOMSON (1996) est proche de celles PROIETTI et LUATI (2008) et GRUN-REHOMME, GUGGEMOS et Dominique 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* :

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

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 :

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

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 Macalauy.

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 Filtrés 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 prédictor 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 prédictor 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 sans 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 Dominique 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 Équivalence avec l'approche FST

5.4.1 Liens entre les critères de Gray et Thomson et ceux de Guggemos *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 Dominique LADIRAY (2018) sont donc équivalentes pour la construction de filtres symétriques.

Notons $x_t = \begin{pmatrix} 1 & t & t^2 & \dots & t^d \end{pmatrix}$, $\beta_t = \begin{pmatrix} \beta_0 & \dots & \beta^d \end{pmatrix}'$.

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 $x_t = \begin{pmatrix} 1 & t & t^2 & \dots & t^d \end{pmatrix}$. 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 $S_{GT}(\theta) = \sigma^2 S_g(\theta)$.

5.4.2 Équivalence avec les moindres carrés pondérés

Du fait de la forme des filtres obtenus par la méthode de GRUN-REHOMME, GUGGEMOS et Dominique 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 obtenu 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).

6 Filtres et Reproducing Kernel Hilbert Space (RKHS)

6.1 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 la moyenne mobile 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 $\langle \cdot, \cdot \rangle$ définit 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 5.1.1.1.

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 :

$$TC_{t+j} = TC_t(j) = a_0 + a_1 j + \dots + a_d j^d + \varepsilon_{t+j}, \quad j \in \llbracket -h, h \rrbracket,$$

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 du 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 Alain BERLINET et THOMAS-AGNAN (2004)). C'est-à-dire qu'il existe 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 ¹⁹.

19. 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 $i \in \llbracket 1, d \rrbracket$.

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 A. BERLINET (1993)) :

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

Par ailleurs, dans le cas discret, la solution (11) s'écrit comme une somme pondérée du 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 ²⁰. 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 $(1 \quad t \quad t^2 \quad \dots \quad t^d)$. 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`")

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

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 :

- 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 aux 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 ($h_{WN}(x) = 1$) 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 et $b_{q,G}$ est défini par $\omega_1 = \pi$ et :

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

— $b_{q,\varphi}$ 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 $\Omega_S = [0, 2\pi/36]$ est l'intervalle de fréquences associé 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 (voir annexe D, graphiques 16, 20 et 24 sur la *timeliness*). Cela affecte uniquement les filtres qui minimisent la *timeliness* et uniquement pour certaines valeurs de q (par exemple $q = 5$ pour un filtre symétrique de 13 termes). Comme notamment montré dans QUARTIER-LA-TENTE et Dominique LADIRAY (2021), la valeur optimal retenue par défaut par `rjdfilters` produit des discontinuités dans l'estimation de la tendance-cycle : dans les cas où il y a plusieurs extremum, il serait donc préférable de retenir celui pour lequel b est maximal. C'est ce qui a été retenu dans la suite de ce rapport, c'est également cohérent avec les valeurs optimales présentés dans DAGUM et BIANCONCINI (2015).

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 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 $w_j = \frac{K_{d+1}(j/b)}{\sum_{i=-h}^p K_{d+1}(i/b)}$.

6.2 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 obtenues 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 :

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

7 Comparaison des différentes méthodes

7.1 Méthodologie

Dans cette partie, on compare les différentes moyennes mobiles en termes de délai pour détecter les points de retournement. Dans ce rapport, on définit de la façon suivante les points de retournement, en distinguant deux cas :

- les redressements (*upturn*) où l'on passe d'une phase de récession à une phase d'expansion de l'économie (comme c'est généralement le cas après une crise économique). C'est le cas à la date t lorsque $y_{t-k} \geq \dots \geq y_{t-1} < y_t \leq y_{t+1} \leq \dots y_{t+m}$.
- les ralentissements (*downturn*) où l'on passe d'une phase d'expansion à une phase de récession. C'est le cas à la date t lorsque $y_{t-k} \leq \dots \leq y_{t-1} > y_t \geq y_{t+1} \geq \dots y_{t+m}$.

En suivant ZELLNER, HONG et MIN (1991), du fait du caractère relativement lisse de la tendance-cycle, on retient les paramètres $k = 3$ and $m = 1$.

Les différents filtres sont comparés sur la base d'Eurostat des indices de production industriels déjà corrigés des jours ouvrables (`sts_inpr_m`, 2 404 séries connues jusqu'en mars 2021). Les points de retournement finaux sont ceux qui sont trouvés sur la composante tendance-cycle estimée avec la méthode X-13ARIMA sur l'ensemble de la période. Le déphasage est défini comme le nombre de mois nécessaires pour détecter le bon point de retournement sans aucune révision future (il est donc possible que ce déphasage ne soit pas calculable si pour une méthode donnée on ne détecte pas le bon point de retournement à la dernière date d'estimation).

Pour chaque série et chaque date, la méthodologie suivante est utilisée :

1. La série est désaisonnalisée avec la méthode X-13ARIMA²¹ pour extraire différents paramètres qui seront fixés entre les différentes méthodologie : la série linéarisée (i.e. : la série initiale corrigée de points atypiques par un Reg-ARIMA), la longueur des filtres saisonniers et du filtre utilisé pour extraire la tendance-cycle, le schéma de décomposition et l'I-C ratio.
2. Une nouvelle désaisonnalisation est effectuée sur la série linéarisée en utilisant l'algorithme de décomposition de X-13ARIMA et en utilisant les mêmes paramètres que précédemment et en changeant uniquement les filtres asymétriques utilisés pour extraire la tendance-cycle²².
Pour contrôler l'impact des estimations des paramètres et des points atypiques dans le modèle Reg-ARIMA model, la série linéarisée utilisée ici correspond à celle calculée sur l'ensemble de la série.
Pour la méthode de référence (X-13ARIMA), la série linéarisée est prolongée par prévision sur un an avant la décomposition mais pour toutes les autres méthodes on n'utilise pas les prévisions dans la décomposition.
Une description plus détaillée de l'algorithme de décomposition est disponible en annexe F.
3. Pour chaque estimation de la tendance-cycle, on calcule les points de redressement et de ralentissement.

Dans les différents résultats, on se restreint aux séries pour lesquelles le filtre symétrique optimal utilisé pour extraire la tendance-cycle est de longueur 13 (valeur majoritairement retenue). Il n'y a pas d'impact du schéma de décomposition sur les résultats.

7.2 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 la méthode X-13ARIMA qu'à l'approche polynomiale locale et aux filtres RKHS. Pour les filtres RKHS on ne retient que ceux qui minimisent la *timeliness* puisque dans notre cas on souhaite minimiser le délai dans la détection de point de retournement.

Le tableau 3 compare les différentes méthodes pour la détection des points de retournement durant la crise du COVID-19 (i.e. : durant l'année 2020). La méthode LC (i.e. : les filtres de Musgrave) semble être, en distribution, plus performante que les autres méthodes (avec des performances similaires à X-13ARIMA) et la méthode la moins performante semble être les filtres RKHS obtenus en minimisant la *timeliness*. La moins bonne performance des filtres QL, CQ et DAF par rapport au filtre LC pourrait être expliquée par la plus grande volatilité des estimations du fait des contraintes plus fortes sur la préservation des tendances polynomiales.

Deux exemples sont présentés dans l'annexe G : un où la méthode X-13ARIMA détecte plus rapidement le point de retournement que les autres et un où c'est l'inverse. On retrouve le fait que les estimations avec les méthodes QL, CQ et DAF sont fortement bruitées avec des estimations qui ne sont pas plausibles (notamment celles autour de points de retournement).

Le tableau 4 montre la distribution de la racine carrée de l'erreur quadratique moyenne de révision

21. On utilise pour cela la fonction `x13()` du package `RJDemetra` avec le paramètre `spec = "RSA3"` (détection automatique des différents paramètres sans correction des jours ouvrables).

22. Cela peut être fait en utilisant la fonction `rjdfilters::x11()`.

TABLE 3 – Déciles du nombre de mois nécessaires pour détecter correctement un point de retournement de l'année 2020 (771 observations).

	X13-ARIMA	RKHS ($b_{q,\varphi}$)	LC	QL	CQ	DAF
Min	2,0	2,0	2,0	2,0	2,0	2,0
D1	2,0	6,0	3,0	2,0	2,0	2,0
D2	3,0	6,0	3,0	3,0	2,0	2,0
D3	3,0	6,0	4,0	3,0	3,0	3,0
D4	3,0	6,0	4,0	4,0	5,0	5,0
Méd.	4,0	6,0	4,0	4,0	6,0	6,0
D6	4,0	6,0	4,0	6,0	6,0	6,0
D7	5,0	6,0	5,0	7,0	7,0	7,0
D8	5,0	7,0	5,0	7,0	8,0	8,0
D9	9,0	8,0	8,0	9,0	9,0	9,0
Max	14,0	14,0	14,0	14,0	14,0	14,0
Moy	4,4	6,6	4,7	5,1	5,4	5,5

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

(RMSE) entre la première et la dernière estimation de la tendance-cycle. La meilleure méthode est cette fois-ci la méthode actuelle (X-13ARIMA), suivie du filtre RKHS et ensuite la même hiérarchie que précédemment se retrouve entre les différentes méthodes polynomiales locales. Toutefois, comme souligné dans la partie 2.3.1, la méthode X-13ARIMA revient à utiliser un filtre asymétrique dont les poids sont optimisés pour minimiser la révision entre la première et la dernière estimation. Par construction, l'indicateur de révision retenu privilégie donc la méthode X-13ARIMA : il faudrait utiliser un indicateur qui prennent également en compte les révisions entre les autres horizons de prévision (par exemple entre la première et la deuxième estimation, deuxième et dernière estimation, etc.).

TABLE 4 – Déciles de la racine carrée de l'erreur quadratique moyenne entre la première et la dernière estimation sur l'année 2020 (771 observations).

	X13-ARIMA	RKHS ($b_{q,\varphi}$)	LC	QL	CQ	DAF
Min	0,1	0,1	0,1	0,2	0,2	0,2
D1	0,3	0,4	0,4	0,5	0,6	0,6
D2	0,5	0,5	0,6	0,8	0,9	0,9
D3	0,6	0,8	0,8	1,1	1,3	1,3
D4	0,8	1,0	1,1	1,5	1,6	1,7
Méd.	1,1	1,4	1,5	2,1	2,3	2,4
D6	1,5	1,9	2,0	2,8	3,1	3,1
D7	2,1	2,6	2,8	3,8	4,1	4,2
D8	3,1	3,7	4,0	5,1	5,7	5,8
D9	4,6	5,8	6,2	8,3	9,7	9,9
Max	25,0	36,6	41,0	54,8	55,2	52,9
Moy	2,1	2,7	2,8	3,9	4,2	4,3

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

Le même exercice peut être fait en analysant le déphasage dans la détection des points de retournement

pendant la crise financière de 2008. En revanche, contrairement à la crise du COVID-19, toutes les séries n’ont pas été impactées au même mois : une analyse plus approfondie devrait être faite pour vérifier si les points détectés sont réellement des points de retournement. Même les résultats du tableau 5 doivent être étudiés avec précaution, on observe la même hiérarchie entre les méthodes que celle trouvée pendant l’étude de la crise du COVID-19. Par ailleurs, pour plus de 80 % des séries et pour toutes les méthodes, le déphasage dans la détection des points de retournement est plus grand que 6 mois. C’est-à-dire que dans la majorité des cas, la détection des points de retournement n’est pas stable, même si l’on utilise un filtre symétrique pour l’estimation de la tendance-cycle. Cela vient notamment de l’algorithme de désaisonnalisation qui, pour estimer les coefficients saisonniers utilisés pour corriger la série où l’on applique les filtres pour extraire la tendance-cycle, a besoin de bien plus que 6 mois dans le futur. Par exemple, en utilisant les paramètres les plus souvent utilisés dans X-13ARIMA, le filtre final symétrique utilisé pour extraire la tendance-cycle a 169 termes !

TABLE 5 – Déciles du nombre de mois nécessaires pour détecter correctement un point de retournement entre 2007 et 2008 (2412 observations).

	X13-ARIMA	RKHS ($b_{q,\varphi}$)	LC	QL	CQ	DAF
Min	2,0	5,0	2,0	2,0	2,0	2,0
D1	6,0	7,0	4,0	7,0	7,0	8,0
D2	15,0	15,0	14,2	15,0	16,0	16,0
D3	21,0	21,0	21,0	21,0	22,0	22,0
D4	26,0	27,0	26,0	27,0	27,0	28,0
Méd.	31,0	33,0	32,0	33,0	33,0	33,0
D6	38,0	41,0	39,0	40,0	40,0	40,0
D7	49,0	51,7	50,0	50,0	50,7	51,0
D8	67,8	71,0	68,0	68,0	68,0	68,0
D9	125,0	129,0	127,0	127,0	127,0	127,0
Max	169,0	169,0	169,0	169,0	169,0	169,0
Moy	45,7	47,7	45,9	46,5	47,0	47,2

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

7.3 Comparaison avec l’approche FST

Pour limiter les degrés de liberté dans le choix des poids dans l’approche FST, une approche peut être de les sélectionner comparativement à une autre moyenne mobile. En effet, si l’on se concentre sur les trois critères de *fidelity*, *smoothness* et *timeliness*, comparativement à une moyenne mobile fixée, un ensemble de poids peut être considéré comme optimal s’il conduit à un filtre de meilleure qualité (selon ces trois critères) avec les mêmes contraintes polynomiales (ou des contraintes polynomiales plus grandes). L’annexe E trace l’ensemble de ces poids pour les filtres locaux polynomiaux et les filtres RKHS mais concentrons-nous ici sur le cas $h = 6$ (filtre symétrique de 13 termes).

Parmi l’ensemble des poids “admissibles” (i.e. : un ensemble de poids minimisant les 3 critères par rapport à un filtre de référence), plusieurs méthodes peuvent être utilisés pour construire les moyennes mobiles asymétriques. Ici nous en retenons trois :

1. on retient la moyenne mobile qui minimise la *timeliness* pour limiter au mieux le déphasage (i.e. :

- celle dont le poids associé à ce critère est le plus fort) ;
2. pour éviter d'introduire trop de variance dans les estimations, on retient la moyenne mobile dont le poids associé à la *timeliness* est la plus faible ;
 3. on adopte une position médiane en retenant la moyenne mobile dont le poids associé à la *timeliness* est médian parmi l'ensemble des poids admissibles.

Parmi les approches polynomiales, le filtre LC est celui qui semble donner les meilleurs résultats : nous nous restreignant donc à la comparaison à cette méthode. Pour le choix du degré de préservation, pour la première méthode, forcer, lorsque c'est possible, la moyenne mobile à préserver les polynômes de degré 1 ne change pas les poids retenus : c'est ce qui a été fait pour cette méthode. Pour les deux autres, nous nous sommes restreints aux moyennes mobiles préservant les constantes uniquement (comme le filtre LC). Par simplicité, le filtre FST optimal associé au filtre RKHS n'a été étudié qu'avec la première méthode. Ces poids sont présentés dans le tableau 6, les coefficients associés sont présentés dans l'annexe B dans les figures 9 à 12. Le filtre utilisé pour l'estimation en temps-réel ($q = 0$) qui minimise la *timeliness* par rapport au filtre LC accorde, pour l'estimation de la tendance-cycle en t , un poids plus important à la valeur observée en $t - 1$ qu'à celle observée en t (figure 9) : cela ne paraît pas plausible économique et justifie de ne pas se restreindre uniquement à la première méthode de sélection des poids.

TABLE 6 – Poids retenus pour construire des filtres avec l'approche FST

	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
LC min. timeliness						
Smoothness	0,010	0,010	0,005	0,020	0,005	0,005
Timeliness	0,870	0,910	0,985	0,975	0,990	0,990
Fidelity	0,120	0,080	0,010	0,005	0,005	0,005
LC max. timeliness						
Smoothness	0,005	0,010	0,050	0,265	0,550	0,715
Timeliness	0,845	0,880	0,915	0,730	0,445	0,280
Fidelity	0,150	0,110	0,035	0,005	0,005	0,005
LC timeliness médiane						
Smoothness	0,005	0,010	0,015	0,125	0,205	0,305
Timeliness	0,860	0,890	0,945	0,845	0,750	0,660
Fidelity	0,135	0,100	0,040	0,030	0,045	0,035
RKHS min. timeliness						
Smoothness	0,005	0,010	0,005	0,005	0,005	0,020
Timeliness	0,885	0,925	0,990	0,990	0,990	0,975
Fidelity	0,110	0,065	0,005	0,005	0,005	0,005

En termes de détection des points de retournement (tableau 7), le déphasage semble plus grand pour ces filtres que pour les autres méthodes. Sur les deux exemples étudiés (annexe G), les estimations sont bien moins volatiles que pour les méthodes polynomiales QL, CQ et DAF²³. En revanche les estimations intermédiaires ($q = 4$ ou $q = 5$) semblent assez erratiques. Cela pourrait provenir du fait les

23. Cela pouvait d'ailleurs s'anticiper par l'analyse des fonctions de gains présentées en annexe B. En effet, pour les autres méthodes, pour les filtres asymétriques utilisés lors de l'estimation en temps réel ($q = 0$), la fonction de gain est supérieure à 1 pour de nombreuses fréquences, ce qui signifie que le bruit est amplifié. Toutefois, pour les basses fréquences, les fonctions de gains décroissent plus vite pour les filtres issus de la méthode FST, ce qui signifie que la composante tendance-cycle est atténuée.

coefficients associés à ces moyennes mobiles donne un poids relativement faible à l'estimation courante (les coefficients associés aux observations en $t - 1$, t et $t + 1$ sont très proches).

TABLE 7 – Quartiles du nombre de mois nécessaires pour détecter correctement un point de retournement de l'année 2020 (771 observations).

		FST - LC			FTS - RKHS
	X13-ARIMA	Min.	Max.	Méd.	Min.
Min	2,0	2,0	2,0	2,0	2,0
Q1	3,0	6,0	6,0	6,0	6,0
Méd.	4,0	7,0	6,0	6,0	6,0
Q3	5,0	7,0	6,0	7,0	7,0
Max	14,0	14,0	14,0	14,0	14,0
Moy	4,4	7,0	6,5	6,6	6,7

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

En revanche, en termes de révisions entre la première et la dernière estimation (tableau 8), ces filtres asymétriques sont plus performants que les méthodes polynomiales locales et les filtres RKHS.

TABLE 8 – Quartiles de la racine carrée de l'erreur quadratique moyenne entre la première et la dernière estimation sur l'année 2020 (771 observations).

		FST - LC			FTS - RKHS
	X13-ARIMA	Min.	Max.	Méd.	Min.
Min	0,1	0,1	0,1	0,1	0,1
Q1	0,6	0,6	0,6	0,6	0,6
Méd.	1,1	1,3	1,2	1,3	1,3
Q3	2,5	2,9	2,8	2,8	3,0
Max	25,0	31,5	30,2	29,8	31,9
Moy	2,1	2,5	2,3	2,4	2,5

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

Notons par ailleurs que la hiérarchie entre les méthodes sur les révisions entre la première et la dernière observation, estimée empiriquement dans les tableaux 4 et 8, se retrouve dans les indicateurs théoriques, définis dans la section 4. C'est ce qui montré par le tableau 9, qui compare les filtres asymétriques utilisés pour l'estimation en temps-réel ($q = 0$), en supposant que les séries sont des bruits blancs.

Plusieurs conclusions peuvent être tirées de ces résultats :

- Minimiser la *timeliness* a surtout du sens lorsque peu de données dans le futur sont connues : dès lors que l'on s'approche du cas symétrique, cela ne parait plus pertinent et peut conduire à des résultats indésirables (estimations erratiques).
- Il n'est pas forcément pertinent d'utiliser la même méthode de construction pour les filtres utilisée en temps réel ($q = 0$) et pour ceux utilisés lorsque l'on s'approche du cas symétrique.
- La méthode utilisée pour mesurer le déphasage privilégie les filtres qui minimisent la révision au filtre symétrique (filtres polynomiaux locaux). En effet, pour les autres (RKHS et filtres FST) il faut, dans la majorité des cas, au moins 6 mois pour détecter correctement le poids de

TABLE 9 – Décomposition de l’erreur quadratique moyenne des filtres asymétriques utilisés en temps-réel ($q = 0$).

	LC	QL	CQ	DAF	$b_{q,\varphi}$	FST - LC			FTS - RKHS
						Min.	Max.	Méd.	Min.
Accuracy	0,4	0,4	0,4	0,4	0,4	0,4	0,5	0,4	0,4
Smoothness	0,6	2,2	3,8	4,0	0,7	0,7	0,7	0,8	0,9
Timeliness	0,5	0,6	0,6	0,6	0,5	0,5	0,5	0,5	0,5
Residual	0,7	0,5	0,5	0,5	0,6	0,5	0,5	0,5	0,5
RMSE	2,2	3,7	5,2	5,5	2,2	2,2	2,1	2,2	2,3

Note : Indicateurs calculés en supposant que la série initiale est un bruit blanc ($h(\omega) = 1$).

retournement : cela signifie qu’il y a eu une révision du point de retournement au moment de l’utilisation du filtre symétrique. Cela rejoint les deux points précédents : les moyennes mobiles asymétriques utilisées lorsque l’on s’approche du cas symétrique sont sûrement sous-optimales.

Une piste d’amélioration pourrait donc être d’étudier les filtres LC calculés en rajoutant la *timeliness* dans le programme de minimisation, comme présenté dans la section 5.2. Plus d’études devront être menées pour savoir comment calibrer le poids associé à cet indicateur.

7.4 Impact du noyau

Dans cette section, nous vérifions l’a priori que le choix du noyau n’est pas fondamental et que dans les approches polynomiales et RKHS, on peut se restreindre à l’étude d’un seul noyau. Pour cela, nous comparons, en utilisant uniquement la méthode *Linear-Constant* (LC), la performance des filtres asymétriques obtenus en utilisant les différents noyaux présentés dans la section 5.1.1.1. En termes de déphasage pour la détection de points de retournement (tableau 10), tous les noyaux ont des performances similaires, sauf pour le noyau uniforme (qui est sous-optimal car accorde un poids équivalent à toutes les observations). Les résultats sont également très proches en termes de révision (tableau 11) : le choix du noyau ne semble donc pas fondamental.

TABLE 10 – Quartiles du nombre de mois nécessaires pour détecter correctement un point de retournement de l’année 2020 (399 observations).

	Henderson	Biweight	Epanechnikov	Triangular	Tricube	Triweight	Uniform
Min	2,0	2,0	2,0	2,0	2,0	2,0	2,0
Q1	3,0	4,0	4,0	4,0	4,0	3,0	5,0
Méd.	4,0	4,0	5,0	4,0	4,0	4,0	6,0
Q3	5,0	5,0	5,0	5,0	5,0	5,0	6,0
Max	12,0	12,0	11,0	11,0	12,0	14,0	13,0
Moy	4,3	4,4	4,9	4,6	4,5	4,5	6,0

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

TABLE 11 – Quartiles de la racine carrée de l’erreur quadratique moyenne entre la première et la dernière estimation sur l’année 2020 (399 observations).

	Henderson	Biweight	Epanechnikov	Triangular	Tricube	Triweight	Uniform
Min	0,1	0,1	0,1	0,1	0,1	0,1	0,1
Q1	0,6	0,6	0,7	0,7	0,6	0,6	0,7
Méd.	1,5	1,5	1,6	1,5	1,5	1,5	1,5
Q3	3,2	3,2	3,3	3,2	3,3	3,1	3,4
Max	41,0	41,1	39,4	40,3	41,3	39,5	39,7
Moy	2,8	2,8	2,9	2,8	2,8	2,8	2,9

Note : Les statistiques ne sont calculés que pour les séries dont le filtre tendance-cycle optimal est de longueur 13.

7.5 Discussion

Comme nous l’avons vu dans les sections précédentes, la méthodologie et les indicateurs utilisés pour comparer les différentes méthodes sont perfectibles. Toutefois, les différents résultats montrent que, pour la construction des moyennes mobiles asymétriques, on peut se restreindre à celles qui conserve les constantes : chercher à imposer des contraintes polynomiales plus élevées augmente la variance des estimations et le délai pour détecter des points de retournement. Ainsi, dans les méthodes polynomiales locales, on peut restreindre les prochaines études à la méthode *Linear-Constant* (LC). Par ailleurs, l’estimation en temps-réel des méthodes utilisant les filtres asymétriques direct (DAF) peut être facilement améliorée, comme c’est par exemple le cas de la méthode de désaisonnalisation STL (*Seasonal-Trend decomposition based on Loess*) proposée par R. B. CLEVELAND et al. (1990).

Sur les méthodes ici étudiées, il serait intéressant de continuer l’exploration pour essayer de comprendre quand une méthode est plus performante qu’une autre, ce qui permettrait de donner des recommandations sur la méthode à utilisée. Il faudrait également veiller à comparer les méthodes sur d’autres types de bases de données pour vérifier si les résultats ne dépendent pas du type de série en entrée.

Toutes les méthodes de construction des filtres asymétriques sont calibrées en utilisant des paramètres estimés sur l’ensemble de la période. Par exemple, on utilise autant de points dans le passé que le filtre symétrique alors que rien ne suggère que c’est la longueur optimale. Ainsi, le ratio I/C utilisé pour choisir la longueur du filtre de la tendance-cycle et pour calibrer les méthodes locales polynomiales, est calculé sur l’ensemble de la série en excluant les 6 premiers et 6 derniers mois, qui sont justement ceux étudiés ! Pour la construction des filtres utilisés pour l’estimation en temps réel, il pourrait donc être préférable d’utiliser d’autres paramètres, par exemple des estimations locales de I/C (c’était d’ailleurs une des pistes de recherche suggérées par GRAY et THOMSON (1996)). De manière alternative, VASYECHKO et GRUN-REHOMME (2014) proposent de toujours utiliser un filtre de 13 termes : pour l’estimation en temps-réel ($q = 0$) le filtre utilise l’observation courante et les 12 observations précédentes, lorsqu’on connaît un point dans le futur ($q = 1$) on utilise les 11 observations précédant la date courante, etc.

Les indicateurs utilisés dans cette partie semblent avantager, par construction, la méthode actuelle X-13ARIMA. Cela pourrait également être le cas de la méthode d’estimation utilisée. En effet, dans les simulations, nous utilisons une série corrigée des points atypiques par un modèle Reg-ARIMA estimée sur l’ensemble de la période. Pour les simulations avec la méthode X-13ARIMA, pour chaque date, la

série est ensuite prolongée sur un an par un modèle ARIMA. Toutefois, de cette façon les prévisions sont toujours réalisées dans des cas “simples” puisqu’on ne prend pas en compte l’impact, sur la prévision, de la détection et de l’estimation des temps réel des points atypiques. Il serait donc intéressant de comparer les différentes méthodes en réestimant à chaque date la série linéarisée et donc en se plaçant réellement en estimation en temps-réel.

Enfin, dans les articles associés aux méthodes étudiées dans ce rapport, les filtres asymétriques sont appliqués et comparés sur des séries déjà désaisonnalisées. Le déphasage dans la détection des points de retournement est donc d’au plus 5 mois (lorsque le filtre symétrique est de 13 termes). Cette méthode, beaucoup simple que celle utilisée dans ce rapport, peut néanmoins conduire à deux écueils :

1. D’une part, l’estimation de la série désaisonnalisée dépend de la méthode utilisée pour extraire la tendance-cycle et comme nous l’avons vu dans l’analyse des points de retournement entre 2007 et 2008, le choix de la méthode utilisée pour l’estimation de la tendance-cycle a un impact bien au-delà de 13 mois. Les exemples devraient toutefois être approfondis pour comprendre si les points en question correspondent à de vrais points de retournement.
2. D’autre part, les moyennes mobiles étant des opérateurs linéaires, ils sont sensibles à la présence de points atypiques. L’application directe des méthodes peut donc conduire à des estimations biaisées du fait la présence de points atypiques, alors que la méthode X-13ARIMA a un module de correction des points atypiques. Par ailleurs, comme notamment montré par DAGUM (1996), le filtre symétrique final utilisé par X-13ARIMA pour extraire la tendance-cycle (et donc celui indirectement utilisé lorsque l’on applique les méthodes sur les séries désaisonnalisées) laisse passer environ 72 % des cycles de 9 ou 10 mois. Les filtres asymétriques finaux amplifient même les cycles de 9 ou 10 mois. Cela peut avoir pour conséquence l’introduction d’ondulations indésirables, c’est-à-dire la détection de faux points de retournement. Ce problème est réduit par la correction des points atypiques (ces cycles étant considérés comme de l’irrégulier). C’est ainsi qu’est né le *Nonlinear Dagum Filter* (NLDF) qui consiste à :
 - a. appliquer l’algorithme de correction des points atypique de X-13ARIMA sur la série désaisonnalisée, puis de la prolonger par un modèle ARIMA ;
 - b. effectuer une nouvelle correction des points atypiques en utilisant un seuil bien plus strict ($\pm 0,7\sigma$ et $\pm 1,0\sigma$, voir annexe F) et appliquer ensuite le filtre symétrique de 13 termes. En supposant une distribution normale cela revient à modifier 48 % des valeurs de l’irrégulier.

Une piste d’étude serait alors d’étudier plus précisément l’impact des points sur l’estimation de la tendance-cycle et la détection des points de retournement, mais aussi d’explorer de nouveaux types de filtres asymétriques basés sur des méthodes robustes (comme les régressions locales robustes, les médianes mobiles, etc.).

Conclusion

En somme, ce rapport fait une première revue de la littérature des méthodes de construction des filtres asymétriques pour l'analyse du cycle économique. Toutes ces méthodes peuvent se voir comme des cas particuliers d'une théorie générale de construction des moyennes mobiles. 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, il n'est pas nécessaire de chercher à conserver des tendances polynomiales de degré supérieur à un (filtres QL, CQ et DAF) : cela introduit de la variance dans les estimations (et donc plus de révisions) sans aucun gain en termes de détection de point de retournement. Par ailleurs, dans les méthodes polynomiales, le choix du noyau ne semble pas être fondamental.

Deuxièmement, dans l'estimation en temps réel, les méthodes d'extraction de tendance-cycle ne peuvent s'étudier indépendamment du processus de désaisonnalisation. Dans le cas contraire, on néglige notamment l'effet des points atypiques sur l'estimation de la tendance-cycle, ce qui pourrait conduire à des estimations biaisées et à une détection de faux points de retournement.

Enfin, sur les séries étudiées, la méthode utilisée dans X-13ARIMA semble satisfaisante avec des résultats proches en utilisant la méthode LC. Cependant, la moins bonne performance apparente de l'approche FST et de la méthode basée sur les RKHS semblent provenir de l'utilisation de filtres sous-optimaux lorsque que l'on s'approche du cas d'utilisation du filtre symétrique. Plus d'études devraient être faites pour savoir si, pour la construction des filtres asymétriques minimisant le déphasage, on pourrait se concentrer uniquement sur les ceux utilisés lorsque peu d'observations futures sont disponibles. Cela impliquerait notamment de revoir la méthodologie et les indicateurs utilisés.

Cette étude pourrait être étendue de plusieurs manières.

Tout d'abord, elle n'est pas exhaustive et pourrait donc être complétée. 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, 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 : dans X-13ARIMA il y a une forte correction des points atypiques, effectuée sur la composante irrégulière avant d'appliquer les filtres pour extraire la tendance. 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

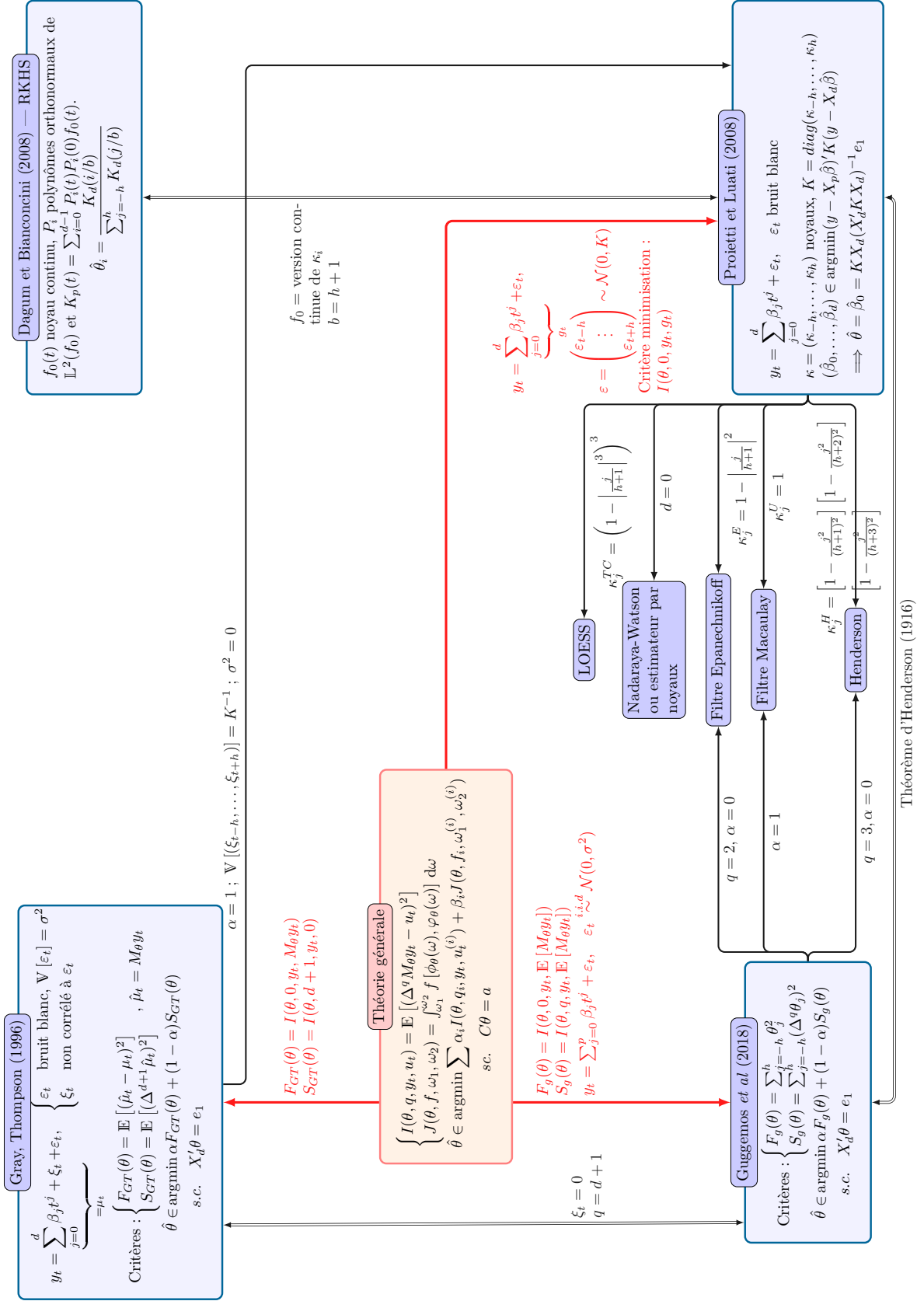


FIGURE 2 – Synthèse des méthodes de construction de moyennes mobiles symétriques $\theta = (\theta_{-h}, \dots, \theta_h)$ de $2h + 1$ termes.

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

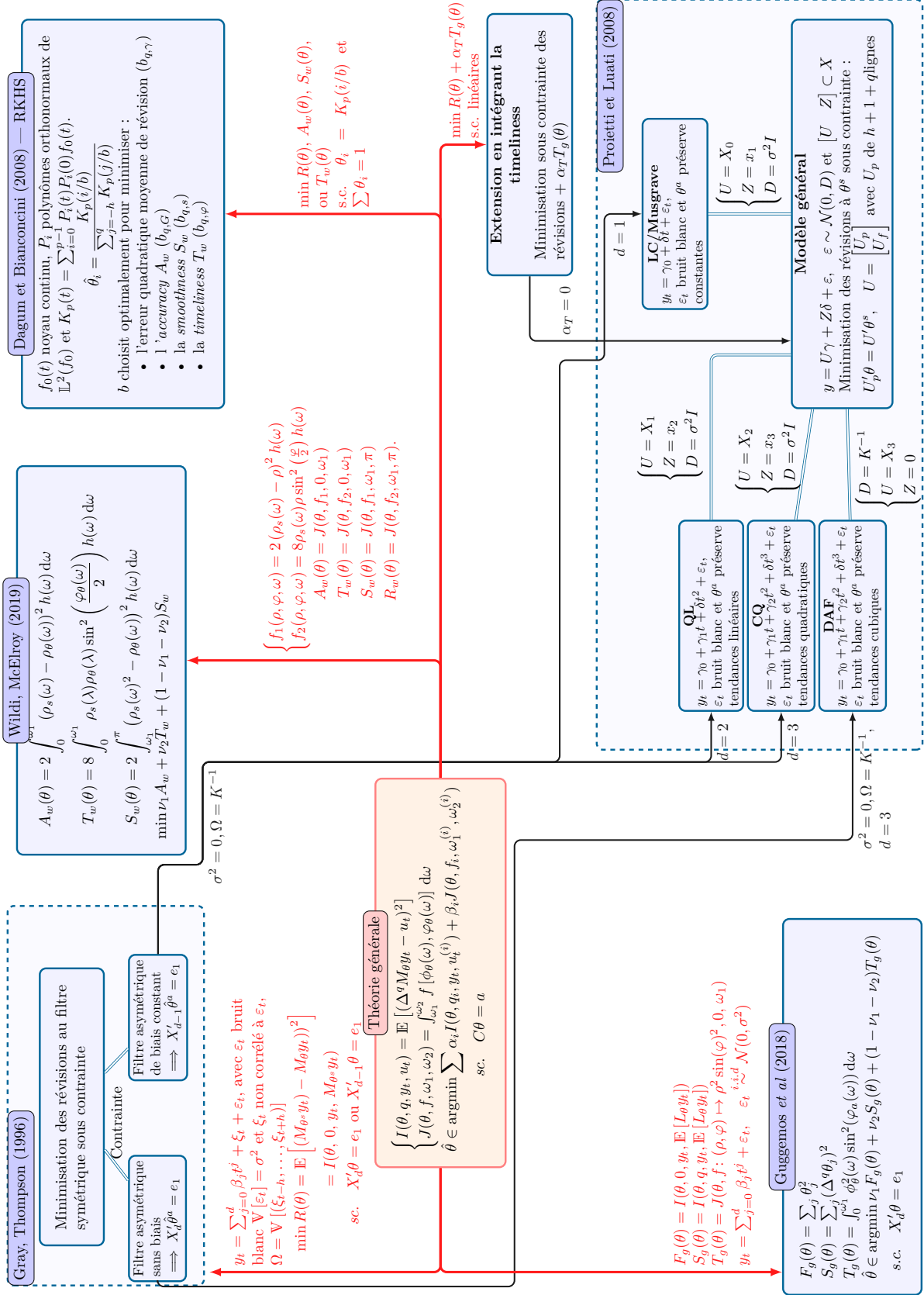


FIGURE 3 – Synthèse des méthodes de construction de moyennes mobiles asymétriques $\theta = (\theta_{-h}, \dots, \theta_q)$, $0 \leq q < h$ avec θ^s le filtre symétrique de référence de $2h+1$ termes.

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

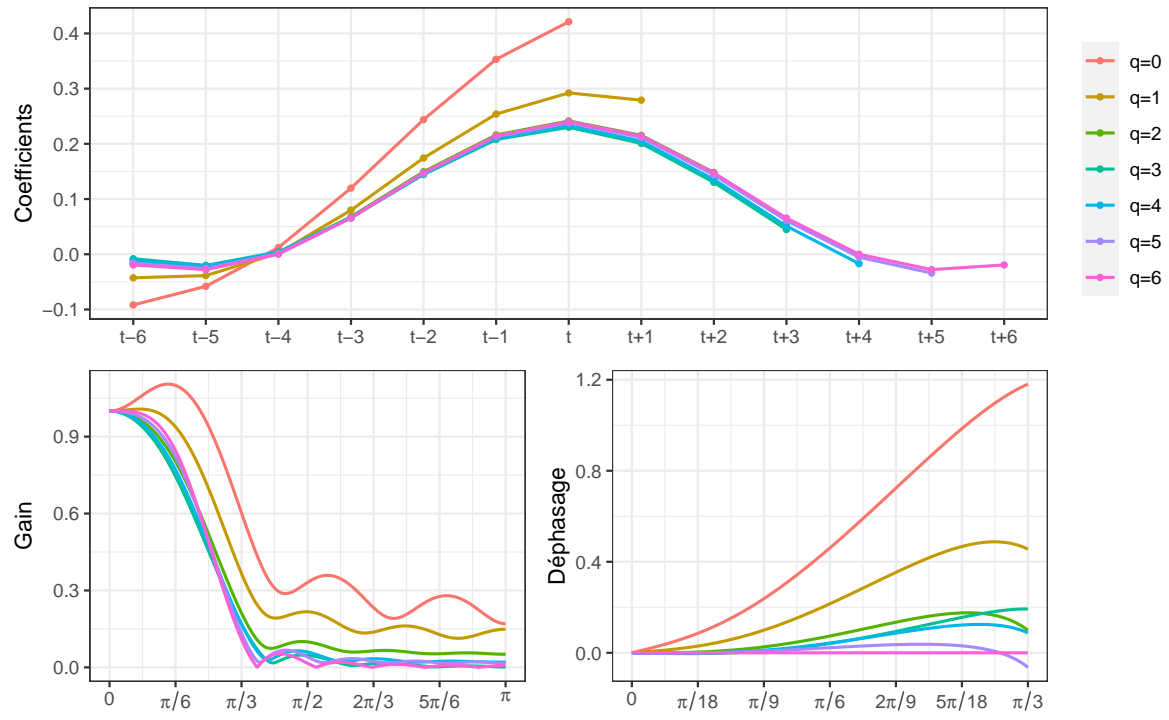


FIGURE 4 – Coefficients, fonctions de gain et de déphasage pour le filtre *Linear-Constant* (LC) avec $I/C = 3,5$.

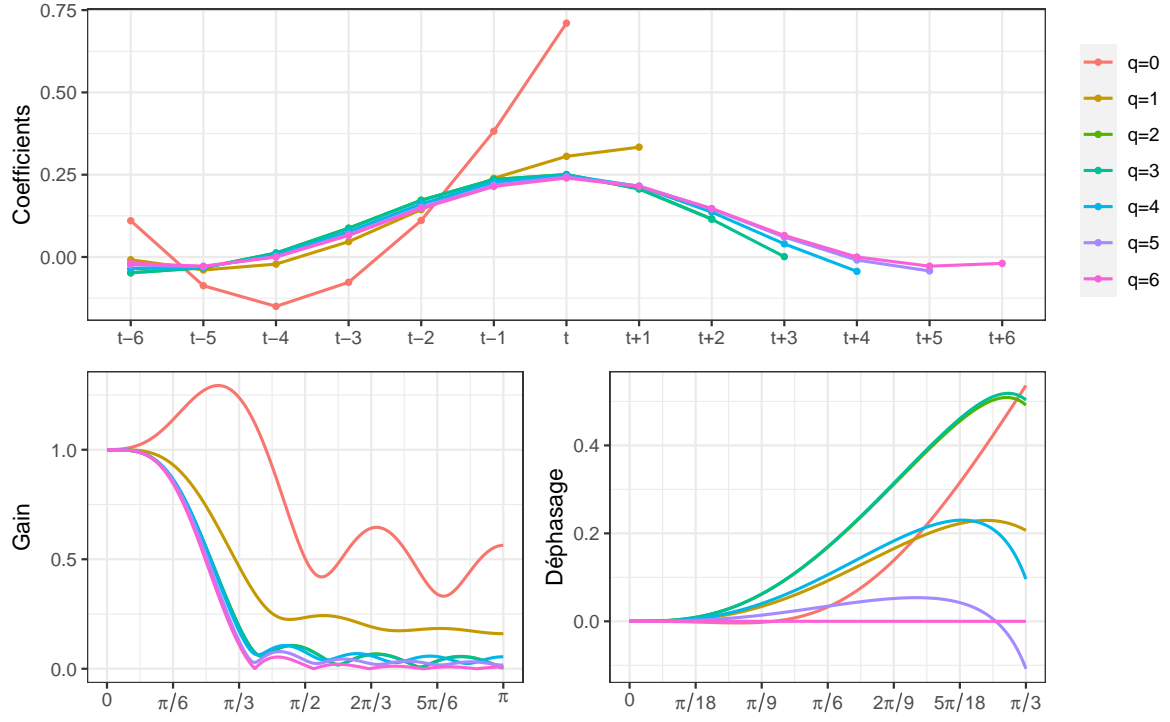


FIGURE 5 – Coefficients, fonctions de gain et de déphasage pour le filtre *Quadratic-Linear* (QL) avec $I/C = 3, 5$.

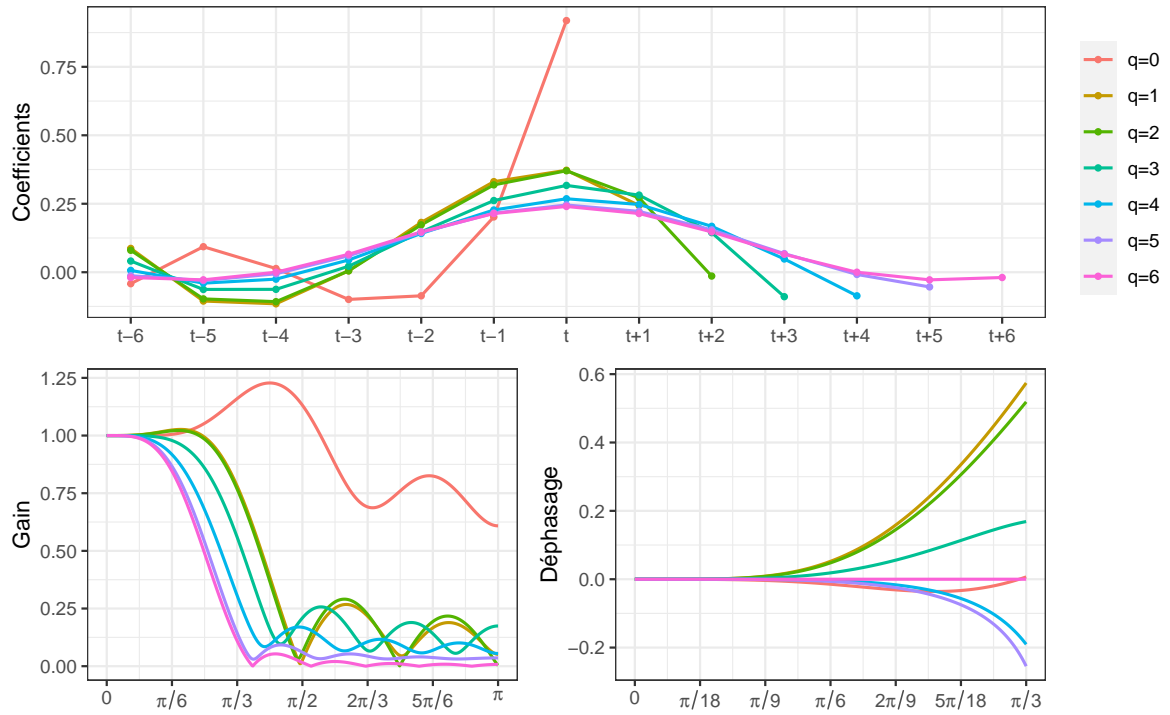


FIGURE 6 – Coefficients, fonctions de gain et de déphasage pour le filtre *Cubic-Quadratic* (QL) avec $I/C = 3, 5$.

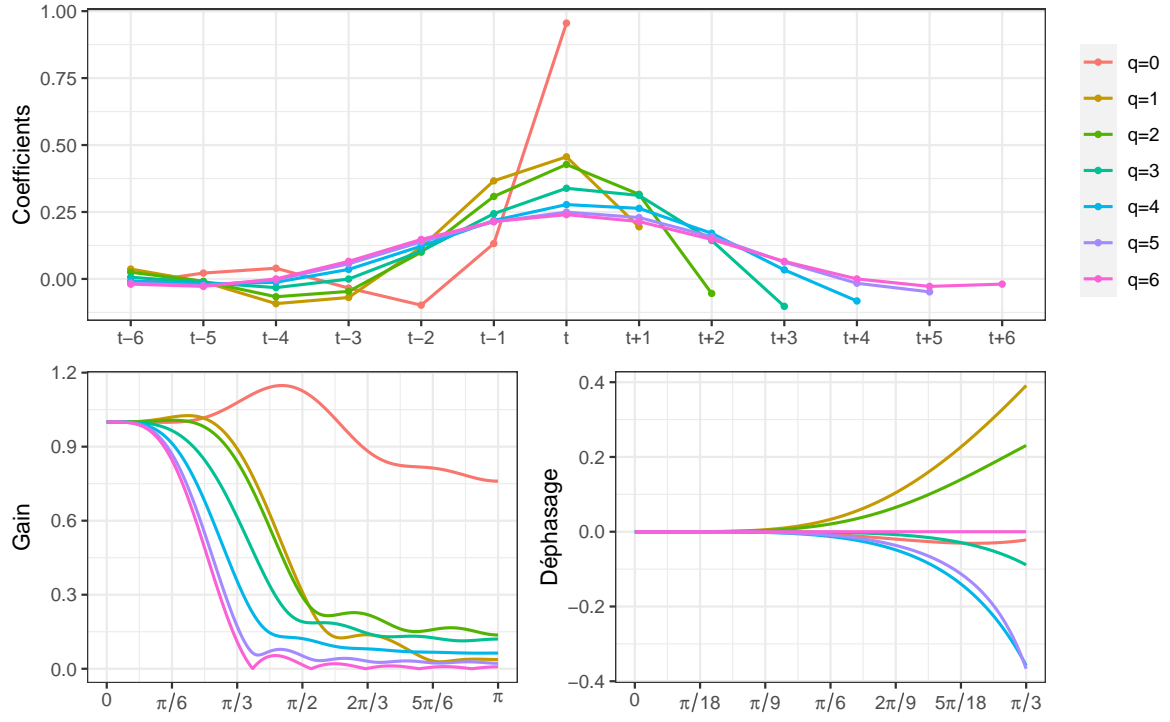


FIGURE 7 – Coefficients, fonctions de gain et de déphasage pour le filtre asymétrique direct (DAF) avec $I/C = 3, 5$.

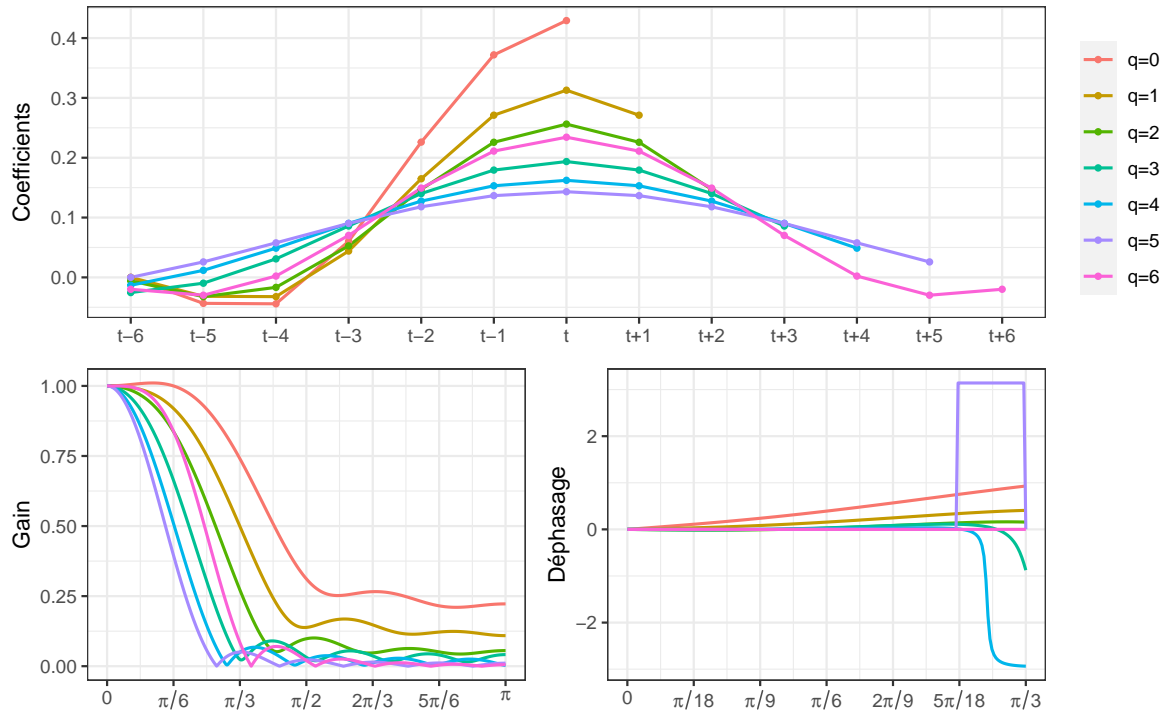


FIGURE 8 – Coefficients, fonctions de gain et de déphasage pour le filtre RKHS $b_{q,\varphi}$.

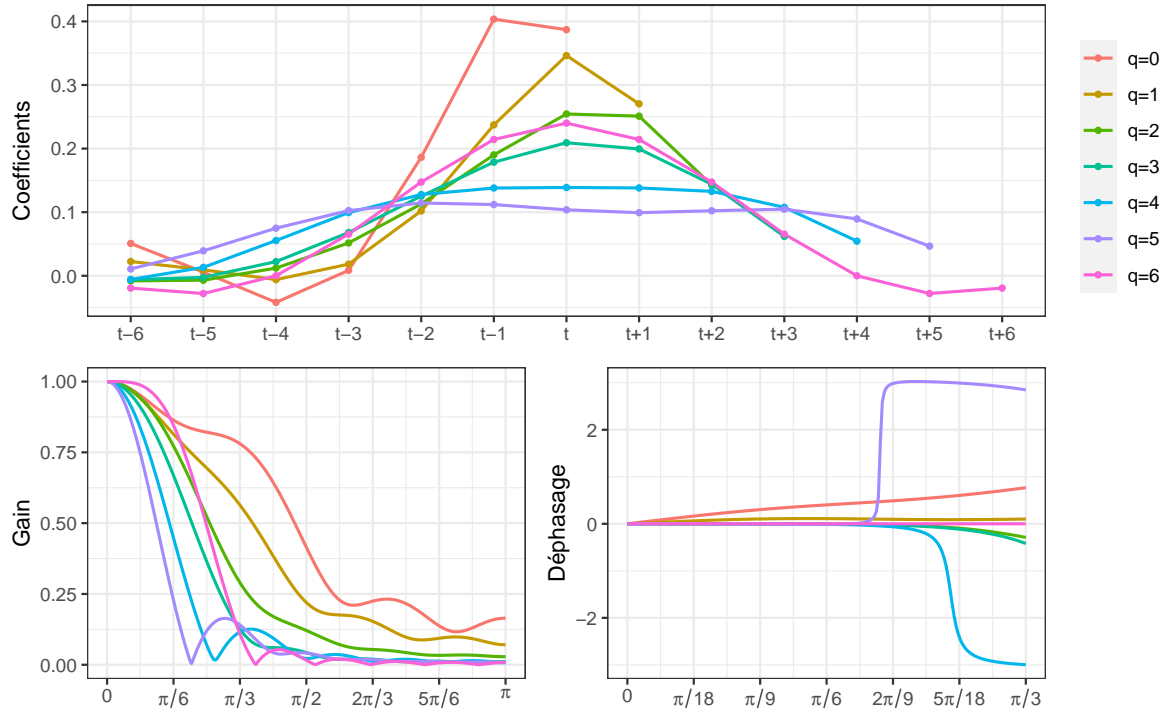


FIGURE 9 – Coefficients, fonctions de gain et de déphasage pour le filtre FST optimal par rapport au filtre LC et minimisant la *timeliness*.

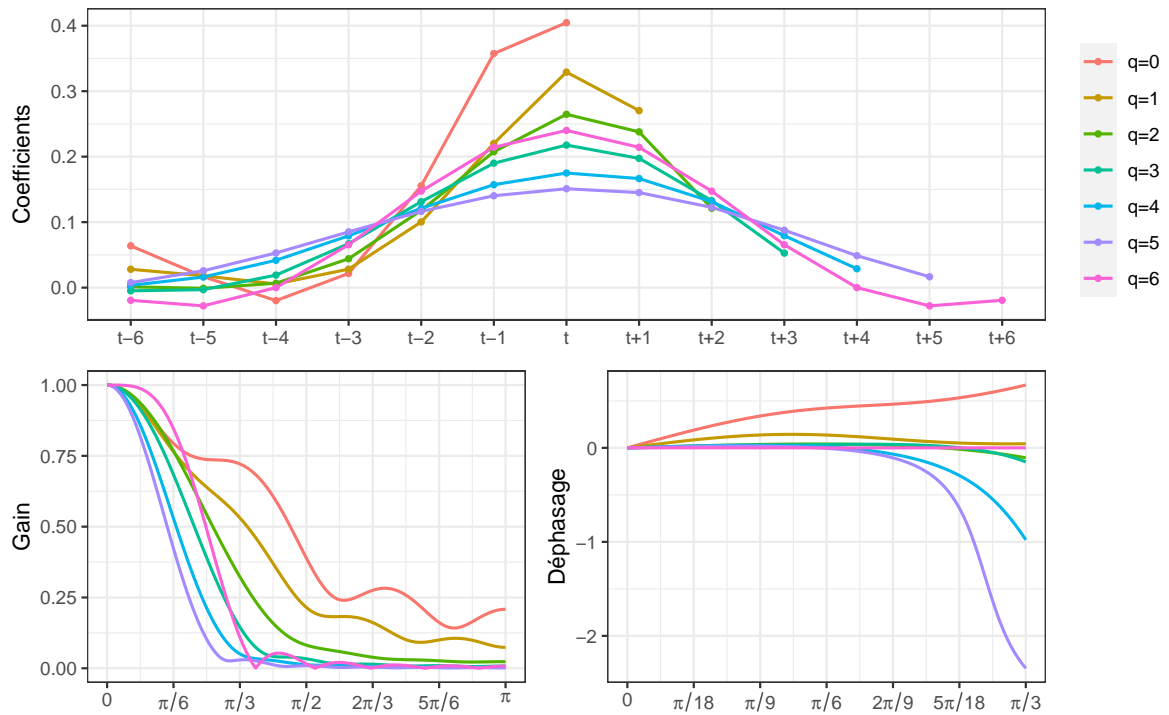


FIGURE 10 – Coefficients, fonctions de gain et de déphasage pour le filtre FST optimal par rapport au filtre LC et ayant la *timeliness* la plus élevée.

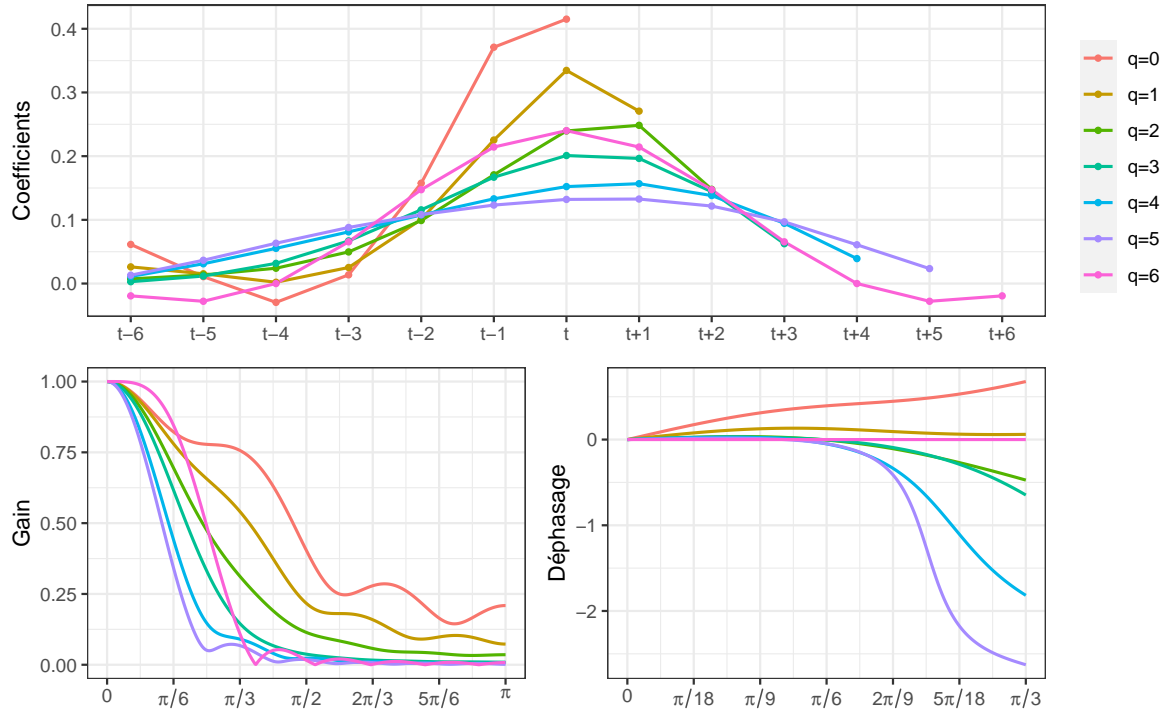


FIGURE 11 – Coefficients, fonctions de gain et de déphasage pour le filtre FST optimal par rapport au filtre LC et ayant une *timeliness* médiane.

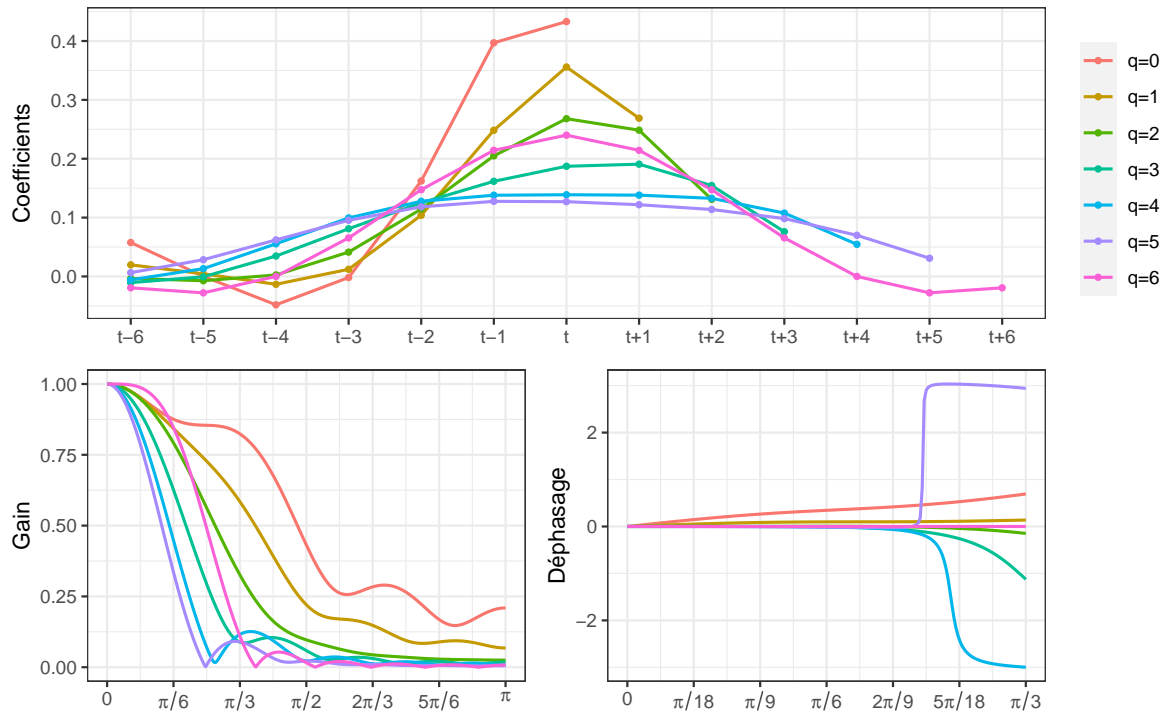


FIGURE 12 – Coefficients, fonctions de gain et de déphasage pour le filtre FST optimal par rapport au filtre RKHS $b_{q,\varphi}$ et minimisant la *timeliness*.

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) et $h = 11$ (filtre symétrique de 23 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 13) 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 testés²⁴.

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.

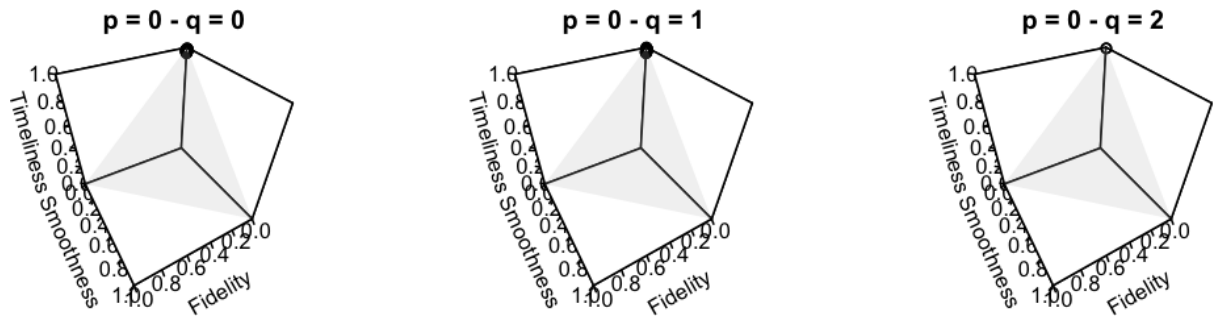


FIGURE 13 – Ensemble des poids pour lesquels la méthode FST n'est pas équivalente aux moindres carrés pondérés pour $h = 6$ (filtre symétrique de 13 termes), sous contrainte de préservation des polynômes de degré au plus 2 ($d = 0, 1, 2$).

24. 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.

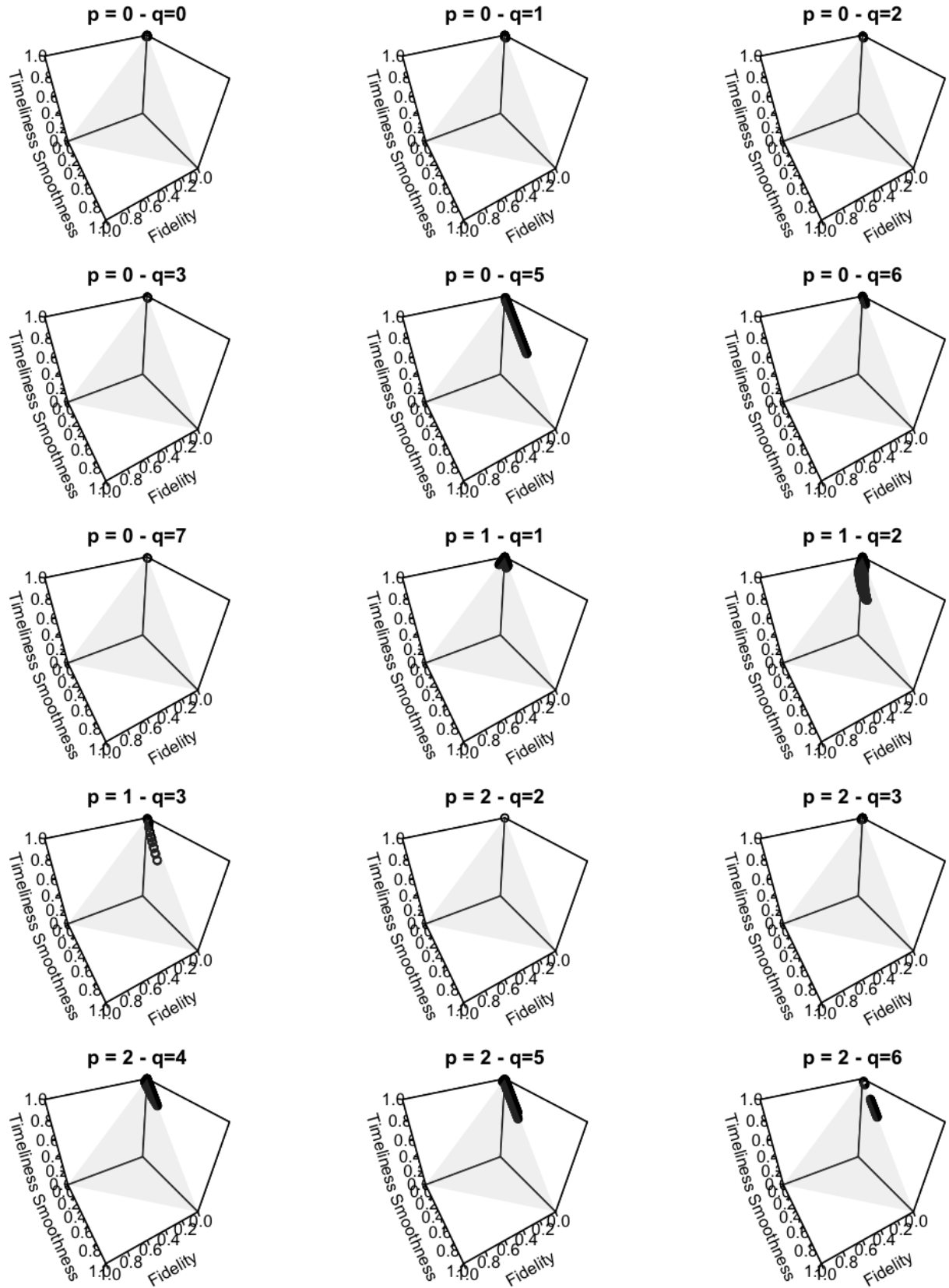


FIGURE 14 – Ensemble des poids pour lesquels la méthode FST n'est pas équivalente aux moindres carrés pondérés pour $h = 11$ (filtre symétrique de 23 termes), sous contrainte de préservation des polynômes de degré au plus 2 ($d = 0, 1, 2$).

D Critères de minimisation dans les filtres RKHS

Cette annexe montre la fonction objectif que l'on cherche à minimiser dans la construction des filtres asymétriques par la méthode des RKHS. Utiliser, dans le calcul des critères, la densité spectrale d'une marche aléatoire a peu d'influence sur la valeur des critères.

D.1 Filtre symétrique de 9 termes ($h = 4$)

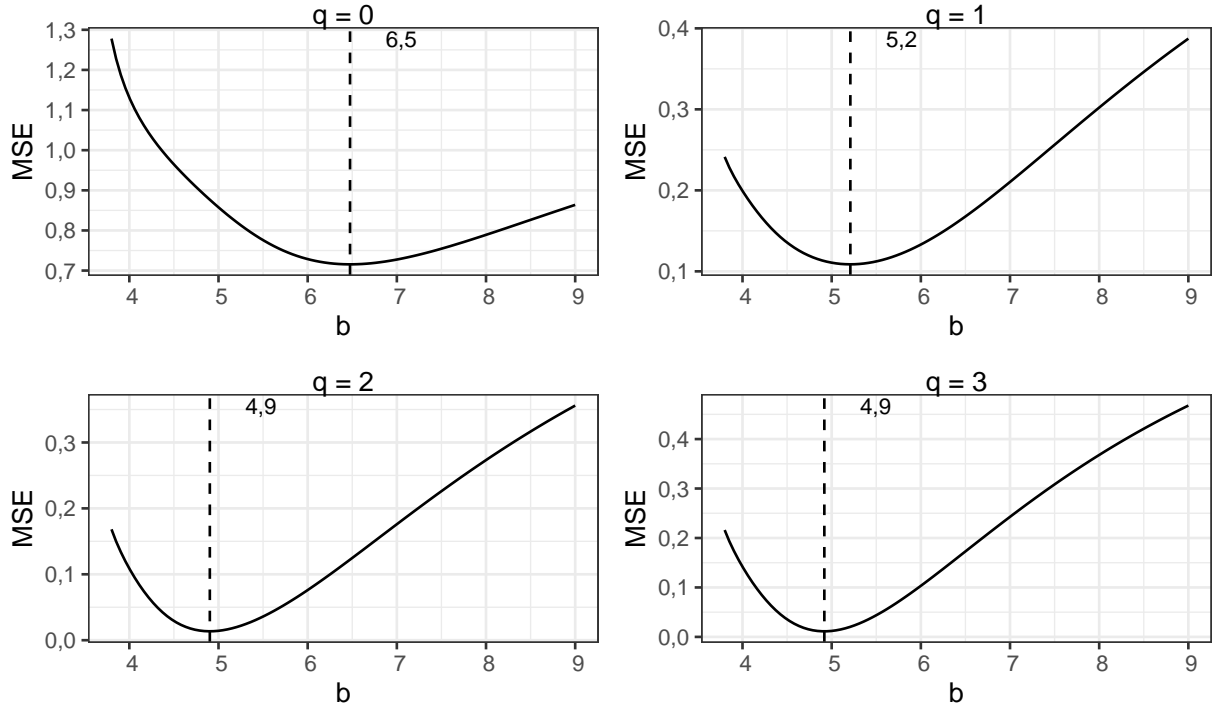


FIGURE 15 – Erreur quadratique moyenne pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 9 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans `rjdfilters`.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

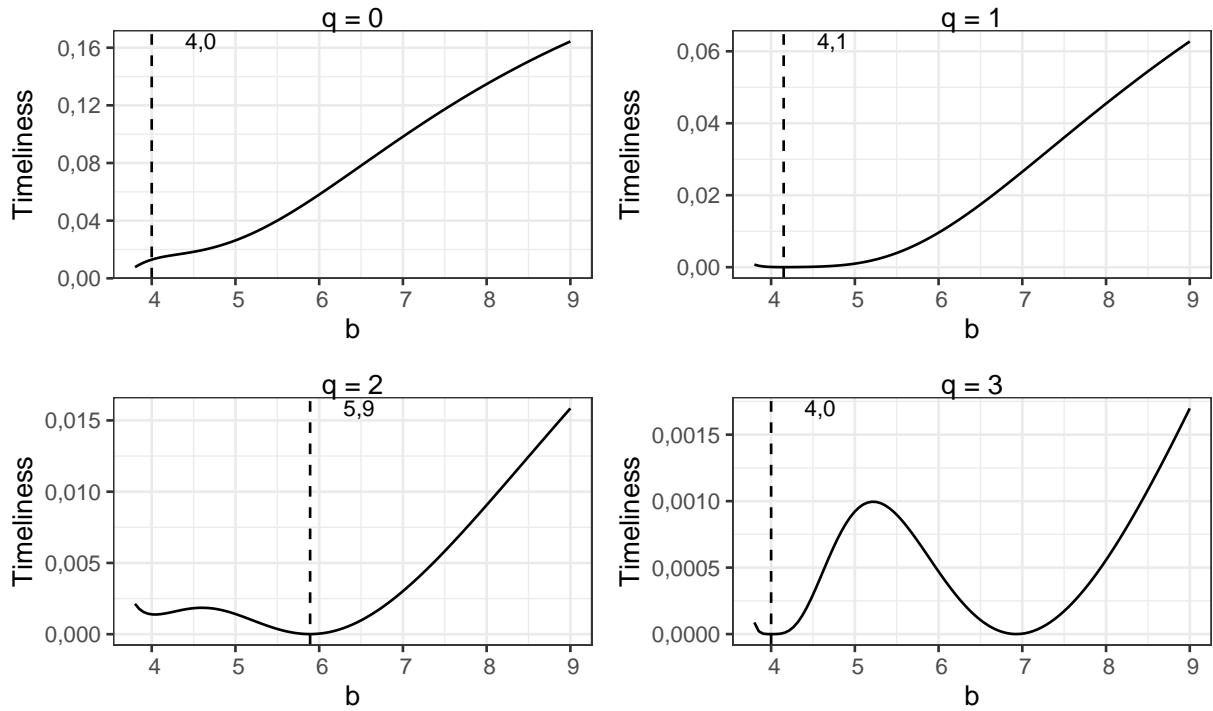


FIGURE 16 – *Timeliness* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 9 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans *rjdfilters*.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

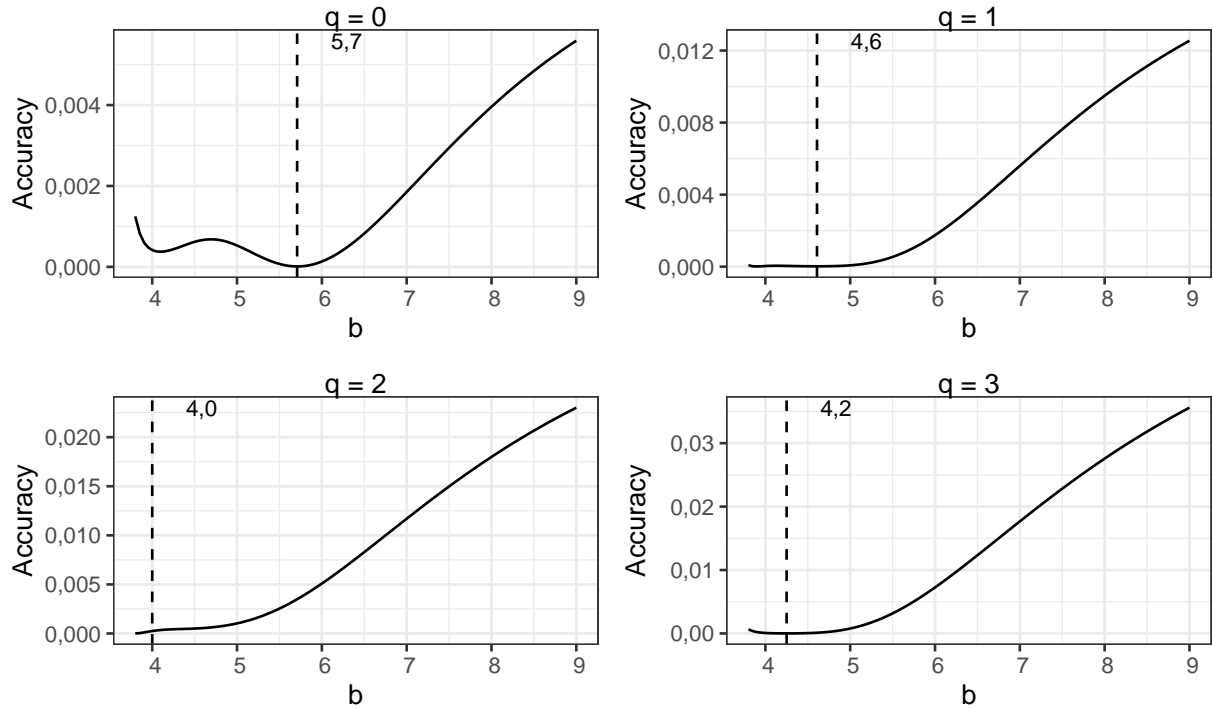


FIGURE 17 – *Accuracy* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 9 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans *rjdfilters*.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

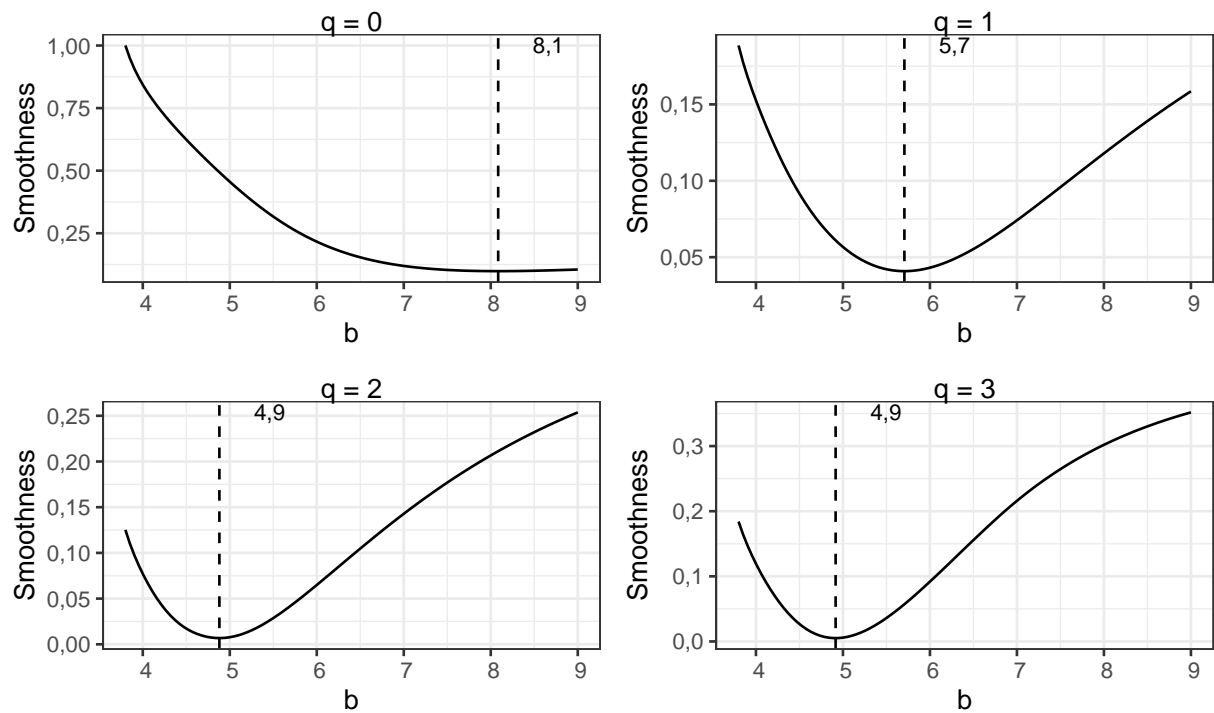


FIGURE 18 – *Smoothness* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 9 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans `rjdfilters`.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

D.2 Filtre symétrique de 13 termes ($h = 6$)

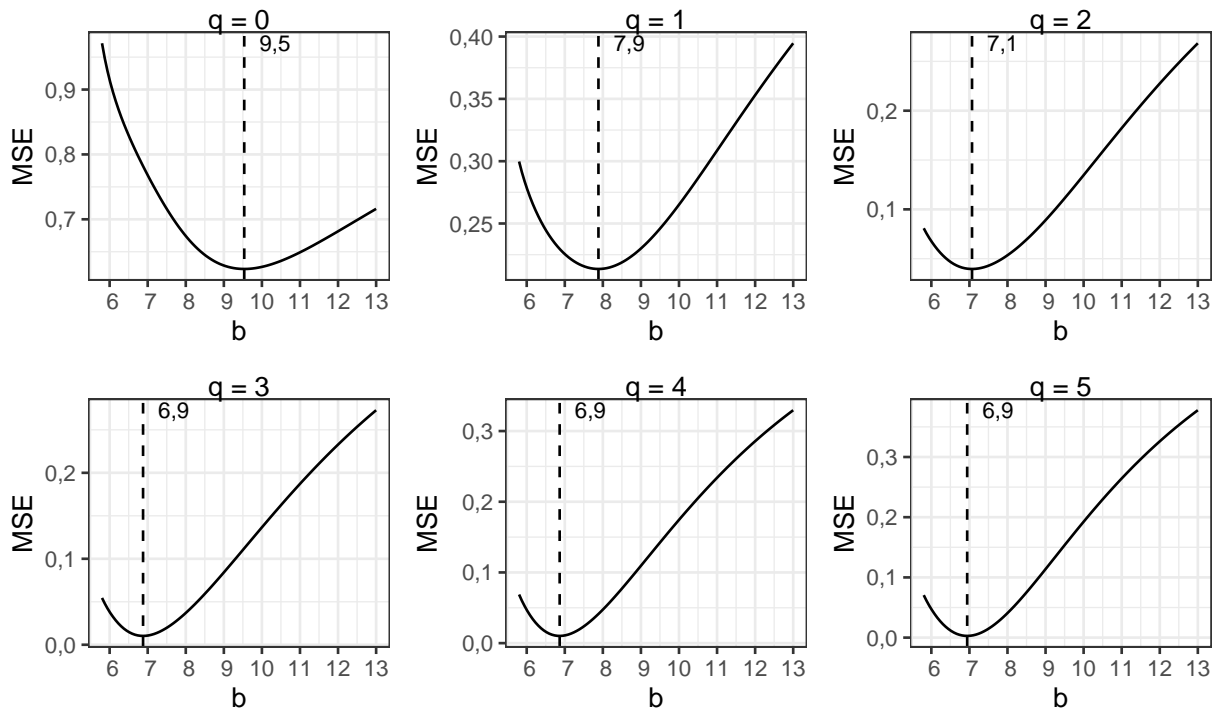


FIGURE 19 – Erreur quadratique moyenne pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 13 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans `rjdfilters`.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

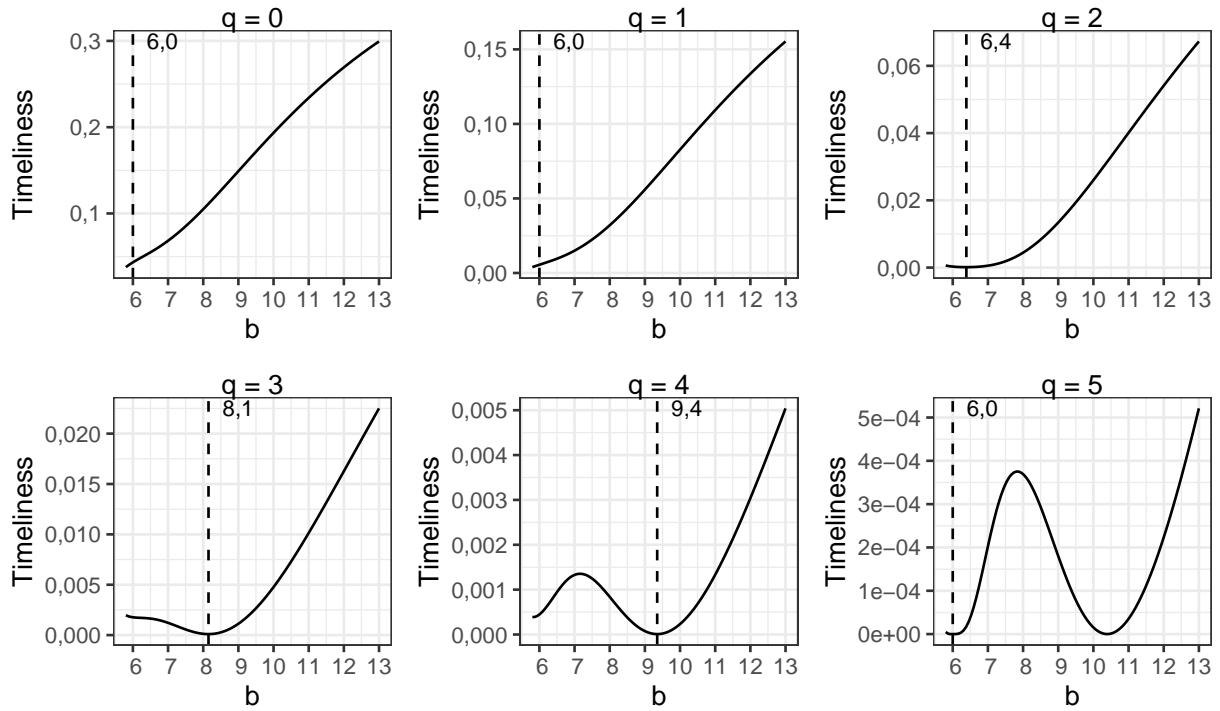


FIGURE 20 – *Timeliness* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 13 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans *rjdfilters*.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

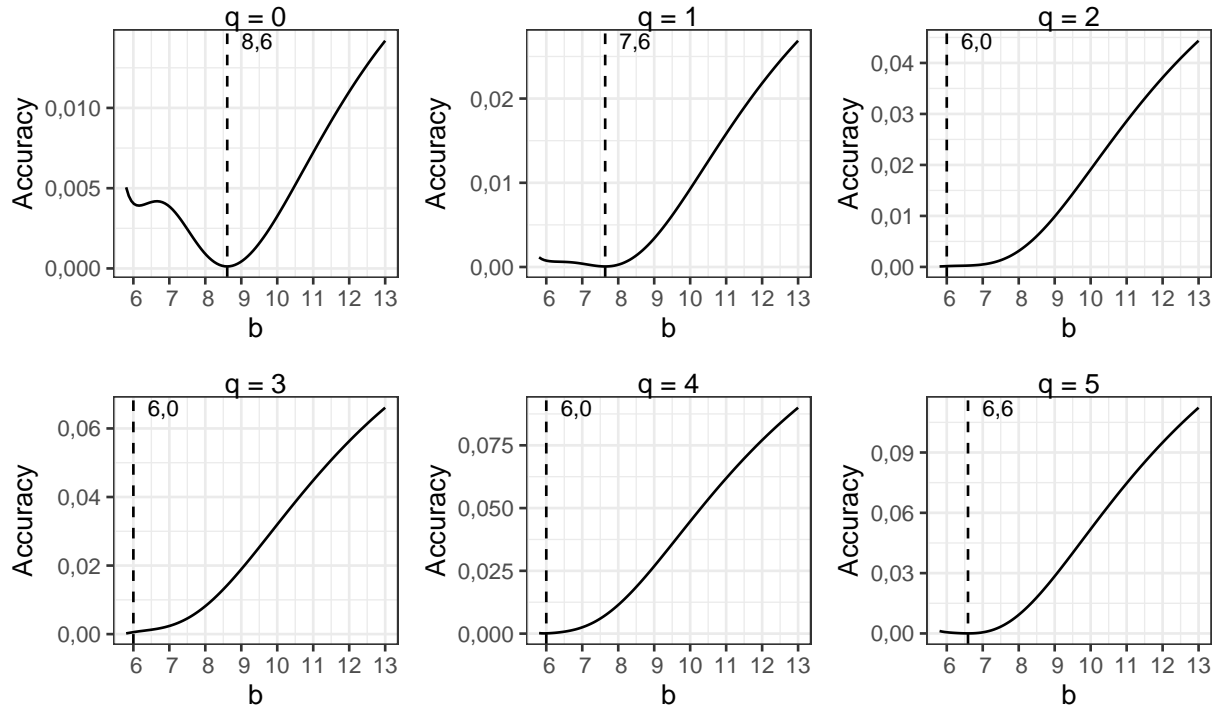


FIGURE 21 – *Accuracy* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 13 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans *rjdfilters*.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

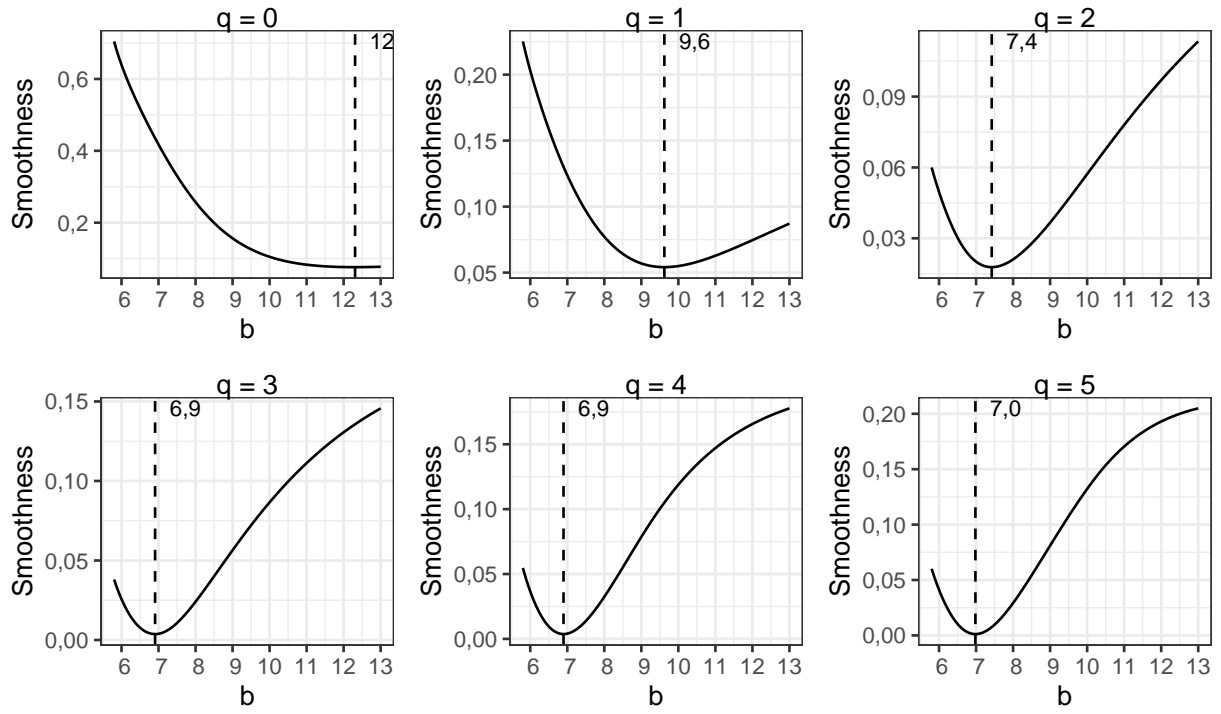


FIGURE 22 – *Smoothness* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 13 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans *rjdfilters*.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

D.3 Filtre symétrique de 23 termes ($h = 11$)

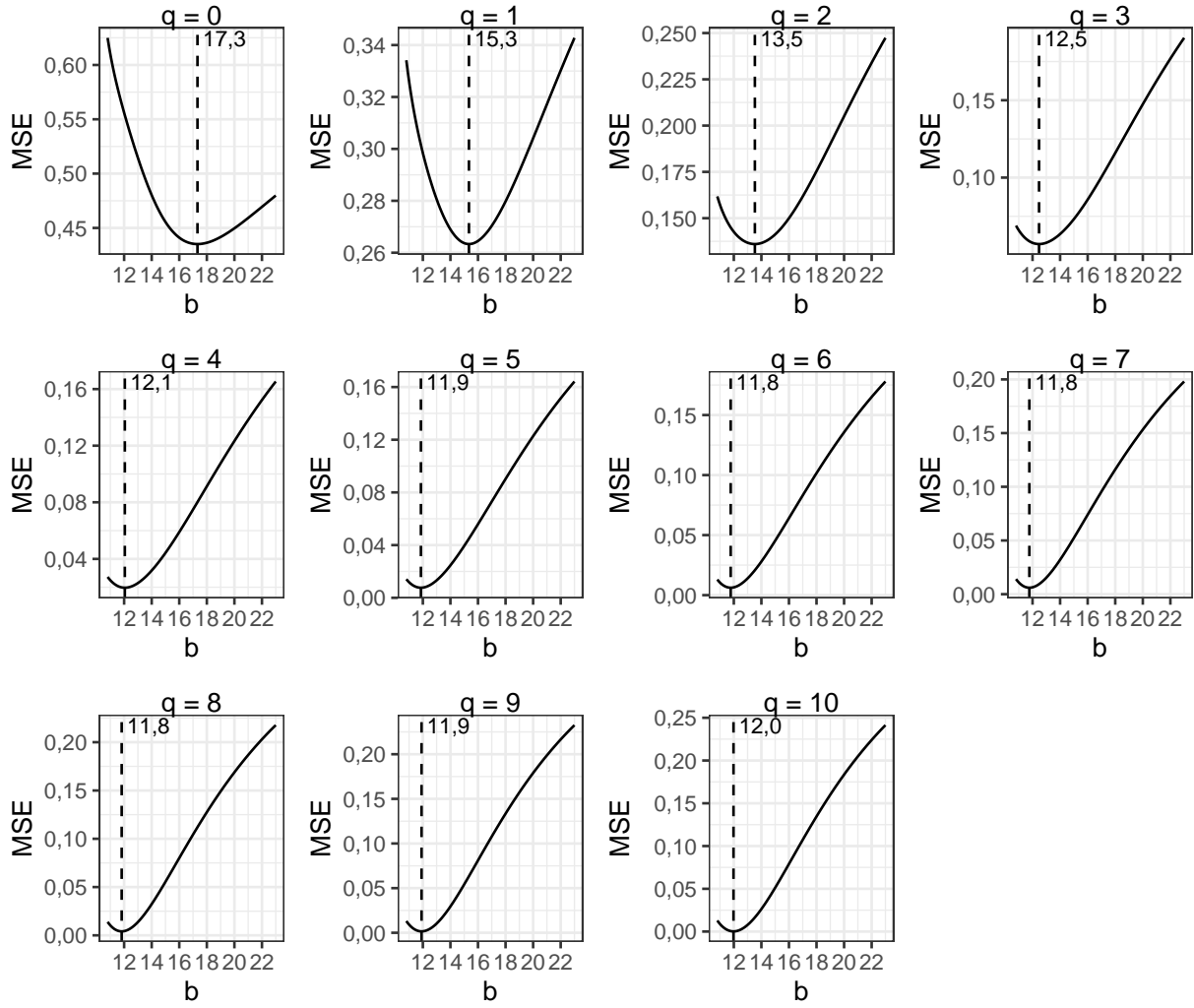


FIGURE 23 – Erreur quadratique moyenne pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 23 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans `rjdfilters`.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

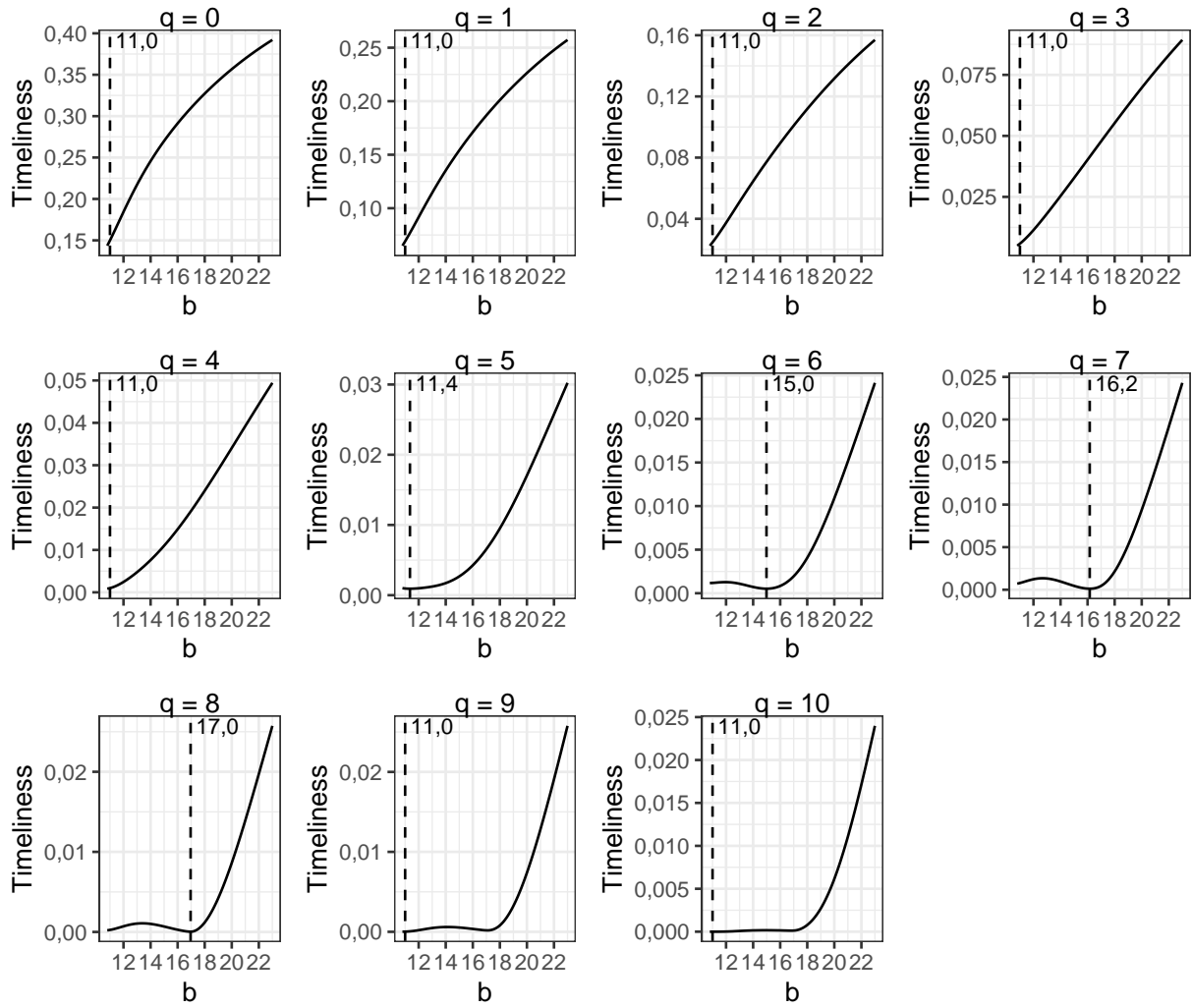


FIGURE 24 – *Timeliness* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 23 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans `rjdfilters`.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

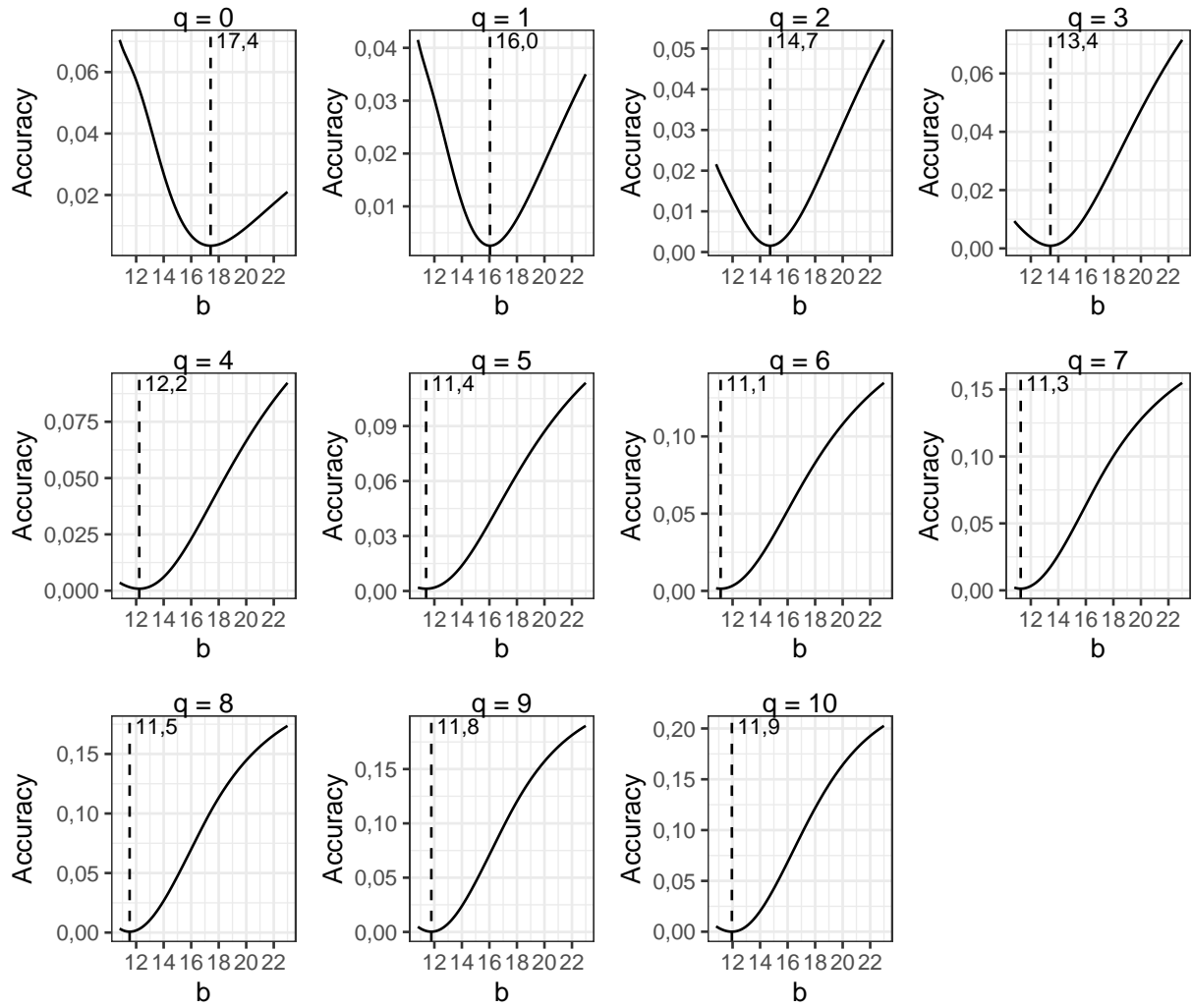


FIGURE 25 – Accuracy pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 23 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans `rjdfilters`.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

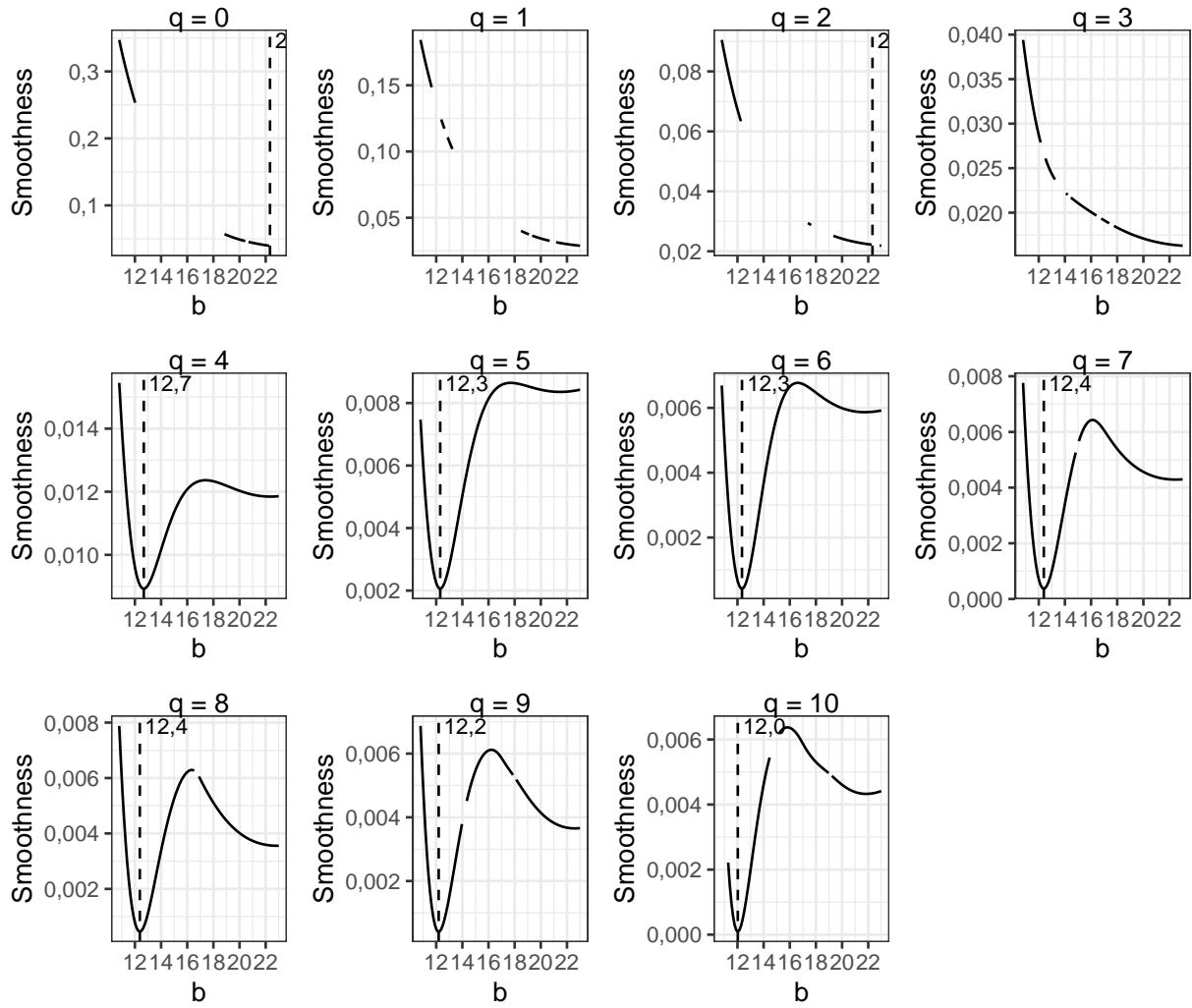


FIGURE 26 – *Smoothness* pour les filtres RKHS avec $h_{WN}(x) = 1$ (série initiale est un bruit blanc) en fonction du paramètre b pour un filtre symétrique de 23 termes.

Notes : Le trait horizontal correspond à la valeur obtenue par défaut dans *rjdfilters*.

Dans la définition des indicateurs $\omega_1 = 2\pi/12$.

E Comparaison avec le filtre FST

Cette annexe synthétise les cas où il est possible de trouver un ensemble de poids tels que les filtres FST ont de meilleures propriétés (en termes de *fidelity*, *smoothness* et *timeliness*) que les filtres polynomiaux ou RKHS, avec les mêmes contraintes sur la préservation des polynômes (ou des contraintes supérieures). Par simplicité, les graphiques ne sont tracés que pour les principales méthodes étudiées : le filtre *Linear-Constant* (LC) et le filtre RKHS $b_{q,\varphi}$ obtenu par minimisation de la *timeliness*. Pour les filtres LC, utiliser un I-C ratio à 3,5 ou 4,5 n'a quasiment aucun impact : pour synthétiser ces valeurs, ne sont tracés que les ensembles de poids qui donnent des filtres optimaux pour ces deux valeurs.

E.1 Filtres polynomiaux locaux

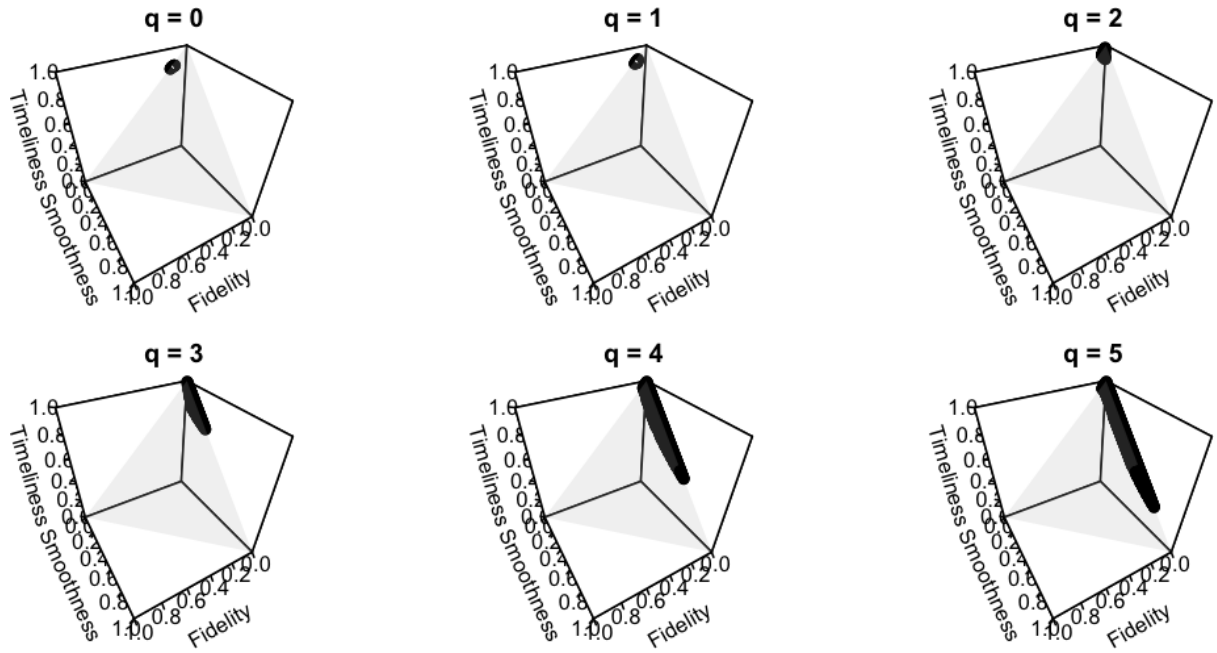


FIGURE 27 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre *Linear-Constant* (LC) sous contrainte de préservation des constantes ($d = 0$) et pour $h = 6$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

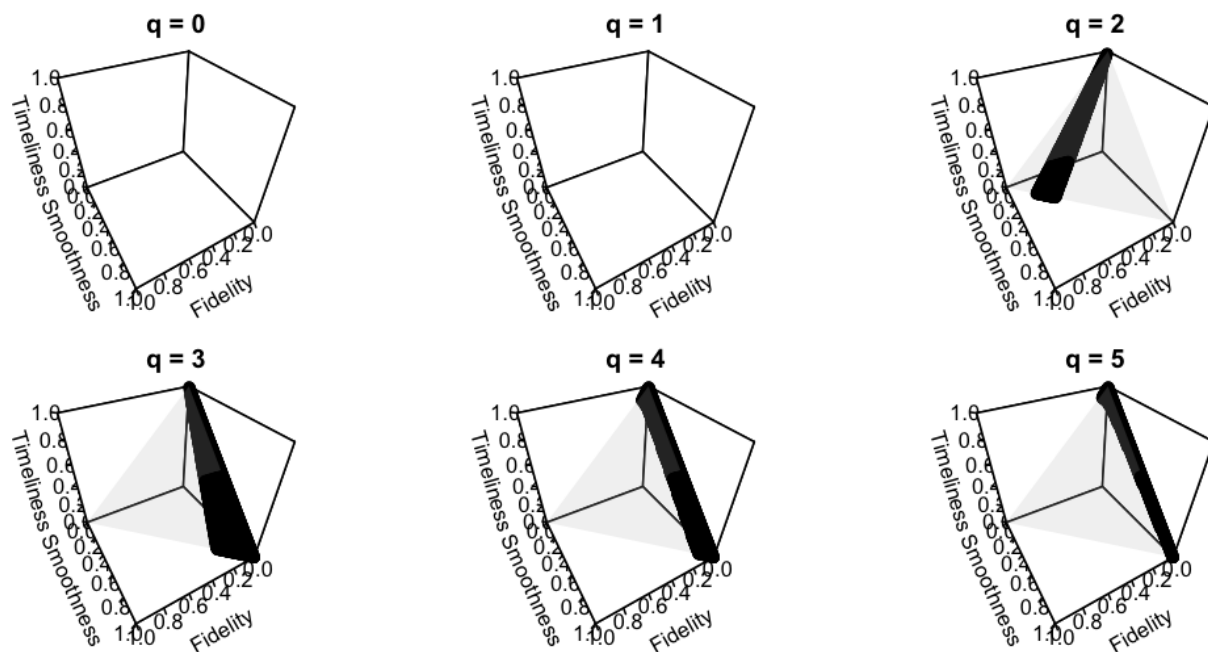


FIGURE 28 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre *Linear-Constant* (LC) sous contrainte de préservation des droites ($d = 1$) et pour $h = 6$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

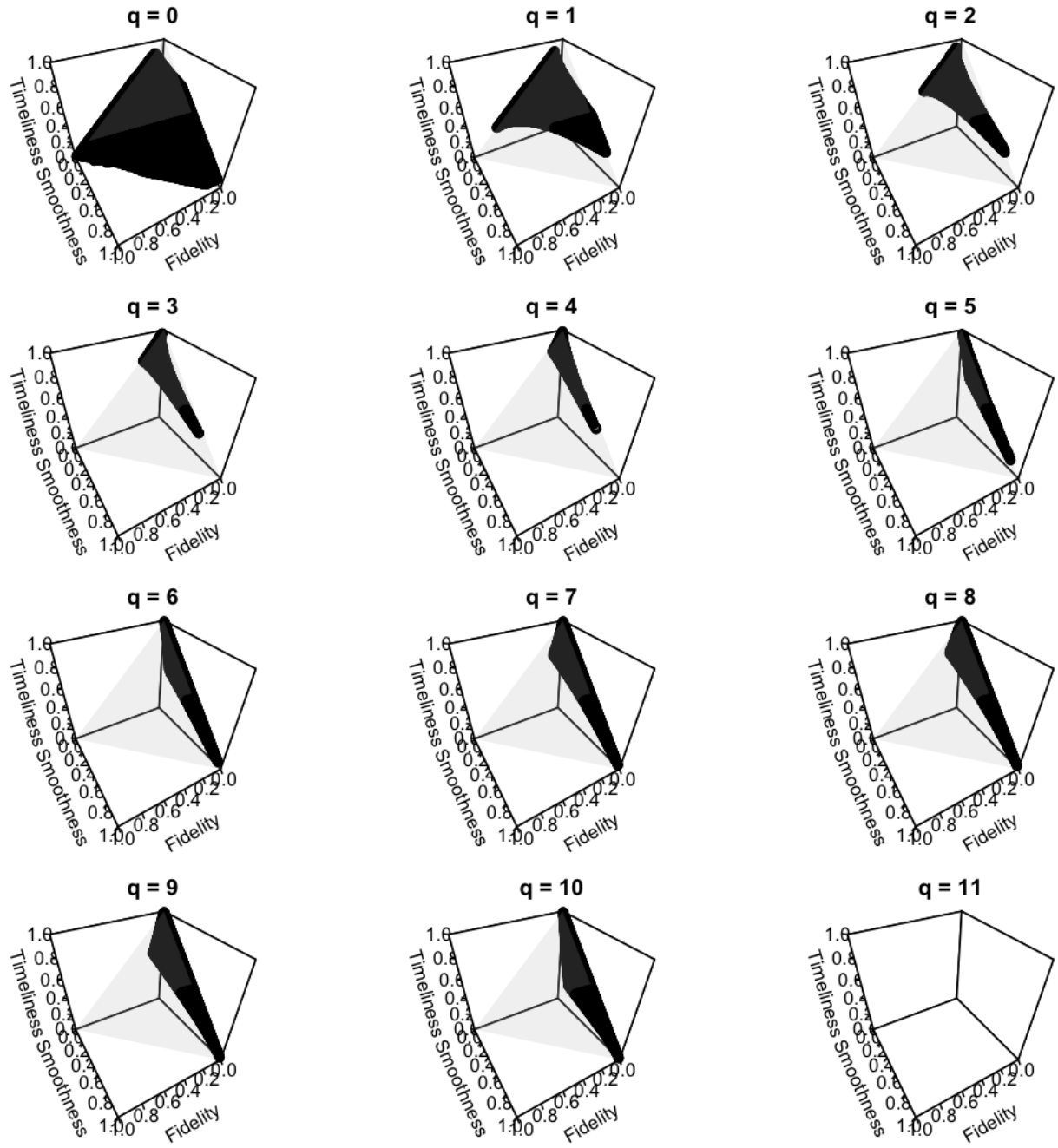


FIGURE 29 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre *Linear-Constant* (LC) sous contrainte de préservation des constantes ($d = 0$) et pour $h = 11$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

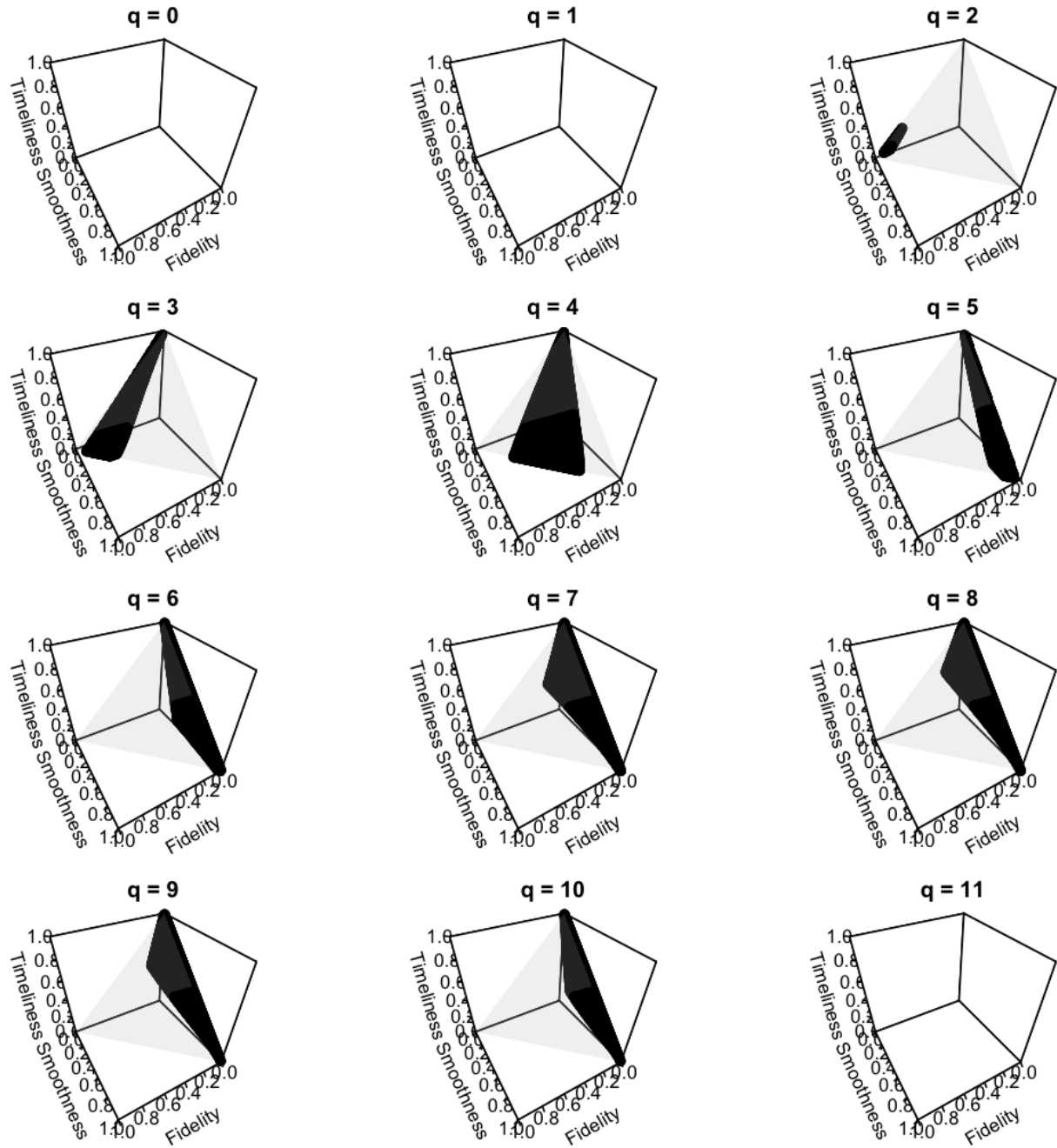


FIGURE 30 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre *Linear-Constant* (LC) sous contrainte de préservation des droites ($d = 1$) et pour $h = 11$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

E.2 Filtres RKHS

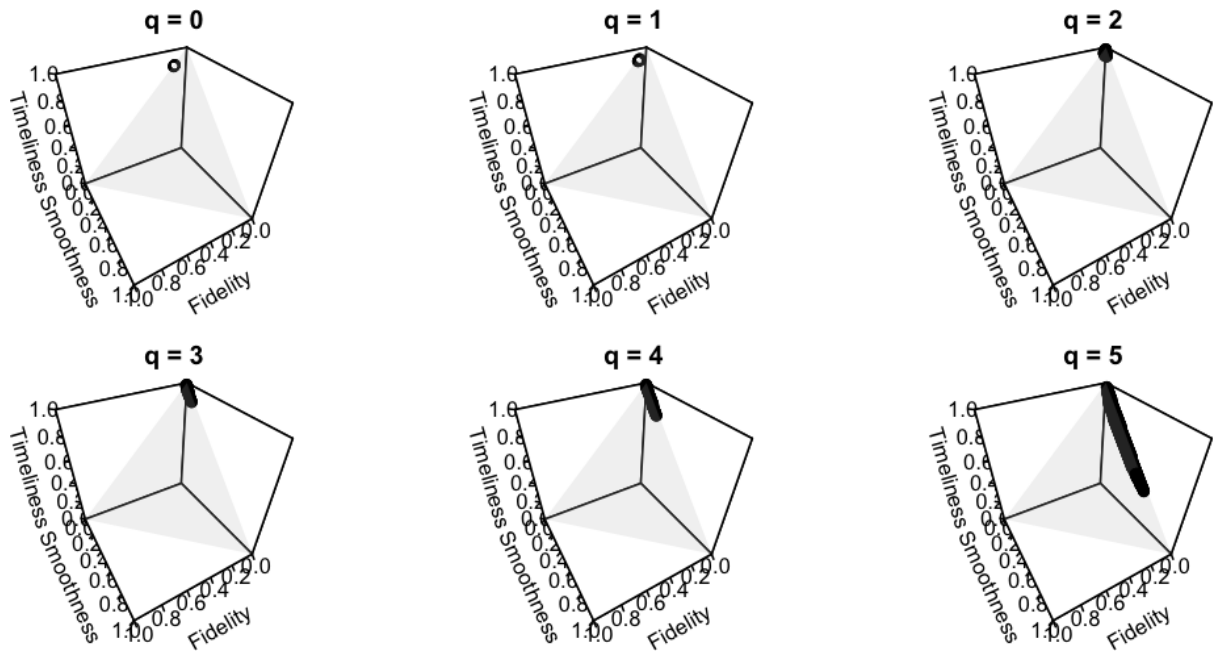


FIGURE 31 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre $b_{q,\varphi}$ sous contrainte de préservation des constantes ($d = 0$) et pour $h = 6$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

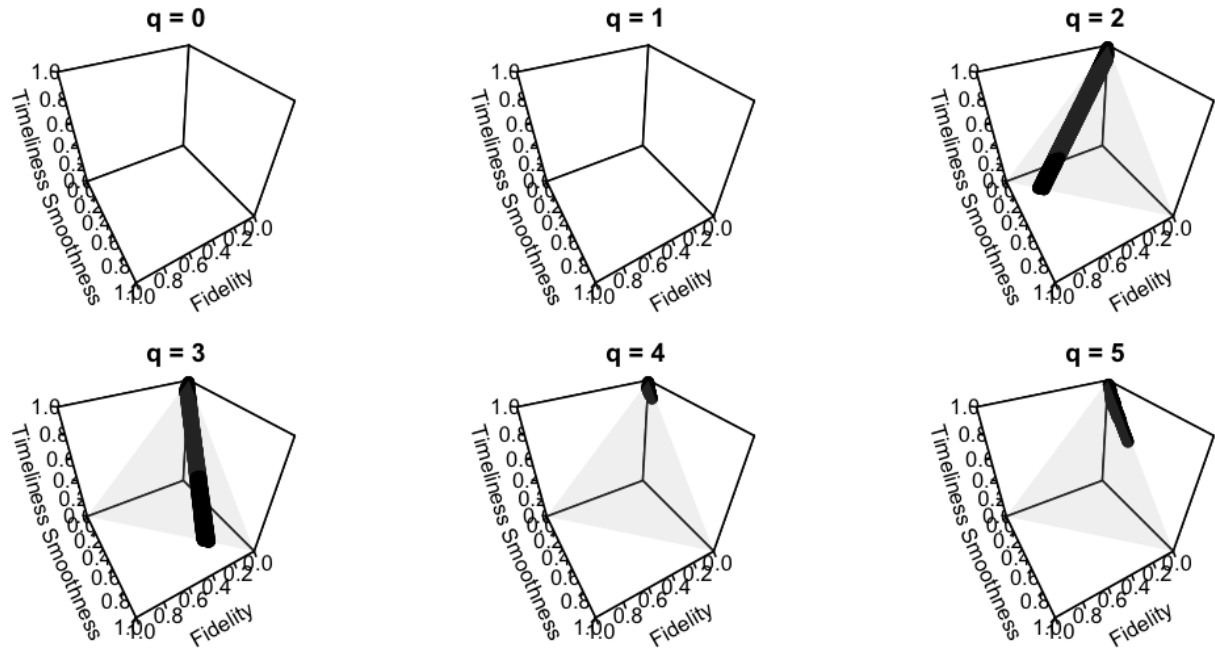


FIGURE 32 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre $b_{q,\varphi}$ sous contrainte de préservation des droites ($d = 1$) et pour $h = 6$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

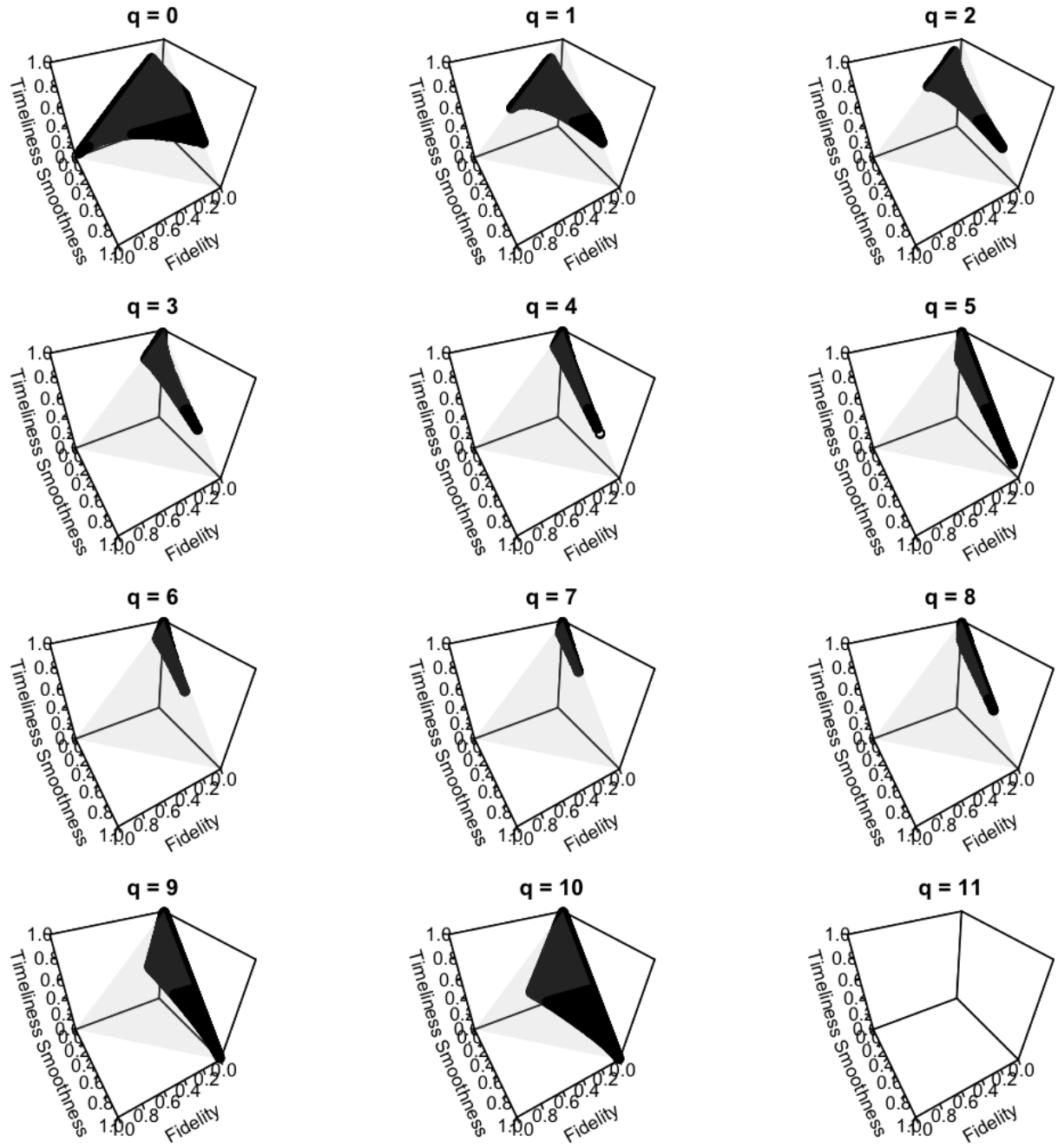


FIGURE 33 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre $b_{q,\varphi}$ sous contrainte de préservation des constantes ($d = 0$) et pour $h = 11$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

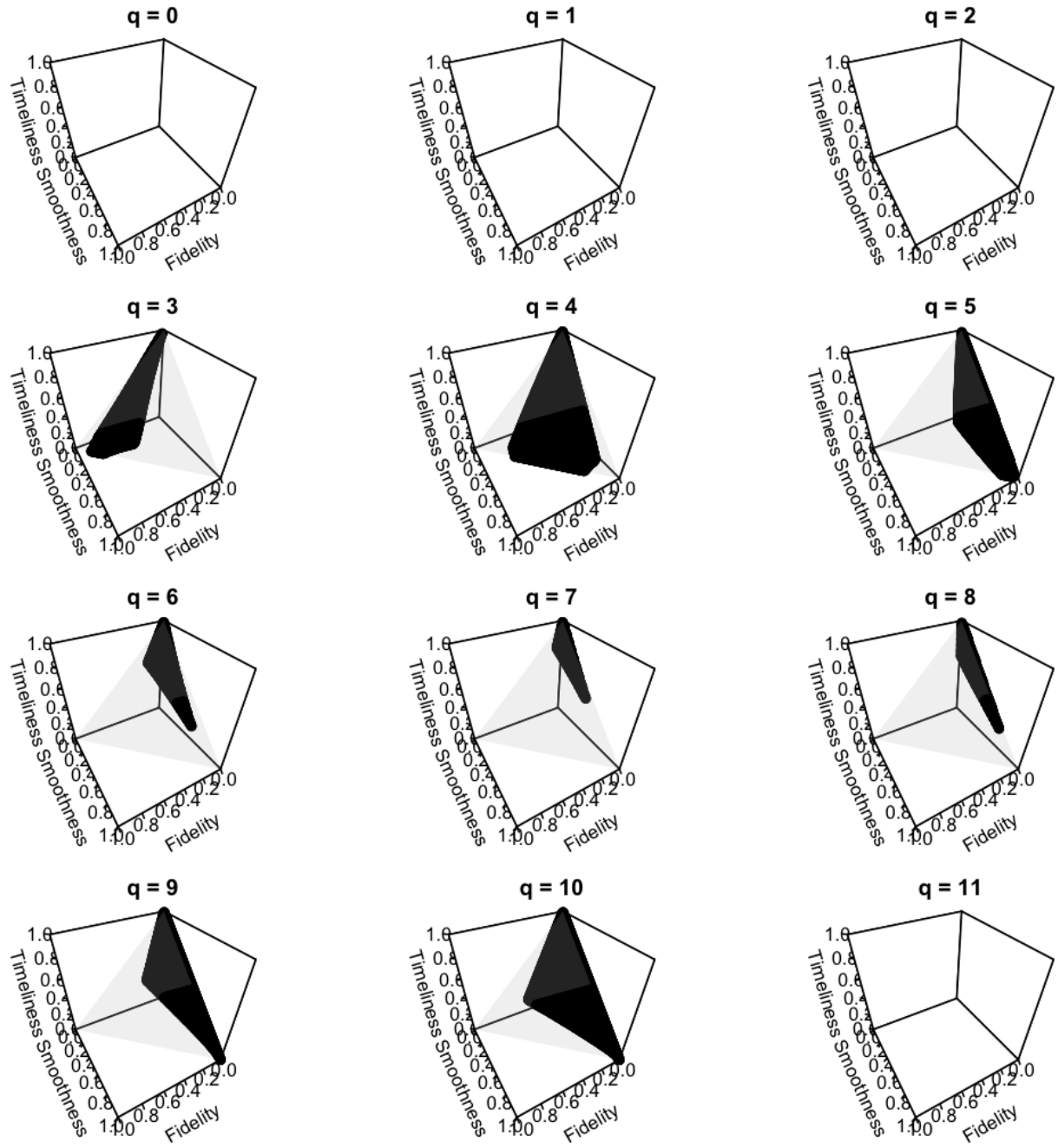


FIGURE 34 – Ensemble des poids pour lesquels la méthode FST donne des filtres de meilleure qualité que le filtre $b_{q,\varphi}$ sous contrainte de préservation des droites ($d = 1$) et pour $h = 11$.

Notes : La zone grisée correspond à l'ensemble $\left\{ (x \ y \ z)' : x + y + z = 1 \right\}$.

Le graphique est vide lorsque aucun poids ne permet de trouver de filtre FST qui minimise la fidelity, smoothness et la timeliness par rapport au filtre étudié.

F Décomposition avec X-13ARIMA

Pour décomposer une série, l'algorithme X-13-ARIMA (et TRAMO-SEATS) fonctionnent en deux grandes étapes :

1. Dans la première étape, la série est corrigée des jours ouvrables et des points atypiques à l'aide d'un modèle ARIMA. La série initiale est également prolongée par prévision sur la dernière année pour la seconde étape. La série corrigée et prolongée est appelée série *linéarisée*.
2. Dans une seconde étape, cette série linéarisée, notée X_t est décomposée en trois composantes saisonnalité S_t , tendance-cycle TC_t et irrégulier I_t .

Dans ce rapport nous nous intéressons essentiellement à cette seconde étape, où la méthode utilisée pour extraire la composante tendance-cycle en fin de période sera modifiée. La méthode de décomposition utilisée dans X-13ARIMA, similaire à celle de X-11, est détaillée de manière très précise dans Dominique LADIRAY et Benoît QUENNEVILLE (1999). Cette annexe reprend de manière synthétique la description faite dans cet article.

Pour la décomposition l'algorithme estime itérativement les différentes composantes en appliquant des moyennes mobiles. Cette algorithme peut se décomposer en l'application de deux blocs successifs. Pour une série mensuelle et dans le cadre d'une décomposition dite additive ($X_t = S_t + TC_t + I_t$)²⁵ :

1. Estimation de la tendance-cycle par une moyenne mobile $M_{2 \times 12}$. C'est une moyenne mobile symétrique qui conserve les tendances linéaires et supprime la saisonnalité (elle est calculée en composant les deux moyennes mobiles simples $1/12(B^6 + B^5 + \dots + B^{-5})$ et $1/12(B^5 + B^4 + \dots + B^{-6})$). À cette étape aucune moyenne mobile asymétrique n'est utilisée, les 6 premiers points et les 6 derniers points ne sont donc pas calculés.

$$TC_t^{(1)} = M_{2 \times 12}(X_t)$$

2. Estimation de la composante saisonnier-irrégulier :

$$(S_t + I_t)^{(1)} = X_t - TC_t^{(1)}$$

3. Estimation de la composante saisonnière par moyenne mobile $M_{3 \times 3}$ (composition de deux moyennes mobiles symétriques simples de 3 termes $1/3(B + I + B^{-1})$) appliquée chaque mois de manière indépendante. En fin de période, des moyennes mobiles asymétriques spécifiques sont utilisées :

$$S_t^{(1)} = M_{3 \times 3} \left[(S_t + I_t)^{(1)} \right]$$

Puis une normalisation est faite entre les mois :

$$Snorm_t^{(1)} = S_t^{(1)} - M_{2 \times 12} \left(S_t^{(1)} \right)$$

25. D'autres schémas de décomposition peuvent être considérés, par exemple une décomposition multiplicative : $X_t = S_t \times TC_t \times I_t$.

En fin de série, la moyenne mobile $M_{2 \times 12} \left(S_t^{(1)} \right)$ ne pouvant être calculée sur les 6 derniers mois, la dernière valeur calculée de $M_{2 \times 12} \left(S_t^{(1)} \right)$ est prolongée sur les 6 derniers mois (et inversement pour le début de période). Les coefficients saisonniers manquants (du fait de la première estimation de la tendance-cycle à l'étape 1) sont imputés en utilisant la valeur la plus proche du mois considéré.

4. Estimation de la série corrigée des variations saisonnières :

$$Xsa_t^{(1)} = (TC_t + I_t)^{(1)} = X_t - Snorm_t^{(1)}$$

Ensuite, les mêmes étapes sont répétées mais en faisant l'estimation de la tendance-cycle sur la série désaisonnalisée. Cela permet notamment d'utiliser des moyennes mobiles ayant de meilleures propriétés en termes de préservation polynomiale (celles notamment présentées dans ce rapport) mais qui ne supprime pas la saisonnalité.

5. Estimation de la tendance-cycle par moyenne de Henderson. L'ordre de la moyenne mobile est déterminée de façon automatique, à partir de l'I-C ratio, entre une moyenne centrée de 9 termes ($h = 4$), 13 termes ($h = 6$, c'est ce qui est retenu dans la majorité des cas) et 23 termes ($h = 11$). En fin de période, des moyennes mobiles asymétriques de Musgrave sont utilisées. C'est à cette étape que dans nos différentes simulations nous modifions les filtres asymétriques utilisés en ne prenant pas en compte les prévisions de la série linéarisée.

$$TC_t^{(2)} = H_{13}(Xsa_t^{(1)})$$

6. Estimation de la composante saisonnier-irrégulier :

$$(S_t + I_t)^{(2)} = X_t - TC_t^{(2)}$$

7. Estimation de la composante saisonnière par moyenne mobile dont l'ordre est détectée automatiquement entre une moyenne mobile $M_{3 \times 3}$, $M_{3 \times 5}$ (ce qui est généralement retenu) et $M_{3 \times 7}$. Sur chaque mois :

$$S_t^{(2)} = M_{3 \times 5} \left[(S_t + I_t)^{(2)} \right]$$

Puis une normalisation est faite (avec même traitement en fin de période qu'à l'étape 3) :

$$Snorm_t^{(2)} = S_t^{(2)} - M_{2 \times 12} \left(S_t^{(2)} \right)$$

8. Estimation de la série corrigée des variations saisonnières :

$$Xsa_t^{(2)} = (TC_t + I_t)^{(2)} = X_t - Snorm_t^{(2)}$$

L'étape 5 est répétée une dernière fois sur cette série désaisonnalisée afin d'avoir l'estimation finale de la tendance-cycle.

Les moyennes mobiles étant très sensibles à la présence de points atypiques, une correction automatique de l'irrégulier est effectuée, à partir de $S_t + I_t$, avant les étapes 3 et 7 :

- Une première estimation de S_t et donc de I_t (par soustraction) est effectuée de la même façon qu'aux étapes 3 et 7.
- Ensuite, pour chaque année, un écart-type mobile de l'irrégulier, noté σ , est calculé sur 5 ans, l'année étudiée étant l'année centrale (pour les deux dernières années, l'écart-type de l'antépénultième année est considérée).
- On affecte, à chaque valeur de I_t , un poids p_t en fonction de son écart à l'écart-type associé (voir graphique 35) :
 - si $|I_t| \leq 1,5\sigma$ un poids de 1 est affecté à la date t (i.e. : le point n'est pas considéré comme atypique) ;
 - si $|I_t| \geq 2,5\sigma$, un poids de 0 est affecté à la date t ;
 - si $1,5\sigma < |I_t| < 2,5\sigma$, un poids variant linéairement entre 0 et 1 est affecté à la date t .

Les seuils de 1,5 et de 2,5 peuvent être modifiés par l'utilisateur.

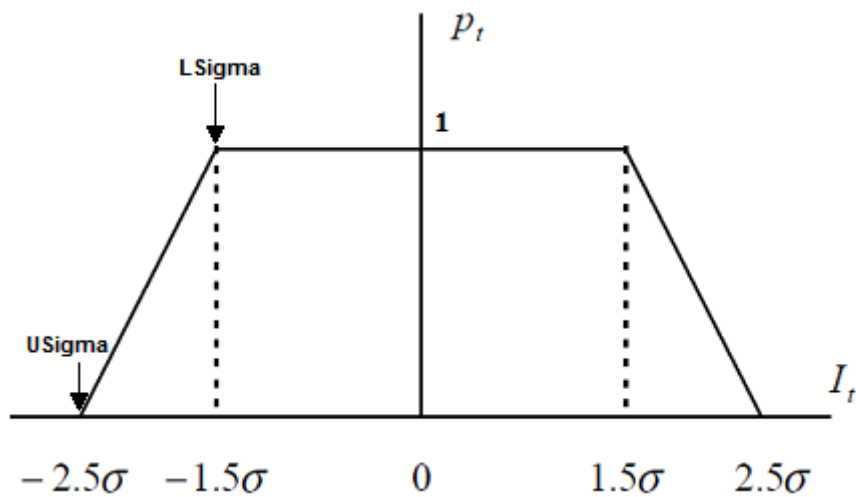


FIGURE 35 – Poids utilisés par X-13ARIMA pour la correction de l'irrégulier.

- Enfin, pour chaque date pour laquelle l'irrégulier n'a pas un poids de 1, $(S + I)_t$ est remplacée par une moyenne pondérée de 5 valeurs : la valeur à la date t avec un poids p_t et les deux valeurs précédentes et suivantes, pour le même mois, ayant une pondération de 1. En fin de période, la moyenne pondérée est calculée à partir de l'observation courante et les quatre valeurs les plus proches du même mois ayant une pondération de 1.

G Exemples d'estimations de la tendance-cycle

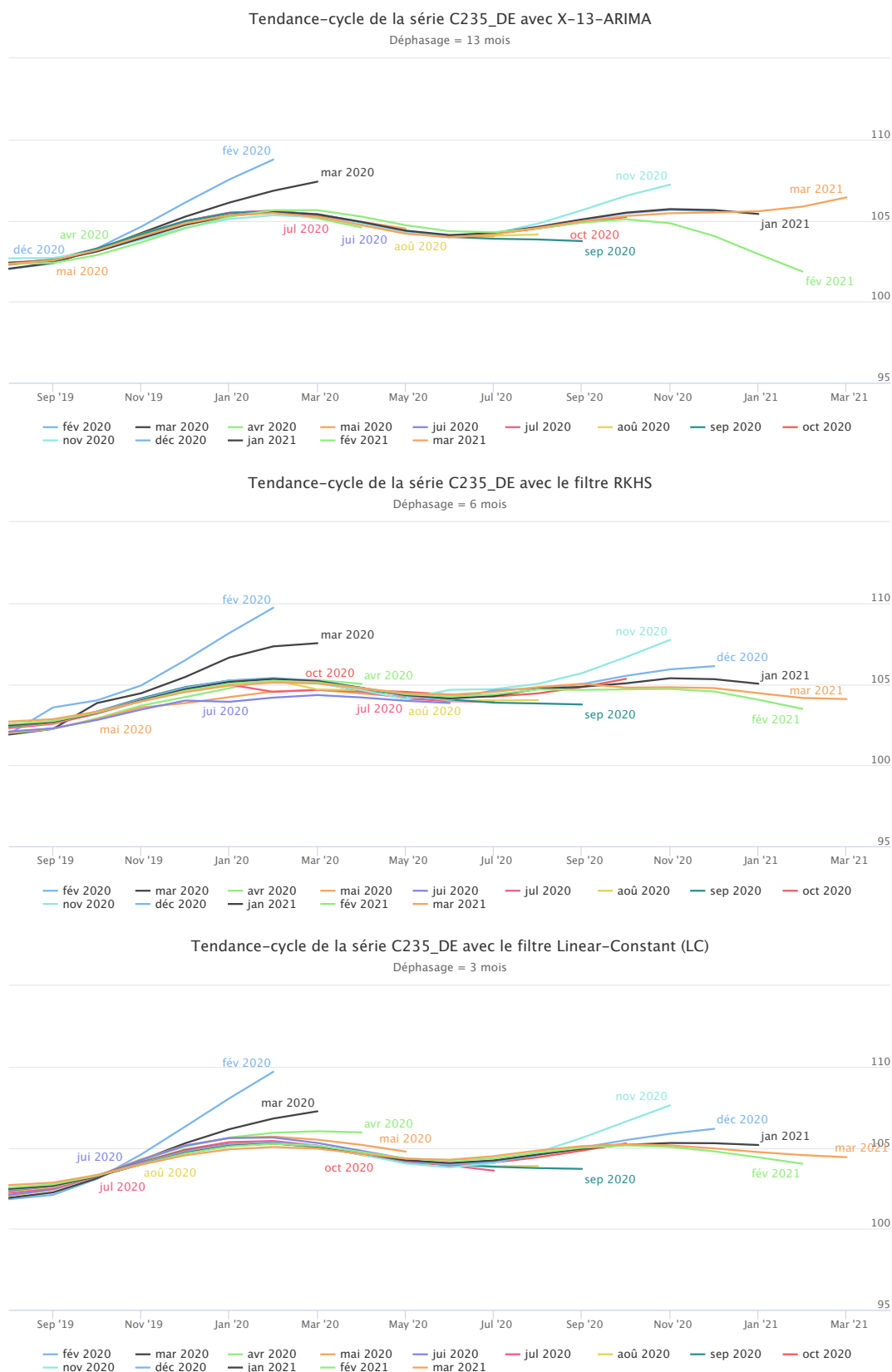


FIGURE 36 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de ciment, chaux et plâtre (C235) en Allemagne (point de retournement en février 2020).

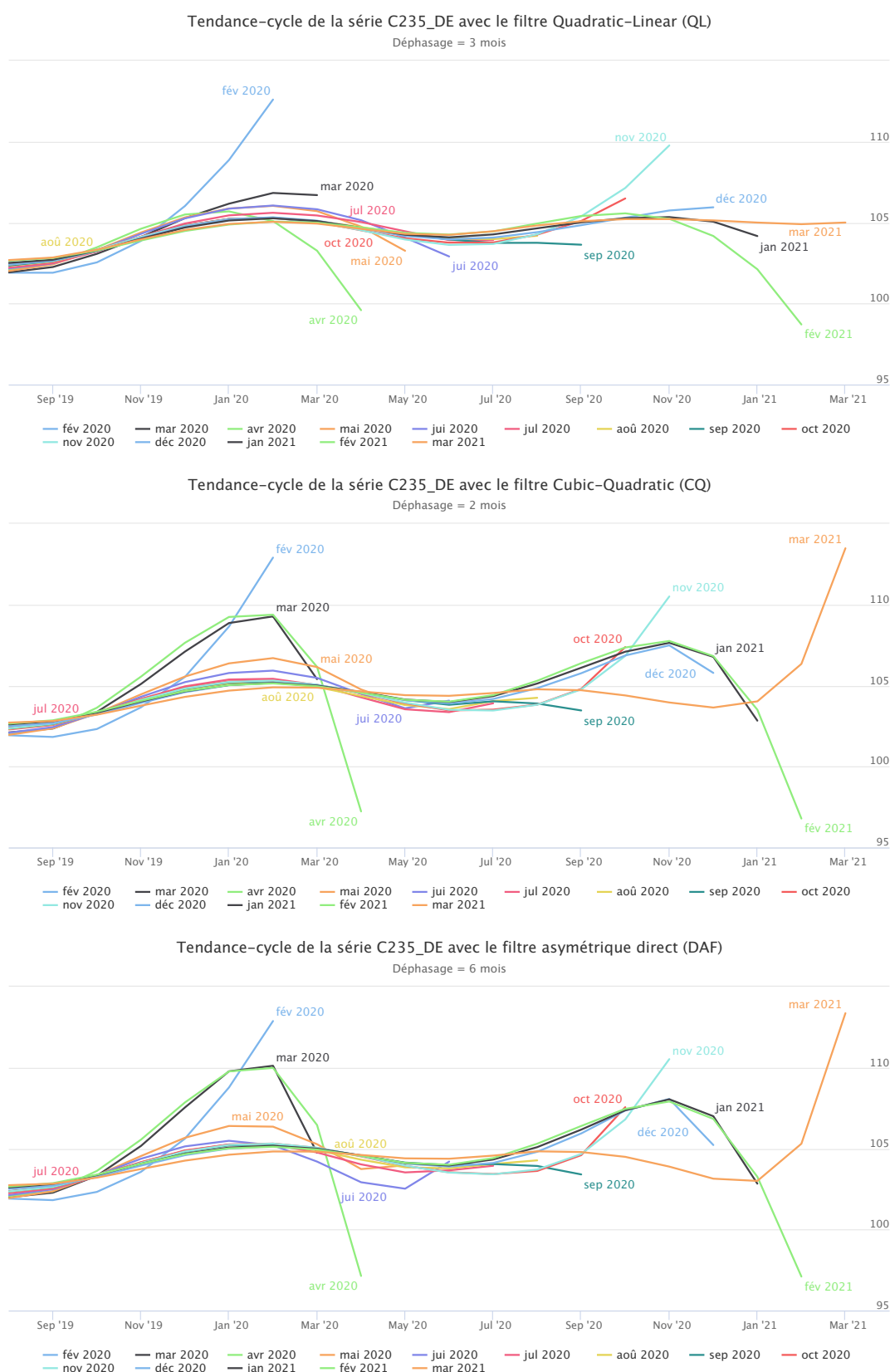
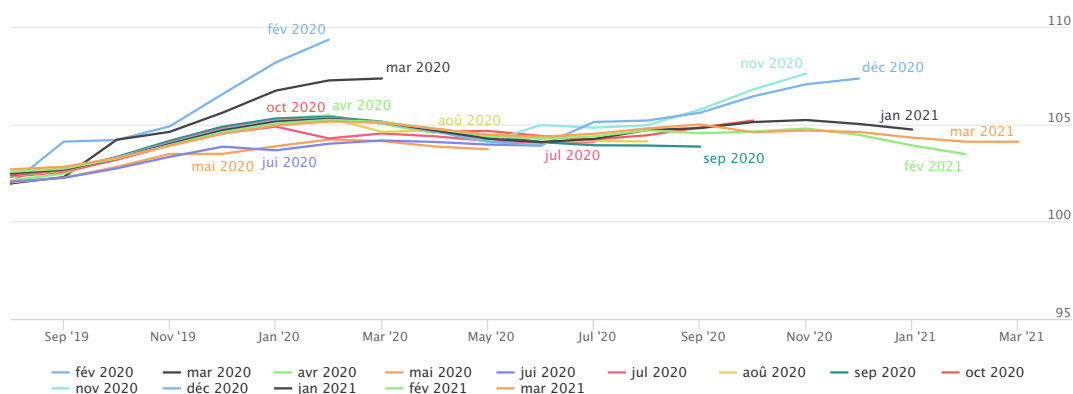
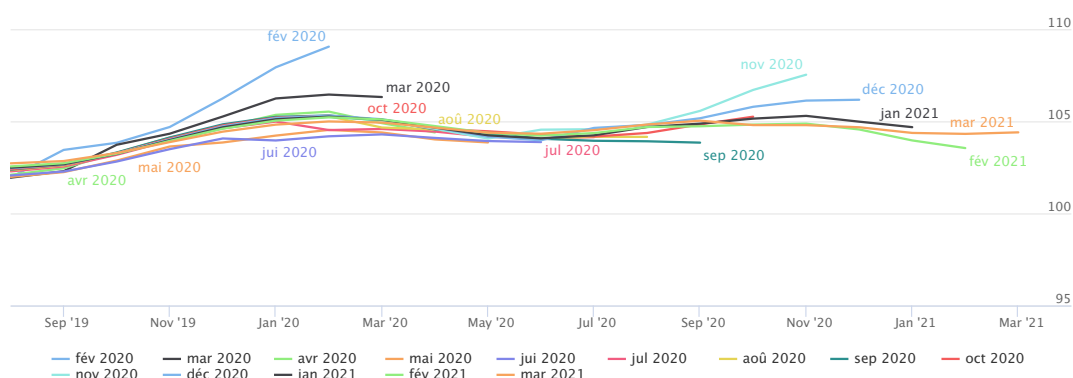


FIGURE 37 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de ciment, chaux et plâtre (C235) en Allemagne (point de retournement en février 2020).

Tendance-cycle de la série C235_DE avec le filtre FST optimal par rapport au filtre LC et minimisant la timeliness
Déphasage = 7 mois



Tendance-cycle de la série C235_DE avec le filtre FST optimal par rapport au filtre LC et ayant la timeliness la plus élevée
Déphasage = 6 mois



Tendance-cycle de la série C235_DE avec le filtre FST optimal par rapport au filtre LC et ayant une timeliness médiane
Déphasage = 6 mois

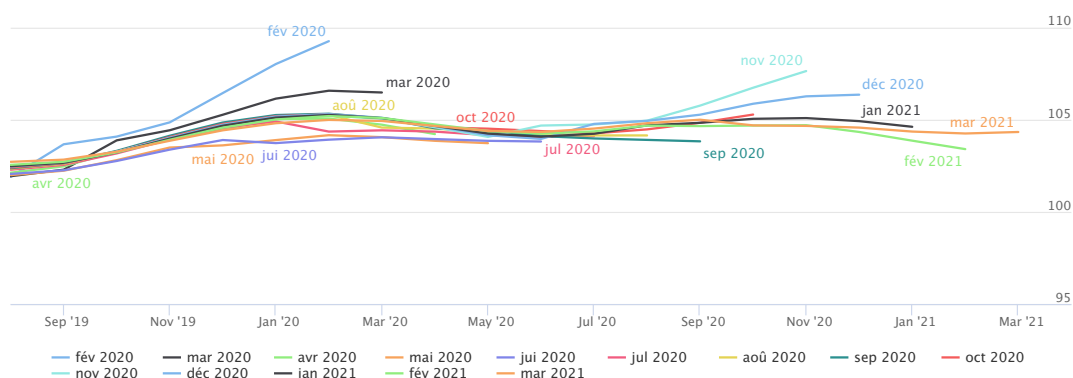


FIGURE 38 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de ciment, chaux et plâtre (C235) en Allemagne (point de retournement en février 2020).

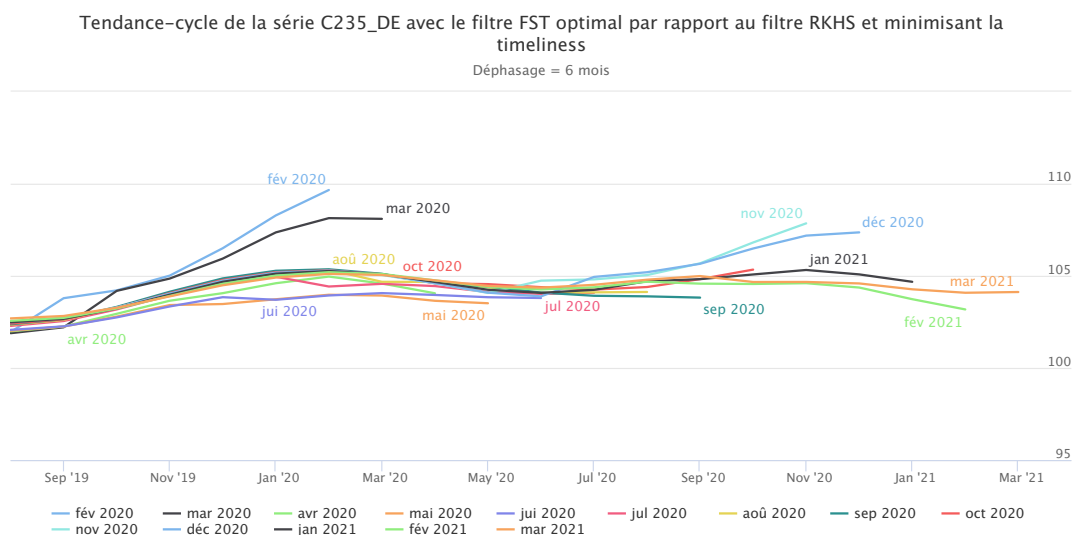


FIGURE 39 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de ciment, chaux et plâtre (C235) en Allemagne (point de retournement en février 2020).

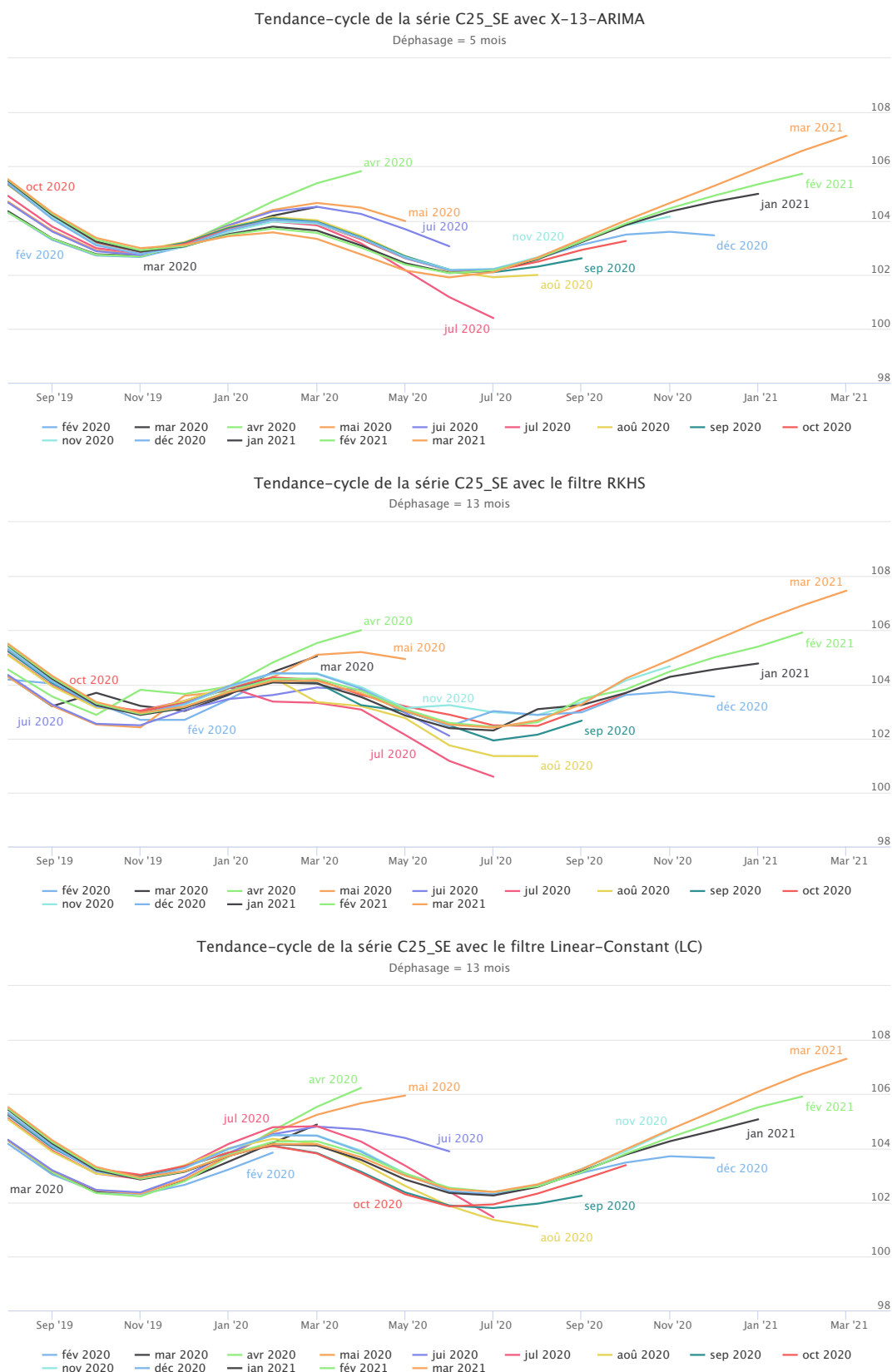


FIGURE 40 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de produits métalliques, à l'exception des machines et des équipements (C25) en Suède (point de retournement en février 2020).

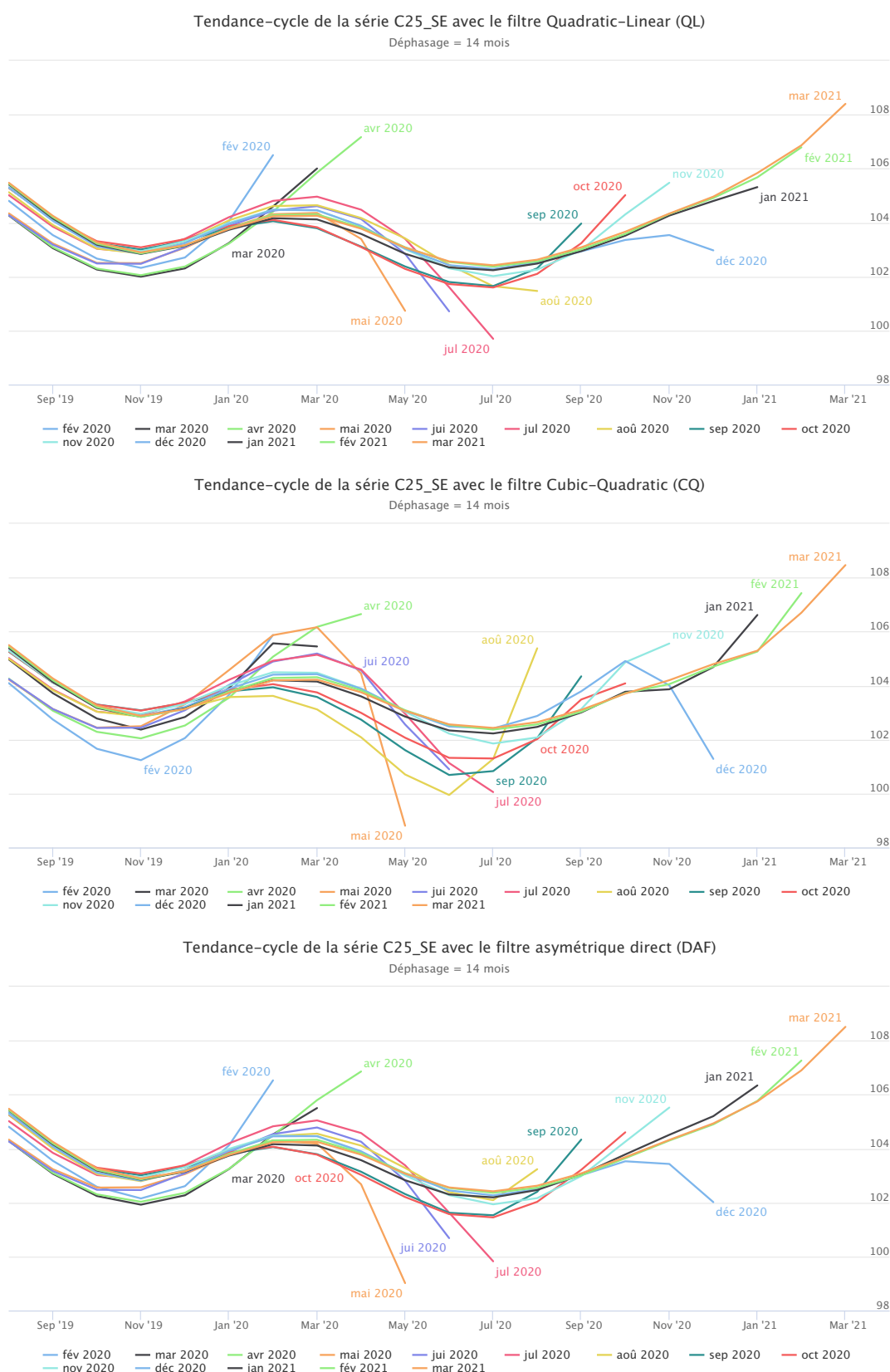


FIGURE 41 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de produits métalliques, à l'exception des machines et des équipements (C25) en Suède (point de retournement en février 2020).

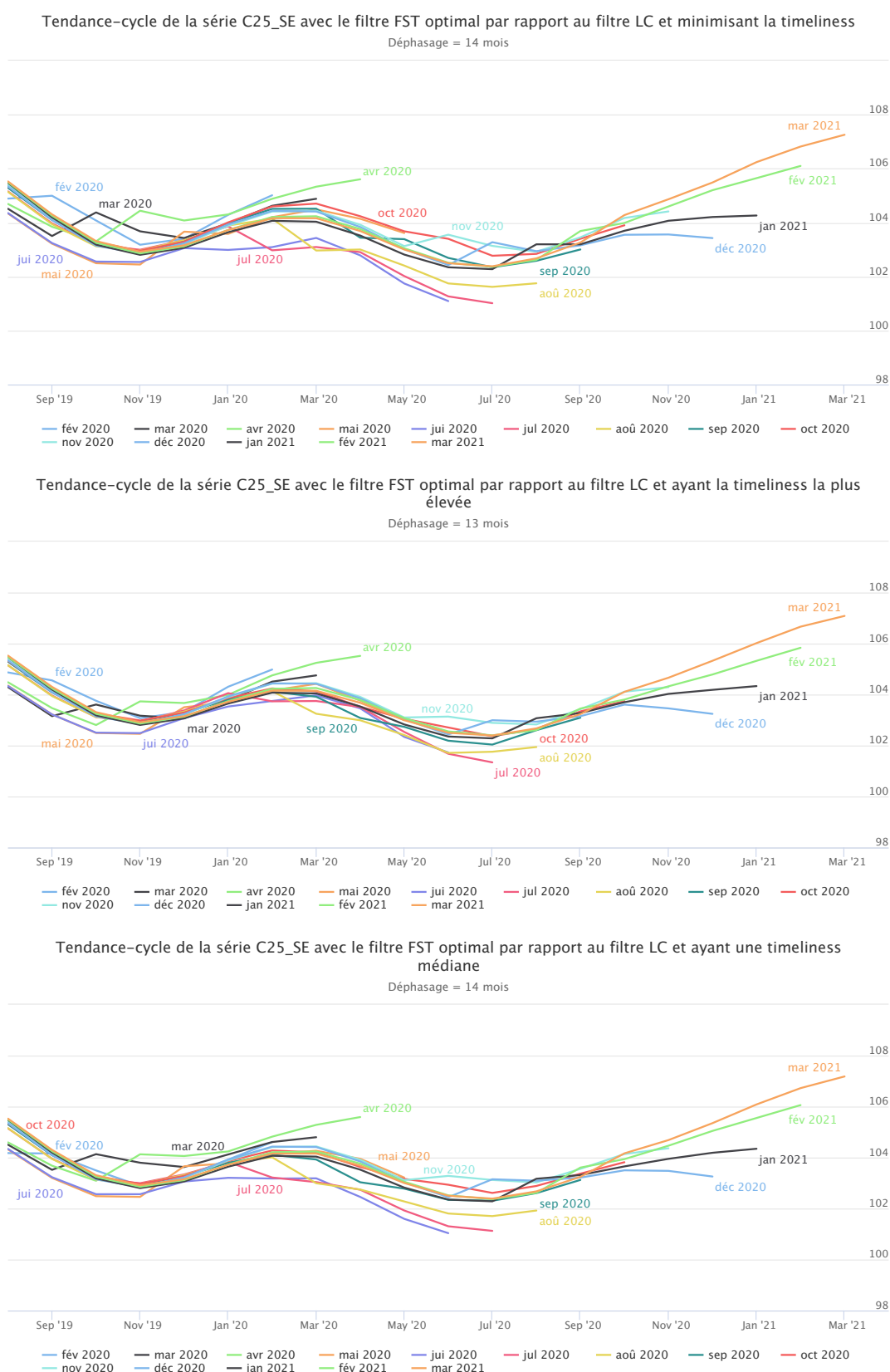


FIGURE 42 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de produits métalliques, à l'exception des machines et des équipements (C25) en Suède (point de retournement en février 2020).

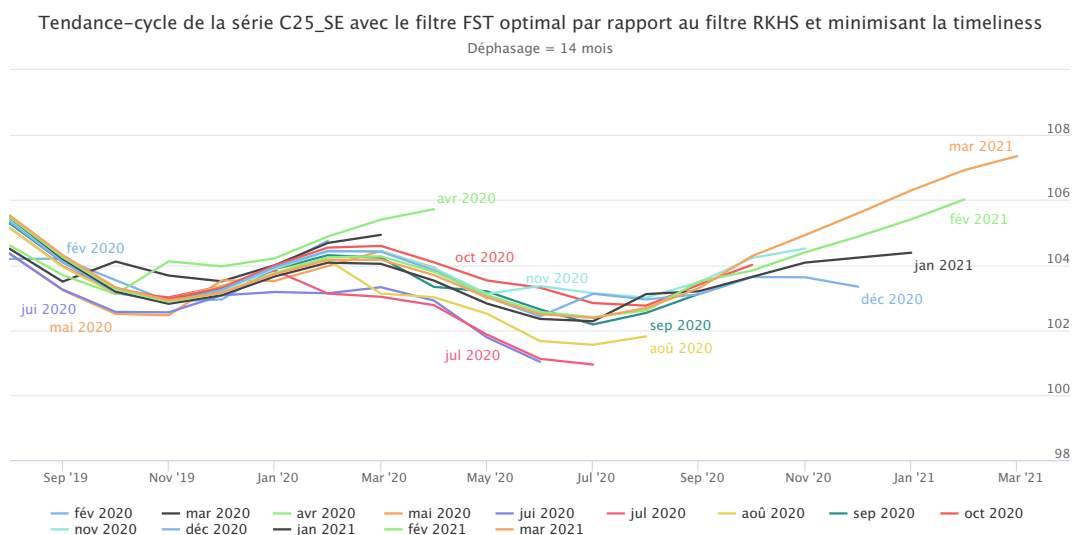


FIGURE 43 – Estimations de la tendance-cycle pour l'indice de production industriel dans la fabrication de produits métalliques, à l'exception des machines et des équipements (C25) en Suède (point de retournement en février 2020).

Références

- ALEXANDROV, Theodore et al. (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, A. (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 al. (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 (1996). “A new method to reduce unwanted ripples and revisions in trend-cycle estimates from X-11-ARIMA”. In : *Survey Methodology* 22, p. 77-84. URL : <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1996001/article/14383-fra.pdf?st=zUxzBC0K>.
- 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.
- 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, D. et B. QUENNEVILLE (2011). *Seasonal Adjustment with the X-11 Method*. Springer. ISBN : 9781461301769.
- 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 (1999). “Comprendre la méthode X11”. In : *Institut National de la Statistique et des Études Économiques*.

- 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 al. (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.
- MCELROY, Tucker S. 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://ideas.repec.org/a/eee/ecosta/v14y2020icp112-130.html) 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>.
- QUARTIER-LA-TENTE, Alain et Dominique LADIRAY (2021). “Asymmetric Linear Filters for Seasonal Adjustment: Applications to the COVID-19 Impact in Monthly Time Series”. In : *Joint Statistical Meetings (JSM)*. URL : <https://ww2.amstat.org/meetings/jsm/2021/onlineprogram/AbstractDetails.cfm?abstractid=317462>.
- 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:eb: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://ideas.repec.org/p/kof/wpskof/04-96.html). 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>.