

Techniques d'enquête

Estimation des indicateurs mensuels de la confiance des consommateurs à l'aide de modèles de séries chronologiques structurels

par Jan van den Brakel, Sabine Krieg et Marc J.E. Smeets

Date de diffusion : le 23 décembre 2025



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par la ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par la ministre de l'Industrie, 2025

L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Estimation des indicateurs mensuels de la confiance des consommateurs à l'aide de modèles de séries chronologiques structurels

Jan van den Brakel, Sabine Krieg et Marc J.E. Smeets¹

Résumé

Dans la présente étude, une procédure d'inférence selon un modèle de séries chronologiques structurel multivarié est élaborée pour la production de chiffres mensuels sur la confiance des consommateurs. Les données d'entrée pour le modèle sont cinq séries d'estimations directes pour les indices mesurant la confiance des consommateurs, tirées de l'Enquête sur la confiance des consommateurs des Pays-Bas. Le modèle améliore l'exactitude des estimations directes, puisqu'il permet une meilleure séparation des erreurs de mesure et d'échantillonnage par rapport aux paramètres ciblés estimés. Les erreurs-types pour les mouvements mois-à-mois sont clairement plus petites dans le modèle de séries chronologiques. Un deuxième problème abordé dans cette étude est lié à la transition vers un nouveau processus d'enquête en 2017. Des modèles de séries chronologiques structurels combinés à un essai parallèle sont appliqués pour estimer les discontinuités causées par le remaniement. Un algorithme conçu pour les variables de la confiance des consommateurs est élaboré pour construire des séries de données d'entrée ininterrompues pour le modèle de séries chronologiques structurel mentionné précédemment. Cette méthode d'inférence a facilité une transition harmonieuse vers un nouveau plan d'enquête et a donné lieu à des séries ininterrompues sur la confiance des consommateurs remontant jusqu'à 1986. Cette méthode est mise en œuvre pour la production de chiffres mensuels officiels sur la confiance des consommateurs aux Pays-Bas.

Mots-clés : Discontinuité de série; Enquête sur la confiance des consommateurs; modèle de séries chronologiques structurel.

1. Introduction

L'Enquête sur la confiance des consommateurs des Pays-Bas mesure la confiance des consommateurs aux Pays-Bas à une fréquence mensuelle. Jusqu'en 2016, les chiffres mensuels étaient obtenus au moyen de méthodes d'inférence fondées sur le plan de sondage appliquées aux données obtenues au moyen d'un échantillonnage probabiliste comprenant un échantillon d'environ 1 000 répondants par mois. La présente étude aborde deux problèmes. Le premier est que les séries des estimations mensuelles sont assez instables. Cela est attribuable à un échantillon relativement petit et au fait que le questionnaire de l'Enquête sur la confiance des consommateurs mesure les émotions relatives à court terme quant à l'opinion du répondant sur le climat financier et économique, malgré le fait que la plupart des questions portent en fait sur les 12 mois précédents ou les 12 mois suivants. L'augmentation de la taille de l'échantillon ne réduirait donc que partiellement l'instabilité de la série. Cela n'est néanmoins pas une option, puisque le bureau de la statistique des Pays-Bas (comme de nombreux autres instituts de statistique) doit réduire les coûts administratifs et le fardeau de réponse. Le deuxième problème est l'interruption de la continuité des séries chronologiques à la suite du remaniement du processus d'enquête ayant eu lieu au début de 2017. Les changements apportés au questionnaire et à la stratégie de travail sur le terrain ont généralement un effet systématique sur les résultats d'une enquête, puisqu'ils ont une incidence sur les erreurs non dues à

1. Jan van den Brakel, Statistics Netherlands, Department of Statistical Methods and Maastricht University, Department of Quantitative Economics.
Courriel : ja.vandenbrakel@cbs.nl; Sabine Krieg et Marc J.E. Smeets, Statistics Netherlands, Department of Statistical Methods.

l'échantillonnage, comme les biais de mesure et les biais de sélection. Les différences systématiques dans des séries chronologiques provoquées par le remaniement d'une enquête seront appelées des discontinuités dans la suite de l'étude. Dans le cadre d'un processus de transition bien conçu, les discontinuités sont quantifiées pour éviter de confondre la variation réelle d'une période à l'autre avec des changements soudains de biais de mesure et d'effets de sélection.

Comme l'Enquête sur la confiance des consommateurs est une enquête à passages répétés, un modèle de séries chronologiques structurel est élaboré comme solution à ces deux problèmes. Des modèles d'équation de séries chronologiques apparemment non liées (Seemingly Unrelated Time Series Equation) sont appliqués à l'Enquête sur la confiance des consommateurs comme forme d'estimation sur petits domaines et pour tenir compte des discontinuités causées par le remaniement de 2017. Les séries sous-jacentes des cinq indices, qui définissent ensemble la confiance des consommateurs, sont les séries de données d'entrée du modèle de séries chronologiques. Les modèles d'équation de séries chronologiques apparemment non liées sont des modèles de séries chronologiques structurels multivariés dans le cadre desquels les relations entre les séries de données d'entrée sont modélisées selon des perturbations corrélées concomitantes des variables d'état (Harvey, 1989, chapitre 8). Au moyen d'un modèle d'équation de séries chronologiques apparemment non liées, les renseignements sur l'échantillon observés au cours de périodes de référence précédentes sont utilisés pour obtenir des estimations plus précises. De plus, les corrélations entre les perturbations de la tendance, par exemple, sont utiles pour emprunter de l'information d'autres variables utilisées dans le concept de confiance des consommateurs. Dans la présente étude, nous discutons également du rôle des corrélations entre les termes irréguliers de population des séries. Des événements réels soudains influencent toutes les séries simultanément, ce qui donne des termes irréguliers de population corrélés.

L'utilisation de la modélisation de séries chronologiques dans le but d'améliorer la précision des données d'enquête a été envisagée par de nombreux auteurs et remonte à Blight et Scott (1973). Cette approche a déjà été appliquée dans le contexte des statistiques officielles; voir, par exemple, Tam (1987), Binder et Dick (1989, 1990), Bell et Hillmer (1990), Tiller (1992), Rao et Yu (1994), Pfeffermann et Burck (1990), Pfeffermann et Bleuer (1993), Pfeffermann (1991), Pfeffermann et Tiller (2006), Harvey et Chung (2000), Feder (2001), Krieg et van den Brakel (2012), van den Brakel et Krieg (2015) ainsi qu'Elliot et Zong (2019).

Les discontinuités causées par un remaniement du processus d'enquête peuvent être quantifiées de différentes façons; voir van den Brakel, Zhang et Tam (2020) pour obtenir un aperçu. L'une des approches consiste à recueillir des données selon l'ancien et le nouveau plan en parallèle pendant un certain temps, ce que l'on appelle également un essai parallèle. La différence des estimations fondées sur les deux plans peut être utilisée comme estimation directe de la discontinuité. Dans le cas d'un essai parallèle suffisamment important, il s'agit d'une approche fiable et rapide. Le principal inconvénient est qu'elle exige une collecte de données supplémentaire, ce qui rend l'approche coûteuse. Par ailleurs, un modèle de séries chronologiques peut être appliqué lorsque les discontinuités sont estimées à l'aide d'une intervention de niveau (van den Brakel et Roels, 2010). Dans le cas d'un petit essai parallèle, les renseignements de l'essai parallèle peuvent être utilisés comme information *a priori* dans le modèle de séries chronologiques, dans le cadre

d'une initialisation informative du filtre de Kalman, par exemple. Cette estimation initiale est encore améliorée par les renseignements des séries chronologiques observées avant et après l'essai parallèle (van den Brakel et Krieg, 2015).

Les estimations des discontinuités de l'Enquête sur la confiance des consommateurs causées par le remaniement sont fondées sur un essai parallèle de trois mois dont la taille de l'échantillon est égale à la taille habituelle de l'échantillon. Ces estimations sont améliorées par un modèle de séries chronologiques structurel, comme il est décrit ci-dessus. Lorsque les estimations des discontinuités sont connues, il est important de communiquer à ce sujet avec les utilisateurs des séries pour éviter une mauvaise interprétation des séries. Dans le cas de l'Enquête sur la confiance des consommateurs, les séries passées sont corrigées pour tenir compte des discontinuités. Les séries sous-jacentes à l'Enquête sur la confiance des consommateurs sont des pourcentages, et une méthode est proposée pour tenter de maintenir les valeurs corrigées dans la fourchette admissible entre 0 % et 100 %. L'approche de modélisation de séries chronologiques élaborée dans la présente étude a été mise en œuvre pour la production des chiffres mensuels officiels sur la confiance des consommateurs de Statistique Pays-Bas, depuis avril 2017.

La structure du présent article va comme suit. La section 2 fournit une description de l'Enquête sur la confiance des consommateurs des Pays-Bas. À la section 3, un modèle de séries chronologiques structurel est élaboré pour estimer des chiffres mensuels de la confiance des consommateurs. Dans cette section, les résultats sont présentés en fonction des observations pour la période précédant le remaniement en 2017. La section 4 décrit le passage au nouveau plan de sondage. Dans cette section, une méthode d'estimation des discontinuités combinant un essai parallèle et une approche de modélisation de séries chronologiques est proposée. En outre, une méthode de correction des séries observées avant le changement pour refléter le niveau des séries observées selon le nouveau plan est décrite. Enfin, les résultats des estimations des discontinuités et des séries corrigées sont présentés. La section 5 résume la façon dont la méthode d'estimation est mise en œuvre pour la production de chiffres officiels mensuels de la confiance des consommateurs. L'article se conclut par une conclusion à la section 6.

2. L'Enquête sur la confiance des consommateurs des Pays-Bas

L'Enquête sur la confiance des consommateurs est une enquête mensuelle. Avant le remaniement de 2017, chaque mois, un échantillon d'environ 2 500 ménages était tiré par échantillonnage stratifié d'une base de sondage d'adresses dérivée du registre municipal néerlandais. La stratification est fondée sur le classement recoupé de 12 provinces et le niveau d'urbanisation dans cinq catégories. Tous les ménages ont des probabilités d'inclusion égales. Les enquêtes mensuelles reposent sur des échantillons transversaux indépendants. Les ménages échantillonnés sont uniquement observés pour un mois. Dans le cas des ménages dont le numéro de téléphone fixe était connu, un intervieweur a communiqué avec eux afin de remplir le questionnaire au moyen d'une interview téléphonique assistée par ordinateur au cours des 10 premiers jours ouvrables du mois. Le questionnaire est rempli par le chef du ménage, qui est le membre du ménage

contribuant le plus au revenu du ménage. Les ménages sans numéro de téléphone fixe connu n'ont pas été contactés. En moyenne, le taux de réponse est d'environ 40 %, ce qui donne un échantillon net d'environ 1 000 ménages répondants. Une grande partie de la non-réponse concernait les ménages pour lesquels aucun numéro de téléphone fixe connu n'était disponible. Le taux de réponse parmi les ménages pour lesquels un numéro de téléphone était disponible était d'environ 60 %.

Pour l'indice de confiance des consommateurs, les questions concernant les cinq variables suivantes sont pertinentes :

1. Opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents, abrégée en « Econ. 12P »;
2. Attentes de changements de la situation économique générale du pays au cours des 12 mois suivants, abrégées en « Econ. 12S »;
3. Opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents, abrégée en « Fin. 12P »;
4. Attentes de changements de la situation financière du ménage au cours des 12 mois suivants, abrégées en « Fin. 12S »;
5. S'agit-il du bon moment pour les gens de faire des achats importants, abrégé en « Achat imp. ».

Les questions pour les quatre premières variables comportent deux options de réponse positive et deux options de réponse négative (« bien meilleure », « un peu mieux », « bien pire » et « un peu moins bonne »). De plus, une option neutre « la même » et une réponse « ne sait pas » étaient possibles. La question pour la cinquième variable comporte une option de réponse positive et une option de réponse négative (« oui, c'est le bon moment maintenant », « non, ce n'est pas le bon moment maintenant »). De plus, une option neutre (« ce n'est ni le bon moment, ni le mauvais moment ») et une réponse « ne sait pas » étaient possibles. Les pourcentages de réponses positives, négatives et neutres (en points de pourcentage du total des réponses), p_i^+ , p_i^- , p_i^0 avec $p_i^+ + p_i^- + p_i^0 = 100$, sont calculés pour chaque question $i = 1, \dots, 5$. Les estimations pour les cinq variables mentionnées précédemment sont obtenues en calculant la différence entre les réponses positives et négatives, c'est-à-dire $y_i = p_i^+ - p_i^-$ pour $i = 1, \dots, 5$. De plus, les cinq variables sont combinées en calculant les moyennes suivantes :

- $y_6 = (y_1 + y_2) / 2$, qui est l'indicateur du *climat économique*,
- $y_7 = (y_3 + y_4 + y_5) / 3$, qui est l'indicateur de la *volonté d'acheter*,
- $y_8 = (y_1 + y_2 + y_3 + y_4 + y_5) / 5$, qui est l'indicateur de la *confiance des consommateurs*.

Les indicateurs $y_1, \dots, y_8$ sont les principales variables cibles de la publication. Jusqu'à la fin de 2016, des moyennes d'échantillon non pondérées ont été utilisées comme estimations pour les variables cibles. Les expressions de la variance des huit séries sont données par van den Brakel, Söhler, Daas et Buelens (2017). La publication des chiffres mensuels a commencé en 1986. Les chiffres originaux et les chiffres

désaisonnalisées des séries d'indicateurs sont publiées. De plus, les séries sous-jacentes des pourcentages sont également publiées.

Le fait de s'adresser à des ménages au moyen d'un numéro de téléphone fixe non secret se traduit par une faible couverture de la population cible. Combinée au faible taux de réponse, la non-réponse sera non manquante de façon aléatoire. Jusqu'à la fin de 2016, ce problème est ignoré, mais il a été l'une des raisons du remaniement du processus d'enquête de l'Enquête sur la confiance des consommateurs. En janvier 2017, cinq changements importants ont été mis en œuvre simultanément : 1) le plan d'échantillonnage est passé d'un échantillon stratifié de *ménages* à un échantillon stratifié de *personnes*. Un échantillon de 2 150 personnes est tiré chaque mois. Toutes les personnes ont des probabilités d'inclusion égales. Avec un taux de réponse d'environ 47 %, l'échantillon net est d'un peu plus de 1 000 répondants; 2) Le mode de collecte des données est passé de l'interview téléphonique assistée par ordinateur à un plan de sondage à mode mixte séquentiel, dans le cadre duquel on demande d'abord aux répondants de remplir un questionnaire sur le Web. Après trois rappels, les non-répondants sur le Web sont interrogés par téléphone (dans la mesure où des numéros de téléphone sont disponibles, y compris des numéros de téléphone mobile); 3) Des changements ont été apportés au questionnaire. Le changement le plus important est la façon dont les catégories de réponses sont offertes. Selon l'ancien questionnaire, le répondant pouvait d'abord choisir entre les options « pire », « neutre » ou « meilleure ». Dans le cas de la sélection de « pire » ou de « meilleure », le répondant devait préciser si la situation était « bien » meilleure ou pire ou « un peu » meilleure ou moins bonne. Dans le nouveau questionnaire, les deux options de réponse positive et négative pour les questions 1 à 4 sont présentées directement; 4) Un autre changement important est qu'un incitatif conditionnel est fourni aux répondants pour améliorer le taux de réponse (une tablette est tirée au sort parmi les répondants); 5) Enfin, les estimations échantillonales de p_i^+ , p_i^0 et p_i^- sont fondées sur l'estimateur par la régression généralisée (Särndal, Swensson et Wretman, 1992) pour corriger, au moins en partie, la non-réponse sélective. Voir également l'annexe pour obtenir une définition détaillée de l'estimateur par la régression généralisée pour l'Enquête sur la confiance des consommateurs. Le passage à un plan séquentiel à mode mixte vise à réduire le sous-dénombrement, mais les taux de réponse demeurent très faibles. On peut s'attendre à ce que la non-réponse soit toujours non manquante de façon aléatoire et que l'estimateur par la régression généralisée ne permette pas de corriger cela suffisamment. D'autres recherches sur les méthodes de travail sur le terrain permettant d'améliorer les taux de réponse et les méthodes d'estimation permettant de corriger la non-réponse sélective sont nécessaires; voir, par exemple, Sverchkov (2008) et Pfeffermann et Sikov (2011) pour obtenir un examen des méthodes de modélisation des probabilités de réponse.

L'un des effets secondaires du remaniement est qu'il cause un changement soudain dans les effets de sélection, puisqu'une autre partie de la population est disposée à répondre au questionnaire de l'enquête lorsqu'un autre mode est appliqué et lorsqu'un incitatif est offert. De plus, il y a des changements soudains du biais de mesure en raison de l'utilisation d'un autre mode de collecte des données (en partie sans intervieweur) et aux modifications apportées au questionnaire. Ces causes sont les soi-disant discontinuités.

3. Inférence pour les chiffres mensuels de l'Enquête sur la confiance des consommateurs à l'aide de modèles de séries chronologiques structurels

Dans la présente section, on élabore un modèle de séries chronologiques structurel à l'aide des séries observées de janvier 1987 à décembre 2016. Dans le cadre d'un modèle de séries chronologiques structurel, une série est décomposée en une composante de tendance, une composante saisonnière, d'autres composantes cycliques, une composante de régression et une composante irrégulière de la variation inexpliquée. Pour chaque composante, on suppose un modèle stochastique. Cela permet d'avoir des composantes de tendance, des composantes saisonnières, des composantes cycliques et des coefficients de régression qui dépendent du temps. Au besoin, des composantes autorégressives-moyennes mobiles peuvent être ajoutées pour tenir compte de l'autocorrélation dans la série au-delà de ces composantes structurelles. Voir une introduction à la modélisation de séries chronologiques structurelle dans Harvey (1989) ou Durbin et Koopman (2012).

3.1 Description du modèle

Chaque mois t , une estimation d'échantillon directe $\hat{y}_{i,t}$ est calculée pour les cinq variables y_i ($i = 1, \dots, 5$) mesurées à l'aide des questions du questionnaire de l'Enquête sur la confiance des consommateurs, comme l'explique la section 2. La construction de modèles de séries chronologiques pour des séries d'estimations échantillonnelles commence par la formulation d'un modèle d'erreur de mesure décomposant l'estimation en un paramètre de population réel, mais inconnu, soit $\theta_{i,t}$, et une erreur d'échantillonnage, soit $e_{i,t}$:

$$\hat{y}_{i,t} = \theta_{i,t} + e_{i,t}, \quad (i = 1, \dots, 5). \quad (3.1)$$

Pour le paramètre de population inconnu, on suppose un modèle de séries chronologiques structurel de base, c'est-à-dire

$$\theta_{i,t} = L_{i,t} + S_{i,t} + I_{i,t}, \quad (i = 1, \dots, 5), \quad (3.2)$$

où $L_{i,t}$ est le niveau d'une composante de tendance stochastique modélisant la faible variation de fréquence de la série, $S_{i,t}$ est une composante stochastique modélisant la fluctuation saisonnière autour de la tendance et $I_{i,t}$ est le terme irrégulier de population. Insérer (3.2) dans (3.1) donne le modèle de séries chronologiques pour la série observée :

$$\hat{y}_{i,t} = L_{i,t} + S_{i,t} + I_{i,t} + e_{i,t}, \quad (i = 1, \dots, 5), \quad (3.3)$$

où $e_{i,t}$ est l'erreur d'échantillonnage de la série observée d'espérance égale à zéro et de variance égale à la variance de la série de données d'entrée, qui peut être dérivée à partir des échantillons d'enquête.

Les cinq séries définies peuvent être combinées en un vecteur $\hat{\mathbf{y}}_t = (\hat{y}_{1,t}, \hat{y}_{2,t}, \hat{y}_{3,t}, \hat{y}_{4,t}, \hat{y}_{5,t})'$ qui peut être modélisé à l'aide d'un modèle d'équation de séries chronologiques apparemment non liées (Harvey, 1989, chapitre 8). Dans un modèle d'équation de séries chronologiques apparemment non liées, chaque série observée apparaît dans une équation; les équations sont corrélées simultanément par les termes de perturbation de leurs variables d'état. Les modèles d'équation de séries chronologiques apparemment non liées pour (3.3) sont définis comme suit :

$$\hat{\mathbf{y}}_t = \mathbf{L}_t + \mathbf{S}_t + \mathbf{I}_t + \mathbf{e}_t \quad (3.4)$$

où $\mathbf{L}_t = (L_{1,t}, L_{2,t}, L_{3,t}, L_{4,t}, L_{5,t})'$, $\mathbf{S}_t = (S_{1,t}, S_{2,t}, S_{3,t}, S_{4,t}, S_{5,t})'$, $\mathbf{I}_t = (I_{1,t}, I_{2,t}, I_{3,t}, I_{4,t}, I_{5,t})'$ et $\mathbf{e}_t = (e_{1,t}, e_{2,t}, e_{3,t}, e_{4,t}, e_{5,t})'$. Les tendances $L_{i,t}$ ($i=1, \dots, 5$) sont modélisées à l'aide du soi-disant modèle de tendance lisse :

$$\begin{aligned} L_{i,t} &= L_{i,t-1} + R_{i,t-1}, \\ R_{i,t} &= R_{i,t-1} + \eta_{R,i,t}, \end{aligned} \quad (3.5)$$

où $L_{i,t}$ est le niveau de la tendance et $R_{i,t}$ est le paramètre de pente. Les termes de perturbation $\eta_{R,i,t}$ sont normalement distribués avec

$$\begin{aligned} E(\eta_{R,i,t}) &= 0, \\ \text{Cov}(\eta_{R,i,t}, \eta_{R,i',t'}) &= \begin{cases} \sigma_{R,i}^2 & \text{si } i=i' \text{ et } t=t' \\ \varsigma_{R,i,i'} & \text{si } i \neq i' \text{ et } t=t' \\ 0 & \text{si } t \neq t'. \end{cases} \end{aligned} \quad (3.6)$$

Il s'agit d'un modèle dynamique de la variation à faible fréquence, qui a la souplesse nécessaire pour prendre en compte la tendance ainsi que les cycles économiques. Il peut donc être interprété comme la tendance à laquelle s'ajoute le cycle économique, que l'on appelle brièvement la tendance. Le modèle de tendance lisse est choisi parce qu'il fournit une estimation de la tendance plus stable que, par exemple, le modèle au niveau local (qui comporte des termes de perturbation pour le paramètre de niveau, mais pas pour le paramètre de pente) ou le modèle de tendance linéaire local (qui comporte des termes de perturbation pour les paramètres de niveau et de pente). Cela correspond à l'objectif de l'Enquête sur la confiance des consommateurs, c'est-à-dire évaluer la confiance économique et financière au cours des 12 mois précédents et suivants. Il convient de noter que dans (3.6), les termes de perturbation de la pente peuvent être corrélés, ce qui est logique, puisque les cinq variables mesurent des phénomènes connexes.

Le soi-disant modèle saisonnier trigonométrique est utilisé pour modéliser la composante saisonnière $S_{i,t}$ ($i=1, \dots, 5$), définie comme suit :

$$S_{i,t} = \sum_{l=1}^6 S_{i,t,l}, \quad (3.7)$$

où

$$\begin{aligned} S_{i,t,l} &= S_{i,t-1,l} \cos(h_l) + S_{i,t-1,l}^* \sin(h_l) + \eta_{S,i,t,l}, \\ S_{i,t,l}^* &= S_{i,t-1,l}^* \cos(h_l) - S_{i,t-1,l} \sin(h_l) + \eta_{S,i,t,l}^*, \quad h_l = \frac{\pi l}{6}, \quad l=1, \dots, 6. \end{aligned}$$

Les perturbations $\eta_{s,i,t,l}$ et $\eta_{s,i,t,l}^*$ sont normalement distribuées avec

$$\begin{aligned} E(\eta_{s,i,t,l}) &= E(\eta_{s,i,t,l}^*) = 0, \\ \text{Cov}(\eta_{s,i,t,l}, \eta_{s,i',t',l'}) &= \text{Cov}(\eta_{s,i,t,l}^*, \eta_{s,i',t',l'}^*) = \begin{cases} \sigma_{s,i}^2 & \text{si } i = i', t = t' \text{ et } l = l' \\ 0 & \text{sinon,} \end{cases} \\ \text{Cov}(\eta_{s,i,t,l}, \eta_{s,i',t',l'}^*) &= 0 \text{ pour tout } i, t, l. \end{aligned}$$

Le modèle (3.7) est un modèle dynamique qui permet aux tendances saisonnières de changer graduellement au fil du temps. Comme dans le cas de la composante de tendance, il est possible de permettre des corrélations non nulles entre les termes de perturbation des composantes saisonnières des cinq séries de données d'entrée. Dans la présente application, les variances des termes de perturbation tendent vers zéro, ce qui sous-entend que les composantes saisonnières sont presque invariantes dans le temps et que leurs termes de perturbation sont presque égaux à zéro. Le fait de permettre des corrélations non nulles entre les termes de perturbation ne contribue pas à améliorer l'ajustement du modèle dans la présente application. Pour cette raison, une structure de covariance diagonale pour (3.7) est choisie à l'avance.

Les termes irréguliers de population de (3.4) sont normalement distribués avec

$$\begin{aligned} E(I_{i,t}) &= 0, \\ \text{Cov}(I_{i,t}, I_{i',t'}) &= \begin{cases} \sigma_{1,i}^2 & \text{si } i = i' \text{ et } t = t' \\ \varsigma_{1,i,i'} & \text{si } i \neq i' \text{ et } t = t' \\ 0 & \text{si } t \neq t'. \end{cases} \end{aligned} \quad (3.8)$$

Les erreurs d'échantillonnage de (3.4) sont normalement distribuées avec

$$\begin{aligned} E(e_{i,t}) &= 0, \\ \text{Cov}(e_{i,t}, e_{i',t'}) &= \begin{cases} \text{Var}(\hat{y}_{i,t}) & \text{si } i = i' \text{ et } t = t' \\ \text{Cov}(\hat{y}_{i,t}, \hat{y}_{i',t'}) & \text{si } i \neq i' \text{ et } t = t' \\ 0 & \text{si } t \neq t'. \end{cases} \end{aligned} \quad (3.9)$$

Il convient de noter que pour $i \neq i'$, les termes irréguliers de population $I_{i,t}$ et $I_{i',t}$ dans (3.8) peuvent être corrélés, tout comme les erreurs d'échantillonnage $e_{i,t}$ et $e_{i',t}$ dans (3.9). L'une des motivations de la modélisation des corrélations entre des termes irréguliers de population est que certains événements, par exemple de bonnes ou de mauvaises nouvelles au sujet de l'économie, influencent les réponses à toutes les questions de manière similaire.

Les variances et les covariances pour les erreurs d'échantillonnage définies dans (3.9) sont idéalement estimées à partir des données d'échantillon. Toutefois, les microdonnées des échantillons du présent projet ne sont disponibles que pour la période suivant le remaniement. Les approximations de la variance sont dérivées des estimations ponctuelles comme suit :

$$\begin{aligned} \text{Var}(\hat{y}_{i,t}) &= \text{Var}(\hat{p}_{i,t}^+ - \hat{p}_{i,t}^-) = \text{Var}(\hat{p}_{i,t}^+) + \text{Var}(\hat{p}_{i,t}^-) - 2\text{Cov}(\hat{p}_{i,t}^+, \hat{p}_{i,t}^-) \\ &= \frac{1}{n_t} \left[\hat{p}_{i,t}^+ (100 - \hat{p}_{i,t}^+) + \hat{p}_{i,t}^- (100 - \hat{p}_{i,t}^-) - 2\hat{p}_{i,t}^+ \hat{p}_{i,t}^- \right]. \end{aligned} \quad (3.10)$$

Cette approximation est fondée sur la variance des proportions dans le cadre d'un échantillonnage aléatoire simple avec remise, motivé par la petite fraction d'échantillonnage des échantillons mensuels. Voir un calcul de l'expression (3.10) dans l'annexe. L'effet d'ignorer la stratification et l'utilisation de l'estimateur par la régression généralisée, introduit lors du remaniement en 2017, dans les approximations de la variance est évalué pour la période de janvier 2017 à décembre 2019 pour laquelle les microdonnées des échantillons sont disponibles. Ignorer la stratification entraîne une surestimation des erreurs-types de 0,2 % en moyenne; les valeurs varient de -0,1 % à 0,6 %. Ignorer l'utilisation de l'estimateur par la régression généralisée entraîne une sous-estimation des erreurs-types de -0,2 % en moyenne; les valeurs varient de -3 % à 2 %. Compte tenu de ces résultats, on a décidé que (3.10) fournissait une bonne approximation des erreurs-types pour les séries de données d'entrée. On obtient des approximations des covariances de (3.9) comme suit :

$$\text{Cov}(\hat{y}_{i,t}, \hat{y}_{i',t}) = \rho_{ii'} \sqrt{\text{Var}(\hat{y}_{i,t})} \sqrt{\text{Var}(\hat{y}_{i',t})}, \quad (3.11)$$

où $\rho_{ii'}$ est la moyenne des corrélations mensuelles au cours de la période de 2017 à 2019.

Grâce à la composante de tendance et à la composante saisonnière du modèle d'équation de séries chronologiques apparemment non liées (3.4) à (3.9), des renseignements passés au sujet de l'évolution à long terme et des fluctuations saisonnières sont utilisés pour améliorer les estimations d'échantillon directes. Les covariances entre les termes de perturbation de la pente et les termes irréguliers de la population des cinq séries de données d'entrée du modèle d'équation de séries chronologiques apparemment non liées améliorent encore l'ajustement du modèle.

Les versions de modèle suivantes (3.4) à (3.9) seront comparées pour examiner l'influence des corrélations entre les termes de tendance et les termes irréguliers de population :

- Modèle 1 : le modèle décrit par les équations (3.4) à (3.9), c'est-à-dire qu'il y a des corrélations entre les perturbations de la pente de la tendance et entre les termes irréguliers de la population des cinq séries.
- Modèle 2 : semblable au modèle 1, mais sans corrélation entre les termes irréguliers de la population des cinq séries.
- Modèle 3 : semblable au modèle 1, mais sans corrélation entre les perturbations de la pente et sans corrélation entre les termes irréguliers de la population des cinq séries.

La façon générale de procéder est de placer le modèle de séries chronologiques structurel dans la représentation de l'espace d'états. Voir la représentation de l'espace d'états du modèle (3.4) à (3.9) dans l'annexe. Ensuite, le filtre de Kalman peut être appliqué pour obtenir des estimations optimales pour le vecteur d'état. Le filtre de Kalman est une procédure récursive permettant d'obtenir des estimations optimales pour le vecteur d'état à un moment t en fonction des données jusqu'à la période t , inclusivement; on les appelle estimations filtrées. Ces estimations filtrées des vecteurs d'état antérieurs peuvent être mises à jour si de nouvelles données deviennent accessibles. Cette procédure est appelée lissage. Soit α_t , le vecteur ayant des variables d'état inconnues pour la période t . Soit $\alpha_{t|t'}$, l'estimation des variables d'état pour la période t , selon les observations obtenues jusqu'à la période t' (inclusivement). Si T désigne la longueur de la série entièrement observée, alors $\alpha_{t|t}$ sont les estimations filtrées et $\alpha_{t|T}$ sont les estimations lissées.

Statistique Pays-Bas ne révisant pas les chiffres publiés à la fin du mois, l'intérêt de la présente application est principalement axé sur les estimations du modèle de séries chronologiques fondées sur l'ensemble complet de renseignements qui seraient disponibles dans le processus de production régulier afin de produire des chiffres sur la confiance des consommateurs pour le mois t . On peut en obtenir une approximation à l'aide des estimations filtrées. Les estimations filtrées pour la période t peuvent être mises à jour au moyen des renseignements devenus disponibles après la période t . Pour illustrer le gain supplémentaire lié à l'application d'un filtre de lissage, on compare également les estimations filtrées aux estimations lissées.

Les hyperparamètres, c'est-à-dire les variances et les covariances des termes de perturbation d'état et des erreurs de mesure, sont estimés avec le maximum de vraisemblance, à l'aide de la procédure d'optimisation numérique MaxBFGS (Doornik, 2007). Les estimations du maximum de vraisemblance pour les hyperparamètres sont insérées dans le filtre de Kalman, mais traitées comme s'il s'agissait des valeurs réelles, connues sans erreur. Cela signifie que l'incertitude supplémentaire liée à l'utilisation des estimations du maximum de vraisemblance pour les hyperparamètres est ignorée dans les erreurs-types des estimations filtrées et lissées pour la tendance et le signal des paramètres de l'Enquête sur la confiance des consommateurs. Il s'agit d'une approche standard de modélisation de l'espace d'états qui est acceptable dans la présente application compte tenu des longues séries disponibles; voir une motivation dans Bollinini-Balabay, van den Brakel et Palm (2017) et une approche bootstrap tenant compte de l'incertitude supplémentaire dans Pfeffermann et Tiller (2005). Enfin, les variables d'état sont initialisées au moyen d'une initialisation diffuse, sauf indication contraire. Voir Harvey (1989) ou Durbin et Koopman (2012) pour obtenir des renseignements techniques. Dans la présente étude, on utilise Ssfpack 3.0 (Koopman, Shephard et Doornik, 1999; et Koopman, Shephard et Doornik, 2008) en combinaison avec Ox (Doornik, 2007) pour les calculs.

Une fois le modèle estimé, il est possible de calculer des estimations basées sur un modèle pour les cinq séries. La tendance $L_{i,t}$ et le signal, défini comme $L_{i,t} + S_{i,t}$, sont utilisés comme estimations basées sur un modèle pour les indicateurs de la confiance des consommateurs. Dans l'application d'estimation sur petits domaines, le terme irrégulier de population est souvent compris dans les prédictions des paramètres. Dans la présente application, on décide de supprimer le terme irrégulier de population des estimations des paramètres, puisque l'objectif de l'Enquête sur la confiance des consommateurs est de mesurer l'opinion du répondant sur le climat financier et économique des 12 mois précédents et suivants. Cela se reflète mieux dans les estimations des paramètres sans le terme irrégulier de population, car ce terme mesure les fluctuations à court terme, qui ne correspondent pas à la période de 12 mois des questions posées au répondant.

Les estimations basées sur un modèle des séries combinées, c'est-à-dire le climat économique, la volonté d'acheter et la confiance des consommateurs, sont calculées comme moyennes des estimations pour les cinq séries. Les erreurs-types des estimations de modèle des séries combinées tiennent compte de la corrélation entre les variables d'état ($L_{i,t}$ et $S_{i,t}$) des séries sous-jacentes. Le calcul des indices combinés à partir des estimations de modèle des cinq séries de données d'entrée présente l'avantage que les estimations de modèle sont cohérentes sur le plan numérique. Avec l'initialisation diffuse du filtre de Kalman, les 13 premières observations sont nécessaires pour construire une distribution *a priori* appropriée pour le filtre de Kalman.

3.2 Résultats

Les trois modèles décrits à la section 3.1 sont appliqués aux séries de janvier 1987 à décembre 2016. Les chiffres de la présente section sont limités à la période de 2001 à 2016, afin de favoriser une meilleure interprétation des graphiques et d'éviter de masquer les différences dans les résultats d'estimation en raison de la congestion des graphiques. L'évaluation des modèles est fondée sur l'ensemble de la série.

Les estimations du maximum de vraisemblance pour les écarts-types des termes de perturbation de la pente, des termes de perturbation saisonnière et des termes irréguliers de la population pour les cinq variables de référence sont présentées au tableau 3.1 pour les trois modèles. Les écarts-types des termes de perturbation de la pente du modèle 2 sont supérieurs à ceux des modèles 1 et 3. Pour tous les modèles, il s'ensuit que la composante saisonnière est invariante dans le temps pour la variable « Econ. 12S » et presque invariante dans le temps pour les variables « Econ. 12P », « Fin. 12P » et « Fin. 12S ».

Tableau 3.1

Estimations du maximum de vraisemblance des écarts-types des termes de perturbation de la pente, des termes de perturbation saisonnière et des termes irréguliers de la population

Hyperparamètre	Variable	Écart-type		
		Modèle 1	Modèle 2	Modèle 3
Pente ($\sigma_{R,i}$) (3.6)	Econ. 12P	2,62	3,68	2,61
	Econ. 12S	2,89	4,53	2,86
	Fin. 12P	0,54	0,53	0,45
	Fin. 12S	0,60	0,94	0,55
	Achat imp.	0,94	1,01	0,85
Perturbation saisonnière ($\sigma_{S,i}$) (3.7)	Econ. 12P	0,03	0,02	0,00
	Econ. 12S	0,00	0,00	0,00
	Fin. 12P	0,02	0,02	0,02
	Fin. 12S	0,04	0,04	0,04
	Achat imp.	0,12	0,11	0,12
Terme irrégulier de population ($\sigma_{I,i}$) (3.8)	Econ. 12P	3,47	2,80	3,30
	Econ. 12S	5,94	4,84	5,59
	Fin. 12P	1,90	1,97	1,93
	Fin. 12S	2,24	2,10	1,97
	Achat imp.	2,15	2,21	2,18

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

Les tableaux 3.2 à 3.4 présentent les estimations du maximum de vraisemblance des corrélations des modèles 1 et 2. Les corrélations entre les perturbations de pente sont présentées au tableau 3.2 pour le modèle 1. Comme prévu, des corrélations élevées entre les termes de perturbation de pente sont observées, puisque les opinions sur la situation financière et économique sont liées. La matrice de corrélation du terme irrégulier de population selon le modèle 1 est présentée au tableau 3.3. Les corrélations entre les perturbations de pente sont présentées au tableau 3.4 pour le modèle 2. Les corrélations des perturbations de pente sont généralement plus importantes dans le modèle 2 que dans le modèle 1. En effet, selon le modèle 1, une

partie des mouvements conjoints des séries est considérée comme des corrélations entre les termes irréguliers de la population. Selon le modèle 2, cette variation est interprétée comme des fluctuations de tendance. Comme cela a déjà été mentionné à la section 3.1, il est possible de faire valoir que les termes irréguliers de population sont corrélés. Le modèle 1 est, de ce point de vue, privilégié par rapport au modèle 2, et les corrélations du tableau 3.4 sont probablement surestimées. D'après le tableau 3.1, il s'ensuit que les écarts-types des termes de perturbation de pente sont systématiquement plus élevés pour le modèle 2 que pour le modèle 1 et le modèle 3, en particulier pour les variables « Econ. 12P », « Econ. 12S » et « Fin. 12S », ce qui est une autre indication que les tendances selon le modèle 2 pourraient avoir tendance à entraîner un ajustement excessif des séries observées.

La comparaison des tendances filtrées et lissées pour les cinq séries de données d'entrée et les trois indices qui sont une combinaison linéaire des cinq séries de données d'entrée montre que la tendance filtrée selon le modèle 3 diffère souvent de celle selon les modèles 1 et 2. Les différences sont toutefois faibles et les tendances lisses sont très semblables. Comme les cinq séries de données d'entrée présentent de forts mouvements conjoints, les modèles 1 et 2 améliorent l'exactitude des estimations des tendances, puisqu'ils reposent sur les renseignements des autres séries au moyen de fortes corrélations positives entre les termes de perturbation de pente. À titre d'exemple, les tendances filtrées et lissées selon les trois modèles sont présentées pour la variable « Fin. 12S » à la figure 3.1.

Tableau 3.2

Corrélations des perturbations de la pente, modèle ayant des corrélations entre les perturbations de la pente et des corrélations entre les termes irréguliers de la population (modèle 1)

	Econ. 12P	Econ. 12S	Fin. 12P	Fin. 12S	Achat imp.
Econ. 12P	1,000				
Econ. 12S	0,865	1,000			
Fin. 12P	0,599	0,373	1,000		
Fin. 12S	0,909	0,931	0,585	1,000	
Achat imp.	0,527	0,385	0,768	0,617	1,000

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

Tableau 3.3

Corrélations des termes irréguliers de la population pour un modèle ayant des corrélations entre les perturbations de la pente et des corrélations entre les termes irréguliers de la population (modèle 1)

	Econ. 12P	Econ. 12S	Fin. 12P	Fin. 12S	Achat imp.
Econ. 12P	1,000				
Econ. 12S	0,581	1,000			
Fin. 12P	0,002	-0,169	1,000		
Fin. 12S	0,189	0,373	0,058	1,000	
Achat imp.	0,175	0,214	-0,287	-0,096	1,000

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

Tableau 3.4

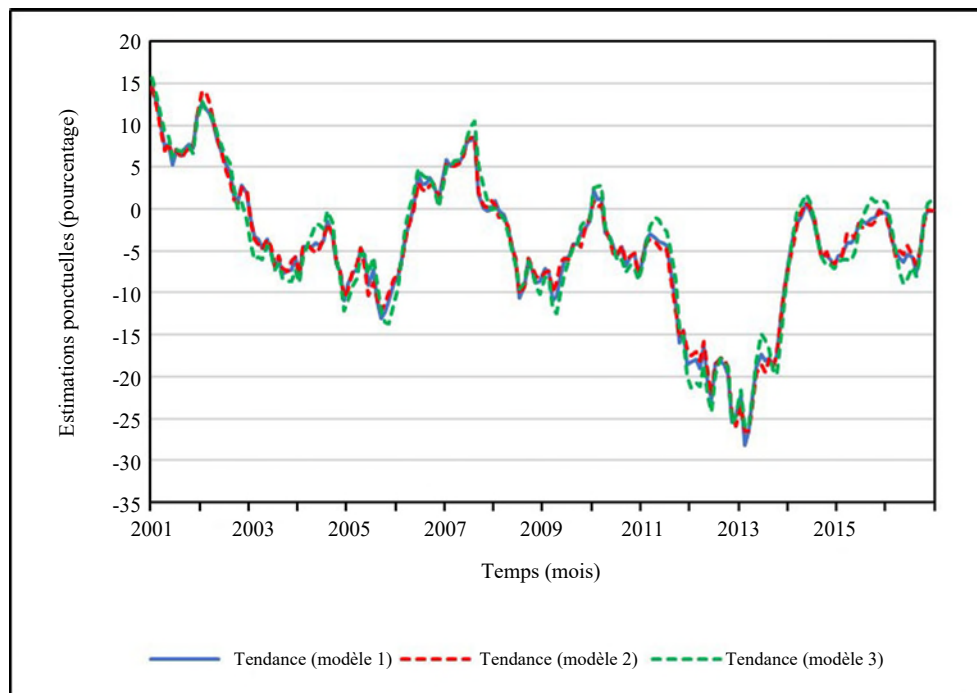
Corrélations des perturbations de la pente pour un modèle ayant des corrélations entre les perturbations de la pente et sans corrélation entre les termes irréguliers de la population (modèle 2)

	Econ. 12P	Econ. 12S	Fin. 12P	Fin. 12S	Achat imp.
Econ. 12P	1,000				
Econ. 12S	0,946	1,000			
Fin. 12P	0,601	0,427	1,000		
Fin. 12S	0,962	0,978	0,566	1,000	
Achat imp.	0,668	0,591	0,730	0,691	1,000

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

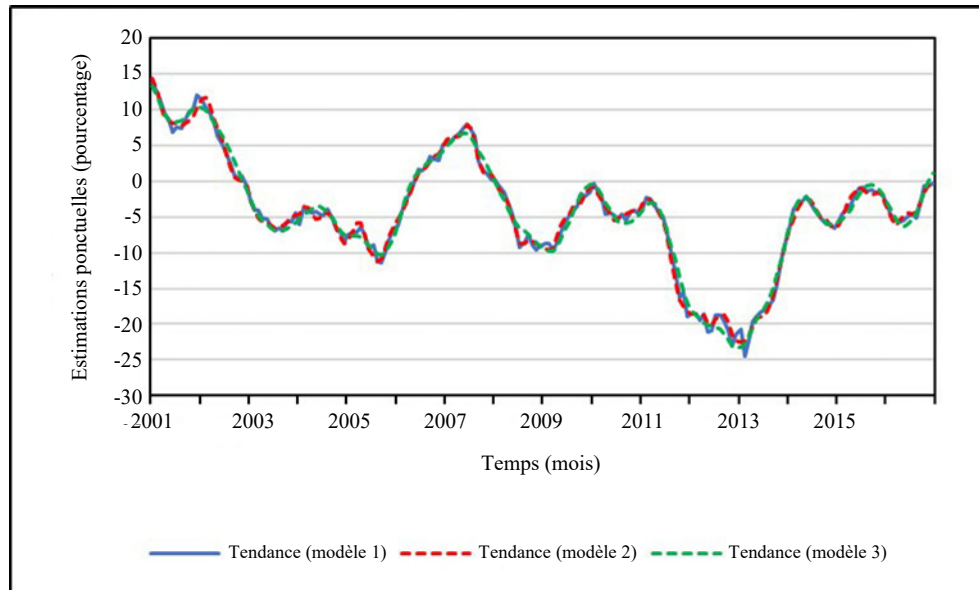
Les différences entre les signaux filtrés et lissés selon les trois modèles sont généralement faibles. Les signaux filtrés sont proches des estimations directes. Les estimations lissées sont clairement plus stables que les estimations filtrées. Bien que les différences entre les estimations lissées et filtrées soient faibles, les estimations filtrées contiennent de petits chocs, qui ne sont pas visibles dans les estimations lissées. Les estimations filtrées illustrent les chiffres sur la confiance des consommateurs publiés à la fin de la période de référence. La publication des estimations lissées plus stables nécessite une politique de révision. Toutefois, la révision continue des chiffres mensuels récemment publiés ne correspond pas à la stratégie de publication de Statistique Pays-Bas. À titre d'exemple, la figure 3.2 compare le signal filtré et lissé selon le modèle 1 aux estimations directes de la confiance des consommateurs.

Figure 3.1 Tendances filtrées (partie supérieure) et tendances lissées (partie inférieure) pour trois modèles pour la variable « Fin. 12S »



Note : La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants.

Figure 3.1 (suite) Tendances filtrées (partie supérieure) et tendances lissées (partie inférieure) pour trois modèles pour la variable « Fin. 12S »



Note : La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants.

Figure 3.2 Estimations directes et signal filtré (partie supérieure) et signal filtré et lissé (partie inférieure) selon le modèle 1 pour la confiance des consommateurs

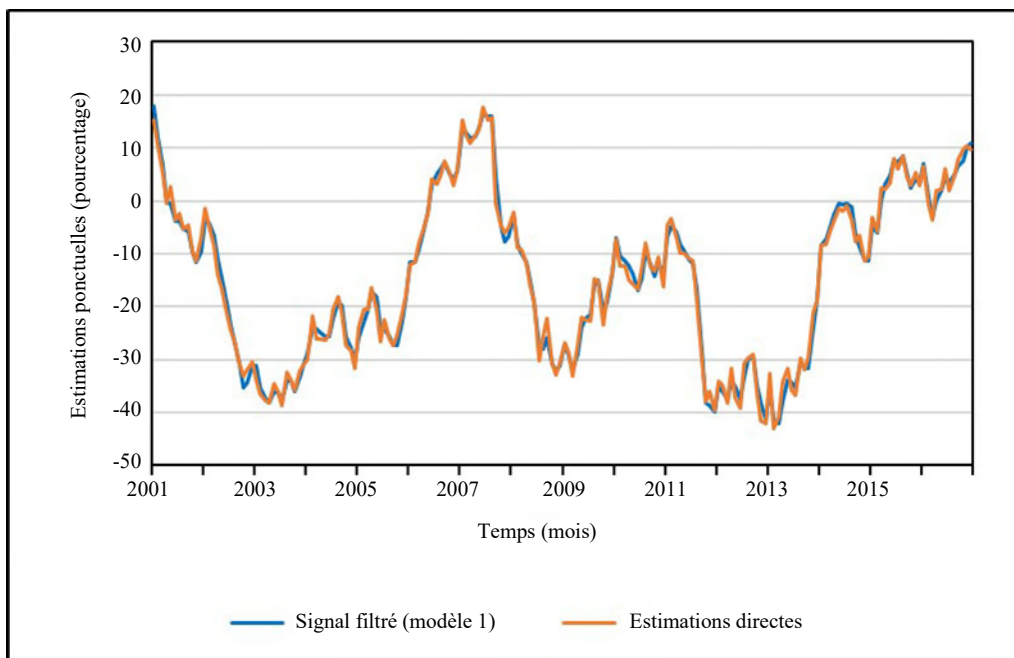
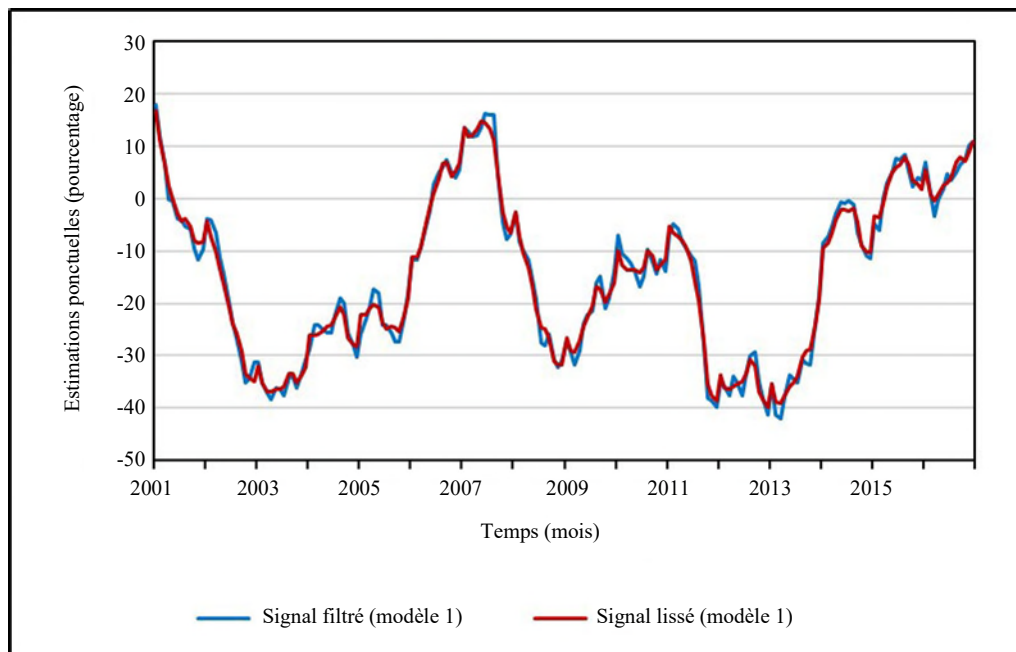
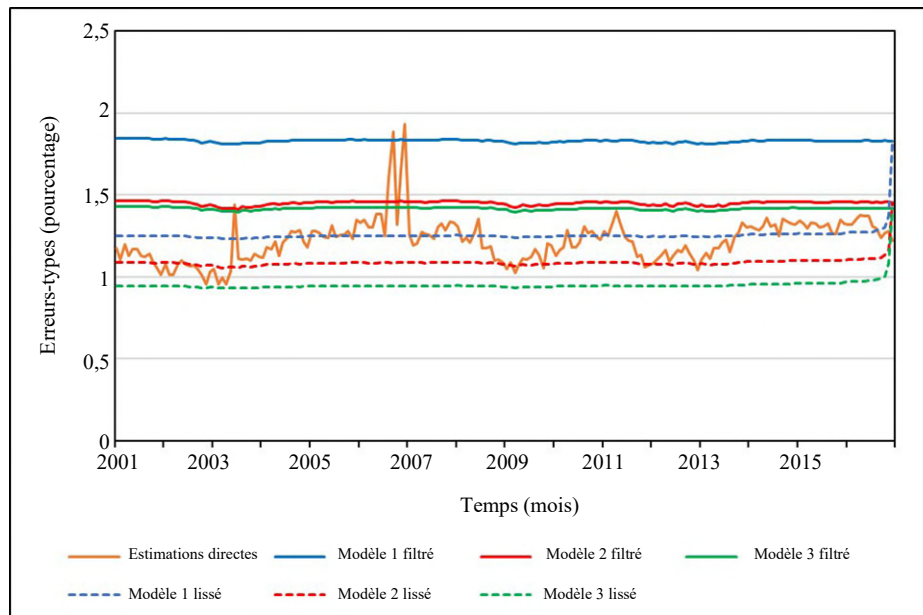


Figure 3.2 (suite) Estimations directes et signal filtré (partie supérieure) et signal filtré et lissé (partie inférieure) selon le modèle 1 pour la confiance des consommateurs



Dans la figure 3.3, les erreurs-types des signaux filtrés et lissés de la confiance des consommateurs dans le cadre des trois modèles sont comparées aux erreurs-types des estimations directes. Les erreurs-types du modèle 1 sont clairement supérieures à celles des modèles 2 et 3. Pour le modèle 1, les erreurs-types des estimations filtrées sont clairement supérieures à celles des estimations directes. Les erreurs-types des estimations lissées du modèle 1 et les erreurs-types des estimations filtrées des modèles 2 et 3 sont plus ou moins semblables à celles des estimations directes. Pour les deux autres séries combinées, c'est-à-dire le climat économique et la volonté d'acheter, les erreurs-types des signaux filtrés et lissés du modèle 1 sont également supérieures à celles des modèles 2 et 3. Pour le climat économique, l'erreur-type des estimations directes est inférieure aux erreurs-types des signaux filtrés dans les trois modèles. Les erreurs-types des signaux lissés dans les modèles 2 et 3 sont plus ou moins semblables aux erreurs-types des estimations directes. Pour ce qui est de la volonté d'acheter, les erreurs-types des signaux filtrés et des signaux lissés dans les trois modèles sont inférieures à celles des estimations directes (résultats non présentés).

Figure 3.3 Erreurs-types des estimations directes et des estimations de modèle ainsi que pour le signal selon les trois modèles de confiance des consommateurs



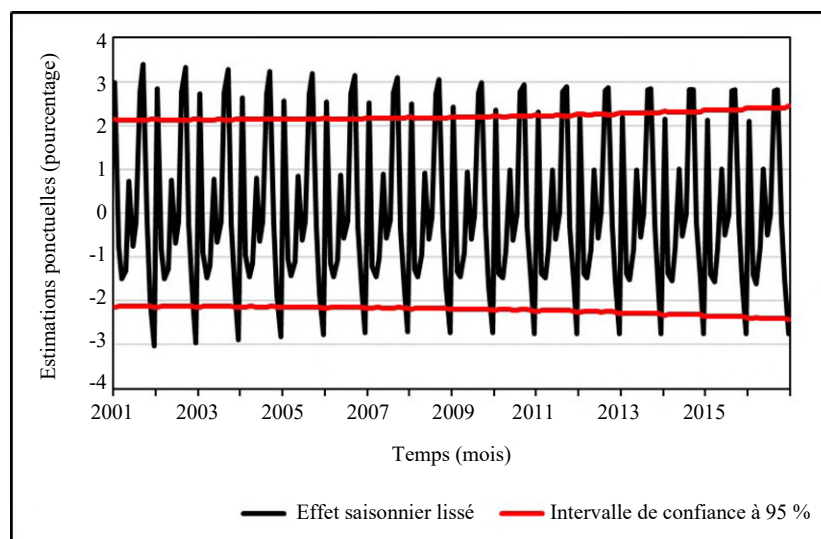
L'erreur-type des signaux filtrés est égale, voire supérieure à celle des estimations directes. Cela ne correspond pas à la constatation générale figurant dans la documentation indiquant que les modèles d'espace d'états appliqués aux séries obtenues au moyen d'enquêtes répétées donnent des estimations de modèle comportant des erreurs-types substantiellement inférieures aux erreurs-types des estimations d'enquête directes; voir, par exemple, Pfeffermann et Bleuer (1993), Pfeffermann et Burck (1990), Pfeffermann et Tiller (2006), Krieg et van den Brakel (2012), van den Brakel et Krieg (2015, 2016), et Boonstra et van den Brakel (2019). Même s'il est difficile de comparer les erreurs-types d'estimations d'échantillon directes à celles de méthodes d'inférence basées sur un modèle, on peut interpréter notre constatation empirique pour l'Enquête sur la confiance des consommateurs des Pays-Bas comme suit. Les estimations du maximum de vraisemblance présentées au tableau 3.1 pour le terme irrégulier de population sont au moins aussi élevées que la variance de l'erreur d'échantillonnage. Cette incertitude supplémentaire liée au terme irrégulier de population se reflète dans l'erreur-type des estimations du modèle de séries chronologiques, qui sont des estimations pour la tendance, L_t , ou le signal, $L_t + S_t$. Les estimations d'échantillon de l'Enquête sur la confiance des consommateurs des Pays-Bas sont des estimations fondées sur le plan pour le paramètre de population θ_t . Leurs erreurs-types, quant à elles, ne contiennent que l'incertitude de l'erreur d'échantillonnage. De plus, les corrélations positives entre les termes irréguliers de population augmentent particulièrement les erreurs-types des indices composites. Il est néanmoins préférable de tenir compte de la corrélation entre les termes irréguliers de population, puisque cela se traduit par une séparation légèrement meilleure entre les fluctuations à court terme saisies par les termes irréguliers de population par rapport à la tendance et au signal.

Les estimations des tendances filtrées sont relativement instables, ce qui indique, lorsque l'on tient compte aussi des termes irréguliers de population relativement grands, que les questions de l'Enquête sur la confiance des consommateurs mesurent les émotions à court terme et ne sont pas interprétées par les

répondants comme une évaluation à long terme de l'économie et de la situation financière pour les 12 mois précédents et suivants. Cette observation est appuyée par le fait que les séries chronologiques ont une tendance saisonnière, qui ne serait pas présente si les questions étaient interprétées comme la situation relative aux 12 mois précédents et suivants. Voir la tendance saisonnière de la variable « Econ. 12P » dans la figure 3.4 à titre d'exemple. Comme on l'a mentionné précédemment, la contribution la plus importante du modèle de séries chronologiques dans la présente application est obtenue avec les estimations lissées, plus stables et présentant des erreurs-types plus faibles.

Les erreurs-types des tendances et des signaux filtrés étant stables au fil du temps, les erreurs-types des cinq séries de référence et des trois séries combinées sont présentées au tableau 3.5 pour décembre 2016 (dernière observation avant le passage au nouveau plan de sondage). Les erreurs-types pour les tendances et les signaux lissés sont également stables au fil du temps, à l'exception des derniers mois des séries. Par conséquent, les erreurs-types pour les signaux lissés sont comprises dans le tableau 3.6 pour décembre 2015. Les erreurs-types des tendances sont très semblables aux erreurs-types des signaux. Pour les cinq séries de référence, l'erreur-type dans le modèle 1 est généralement inférieure à celles du modèle 3. Toutefois, pour les séries combinées, les erreurs-types dans le modèle 1 sont généralement supérieures à celles des modèles 2 et 3. À partir de la comparaison entre les modèles 2 et 3, il s'ensuit que la modélisation des corrélations transversales au moyen de la composante de tendance améliore la précision des estimations de modèle. La modélisation de la corrélation entre les termes irréguliers de population diminue la précision parce que la corrélation positive augmente la variance des termes irréguliers de population des séries combinées. Cela s'applique en particulier aux indicateurs du climat économique et de la confiance des consommateurs. Pour l'indicateur de la volonté d'acheter, les termes irréguliers de population présentent de faibles corrélations négatives, comme le montre le tableau 3.3. Comme cela a été expliqué ci-dessus, il est nécessaire de tenir compte de ces corrélations pour éviter de sous-estimer l'incertitude des prédictions de modèle.

Figure 3.4 Effet saisonnier lissé de la variable « Econ. 12P » et son intervalle de confiance à 95 %



Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents.

Tableau 3.5

Erreur-type des estimations filtrées de la tendance et du signal pour la dernière période (décembre 2016) (décembre 2015) pour huit séries

	Modèle 1		Modèle 2		Modèle 3	
	tendance	signal	tendance	signal	tendance	signal
Econ. 12P	3,17	3,00	3,23	2,75	3,01	3,10
Econ. 12S	4,77	4,67	4,79	4,00	4,25	4,68
Fin. 12P	1,47	1,45	1,74	1,71	1,68	1,76
Fin. 12S	1,55	1,59	1,84	1,62	1,50	1,85
Achat imp.	2,07	1,84	2,34	2,22	2,30	2,22
Climat économique	3,60	3,61	2,95	2,74	3,05	2,86
Volonté d'acheter	1,05	1,20	1,23	1,23	1,26	1,22
Confiance des consommateurs	1,76	1,83	1,44	1,46	1,61	1,42

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

Tableau 3.6

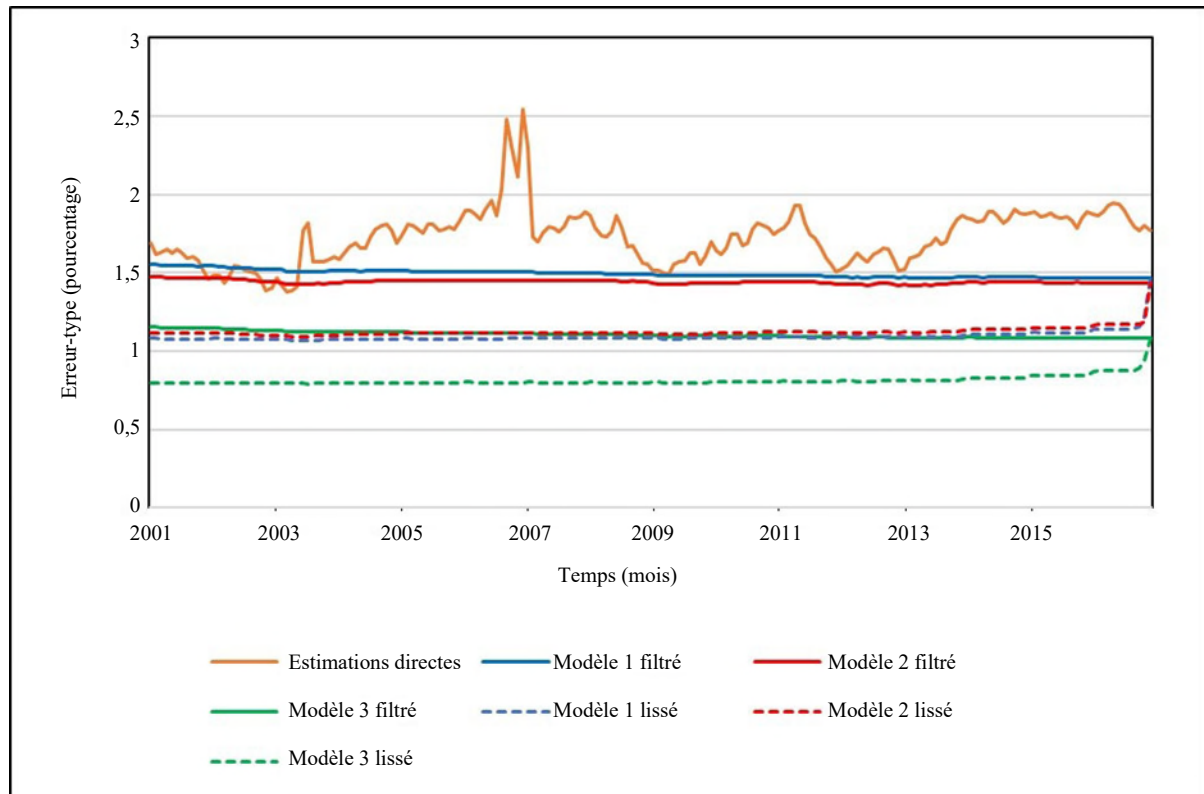
Erreurs-types des estimations lissées de la tendance et du signal pour la dernière période (décembre 2015) pour 8 séries

	Modèle 1		Modèle 2		Modèle 3	
	tendance	signal	tendance	signal	tendance	signal
Econ. 12P	2,23	2,20	2,22	2,05	2,19	2,15
Econ. 12S	3,10	3,16	2,99	2,82	3,06	3,12
Fin. 12P	1,04	1,12	0,96	1,14	0,99	1,16
Fin. 12S	1,11	1,29	0,90	1,22	1,05	1,34
Achat imp.	1,34	1,57	1,30	1,81	1,31	1,81
Climat économique	2,44	2,46	2,33	2,08	1,91	1,93
Volonté d'acheter	0,76	0,90	0,76	0,92	0,69	0,89
Confiance des consommateurs	1,22	1,27	1,19	1,11	0,90	0,97

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

L'un des principaux avantages de l'inférence basée sur les modèles de séries chronologiques est le gain de précision des variations d'une période à l'autre. Pour illustrer cela, la figure 3.5 compare les erreurs-types de l'évolution d'un mois à l'autre pour les signaux filtrés et lissés relatifs à la confiance des consommateurs selon les trois modèles et les estimations directes. La variation d'une période à l'autre et leurs erreurs-types sont obtenues en calculant la combinaison linéaire de $\Delta_t = L_{t|t} - L_{t-1|t} + S_{t|t} - S_{t-1|t}$ au moyen de la récursion par filtre de Kalman. Pour ce faire, les variables d'état pour L_{t-1} et S_{t-1} demeurent dans le vecteur d'état. Pour les estimations directes, ces erreurs-types sont plus élevées, puisque les estimations directes de deux périodes différentes dans une enquête transversale sont indépendantes. Toutefois, pour les estimations de modèle, ces erreurs-types sont plus faibles principalement en raison de la forte corrélation positive entre les niveaux de tendance de deux périodes qui se suivent.

Figure 3.5 Erreurs-types pour l'évolution d'un mois à l'autre des estimations directes et des estimations filtrées du signal selon trois modèles de la confiance des consommateurs



3.3 Évaluation du modèle

Les hypothèses sous-jacentes du modèle d'espace d'états sont évaluées en vérifiant si les innovations standardisées suivent une distribution standard normale et indépendante; voir Durbin et Koopman (2012), sections 2.12 et 7.5. Différents tests (les tests de normalité de Bowman-Shenton, les tests F pour l'hétéroscédasticité, les graphiques quantile-quantile, les graphiques des innovations standardisées et les corrélogrammes d'échantillon, et le test de Durbin-Watson) indiquent de petits non-respects de ces hypothèses pour tous les modèles. Les résultats en matière de normalité et d'hétéroscédasticité sont comparables selon tous les modèles. Les corrélogrammes présentent une autocorrélation très faible pour les 12 premiers décalages de certaines séries et n'indiquent pas de présence d'autocorrélation résiduelle. D'autres modèles de tendance, c'est-à-dire un modèle au niveau local ou un modèle de tendance linéaire local, et une composante saisonnière ayant des composantes de variance distinctes pour les harmoniques de (3.7) ou des composantes autorégressives (1), de moyennes mobiles (1) ou autorégressives-moyennes mobiles (1,1) n'ont pas amélioré ces diagnostics.

Les trois modèles sont imbriqués et sont donc également comparés expressément à l'aide d'un test du rapport de vraisemblance. Le tableau 3.7 présente les logarithmes du rapport de vraisemblance et les

résultats des tests du rapport de vraisemblance. Le test pour le modèle 2 par rapport au modèle 3 indique que la modélisation de la corrélation entre les termes de perturbation de pente améliore significativement l'ajustement du modèle. Le test pour le modèle 1 par rapport au modèle 2 montre que la modélisation de la corrélation entre les termes irréguliers de population améliore encore l'ajustement du modèle de façon significative. Enfin, le test pour le modèle 1 par rapport au modèle 3 montre que le test conjoint sur l'inclusion d'une matrice de covariance complète pour les termes de perturbation de la pente et les termes irréguliers de la population permet de rejeter l'hypothèse nulle selon laquelle les deux modèles sont équivalents. Selon les tests du rapport de vraisemblance, le modèle 1 est sélectionné pour l'inférence de l'Enquête sur la confiance des consommateurs des Pays-Bas.

Tableau 3.7**Logarithmes du rapport de vraisemblance et résultats des tests du rapport de vraisemblance**

Modèle	Logarithme du rapport de vraisemblance	Comparaison	Statistique du rapport de vraisemblance	Degrés de liberté	Valeur de p
Modèle 1	-5 293,09	Modèle 1 par rapport au modèle 2	107,96	10	0000
Modèle 2	-5 335,70	Modèle 2 par rapport au modèle 3	309,20	10	0000
Modèle 3	-5 476,66	Modèle 1 par rapport au modèle 3	417,16	20	0000

4. Discontinuités

Un aspect important de la mise en œuvre du nouveau plan en 2017 est de quantifier les discontinuités dans les résultats de l'Enquête sur la confiance des consommateurs des Pays-Bas résultant de la mise en œuvre d'un nouveau processus d'enquête et de les prendre en compte. Dans la présente application, les discontinuités sont analysées au niveau du pourcentage des catégories de réponses positives, neutres ou négatives aux cinq questions, c'est-à-dire p_i^+ , p_i^- , p_i^0 pour $i=1,\dots,5$, puisque ces pourcentages sont les variables mesurées dans le questionnaire. Ces estimations des discontinuités servent à calculer des séries ininterrompues pour les pourcentages $p_{i,t}^+$, $p_{i,t}^0$, $p_{i,t}^-$ pour $i=1,\dots,5$, en corrigeant les séries observées avant le changement pour refléter le niveau des séries observées dans le cadre du nouveau plan de sondage. Ces séries corrigées sont utilisées lors d'une deuxième étape pour calculer des séries ininterrompues pour $y_{1,t}, \dots, y_{5,t}$. Ces séries corrigées seront utilisées comme données d'entrée pour le modèle 1, déterminé par (3.4) à (3.9), qui est utilisé dans la production des chiffres mensuels officiels sur la confiance des consommateurs. La méthode de correction des séries de pourcentages observées avant le changement est décrite dans la présente section. La méthode proposée garantit que la taille de la correction pour une période donnée dépend de la taille des discontinuités de ces pourcentages et de la part des pourcentages dans les trois catégories à cette période. On soutient qu'une telle correction est plus réaliste qu'une approche permettant d'estimer et de corriger directement les discontinuités au niveau des séries de données d'entrée $y_{1,t}, \dots, y_{5,t}$.

4.1 Quantification des discontinuités

Dans la présente application, les discontinuités sont estimées au moyen d'un essai parallèle combiné à une analyse d'intervention de séries chronologiques. Le budget disponible a permis un essai parallèle de trois mois au premier trimestre de 2017, au cours duquel les tailles d'échantillon des deux plans étaient égales à la taille nette de l'échantillon de l'enquête régulière, soit environ 1 000 personnes. Les estimations directes initiales des discontinuités sont obtenues sous forme de contrastes entre les estimations d'échantillon directes ou fondées sur le plan dans le cadre de l'ancien et du nouveau plan d'enquête. La précision des discontinuités estimées peut être encore améliorée en modélisant les séries observées au moyen d'un modèle de séries chronologiques structurel. Le paramètre de population d'intérêt est modélisé avec une composante de tendance, une composante saisonnière, un terme irrégulier de population et l'erreur d'échantillonnage, comme pour les modèles élaborés à la section 3. La discontinuité est modélisée au moyen d'une variable d'intervention qui passe de zéro à un au moment où l'enquête passe de l'ancien au nouveau plan. En supposant que les autres composantes du modèle de séries chronologiques (composante de tendance et saisonnalité) décrivent correctement l'évolution du paramètre de population, le coefficient de régression de la variable d'intervention peut être interprété comme une estimation de la discontinuité. Harvey et Durbin (1986) ont initialement proposé l'approche d'intervention de niveau avec des modèles d'espace d'états pour estimer l'effet de la loi sur les ceintures de sécurité sur les victimes des accidents de la route au Royaume-Uni. Si aucune information tirée d'un essai parallèle n'est disponible, le coefficient de régression de la variable d'intervention est initialisé dans le filtre de Kalman avec une distribution *a priori* diffuse. L'estimation directe de la discontinuité et de sa variance obtenues à partir de l'essai parallèle peut être utilisée comme une initialisation informative du coefficient de régression de l'intervention de niveau dans le filtre de Kalman. De cette façon, l'estimation des discontinuités découlant de l'essai parallèle est améliorée au moyen des renseignements disponibles pour l'ensemble des séries chronologiques.

Selon l'essai parallèle, il existe deux estimations pour les pourcentages des cinq questions $p_{i,t}^+$, $p_{i,t}^0$, $p_{i,t}^-$, désignées par $\hat{p}_{i,t}^{j,O}$ ($j = \{+, 0, -\}$) pour l'estimation de $p_{i,t}^j$ selon l'ancien plan et par $\hat{p}_{i,t}^{j,N}$ pour l'estimation de $p_{i,t}^j$ selon le nouveau plan. Les estimations directes des discontinuités sont obtenues comme suit :

$$\hat{\Delta}(p_i^j) = \frac{1}{3} \sum_{t=2017(1)}^{2017(3)} [\hat{p}_{i,t}^{j,N} - \hat{p}_{i,t}^{j,O}], \text{ pour } j = \{+, 0, -\}. \quad (4.1)$$

Comme les pourcentages $p_{i,t}^+$, $p_{i,t}^0$, $p_{i,t}^-$ pour chaque question i totalisent 100 %, il s'ensuit que

$$\sum_{j \in \{+, 0, -\}} \hat{\Delta}(p_i^j) = 0. \quad (4.2)$$

La variance des estimations des discontinuités peut être estimée par

$$\widehat{\text{Var}}[\hat{\Delta}(p_i^j)] = \frac{1}{9} \sum_{t=2017(1)}^{2017(3)} [\widehat{\text{Var}}(\hat{p}_{i,t}^{j,O}) + \widehat{\text{Var}}(\hat{p}_{i,t}^{j,N})] \quad (4.3)$$

avec

$$\widehat{\text{Var}}(\hat{p}_{i,t}^{j,d}) = \hat{p}_{i,t}^{j,d} (100 - \hat{p}_{i,t}^{j,d}) / n_t^d, \text{ pour } d \in \{O, N\}$$

et n_t^O, n_t^N , la taille de l'échantillon au mois t selon l'ancien et le nouveau plan.

À l'étape suivante, un modèle multivarié tridimensionnel est appliqué aux trois séries avec des pourcentages de réponses positives, négatives et neutres pour chaque question séparément, c'est-à-dire pour $i = 1, \dots, 5$. Pour simplifier la notation, l'indice i est omis dans les formules. Il est entendu que les observations de chaque série jusqu'en décembre 2016, inclusivement, sont fondées sur l'ancien plan. À compter de janvier 2017, les observations sont fondées sur le nouveau plan. Pour chacune des trois séries, le modèle de séries chronologiques structurel de base est étendu par une intervention de niveau. Comme à la section 3, le modèle comporte une composante distincte pour le terme irrégulier de population et l'erreur d'échantillonnage :

$$\hat{\mathbf{p}}_t = \mathbf{L}_t + \mathbf{S}_t + \boldsymbol{\beta}' \mathbf{x}_t + \mathbf{I}_t + \mathbf{e}_t, \quad (4.4)$$

où $\hat{\mathbf{p}}_t = (\hat{p}_t^+, \hat{p}_t^0, \hat{p}_t^-)'$ est le vecteur des estimations directes des pourcentages jusqu'en 2016 selon l'ancien plan et à partir de 2017 selon le nouveau plan, $\mathbf{L}_t = (L_t^+, L_t^0, L_t^-)'$ est le vecteur des tendances, $\mathbf{S}_t = (S_t^+, S_t^0, S_t^-)'$ est un vecteur des tendances saisonnières, $\mathbf{x}_t = (x_t, x_t, x_t)'$ est la variable d'intervention de niveau (c'est-à-dire que x_t passe de zéro à un en janvier 2017, lors de la mise en œuvre du nouveau plan) et $\boldsymbol{\beta} = (\beta^+, \beta^0, \beta^-)'$ représente les coefficients de régression des interventions de niveau qui peuvent être interprétés comme des approximations des discontinuités. Enfin, $\mathbf{I}_t = (I_t^+, I_t^0, I_t^-)'$ est un vecteur ayant des composantes de termes irréguliers de population et $\mathbf{e}_t = (e_t^+, e_t^0, e_t^-)'$ est un vecteur ayant des erreurs d'échantillonnage. Il est entendu que les coefficients de régression utilisés pour modéliser les discontinuités sont modélisés comme des paramètres invariants dans le temps. Il est possible de les faire varier dans le temps en les modélisant comme des trajets aléatoires. Cela donnerait une autre façon de permettre des corrections variant dans le temps des discontinuités. Toutefois, il est probable que le modèle d'espace d'états ne puisse plus permettre de déterminer la tendance et le coefficient de régression variant dans le temps de la variable d'intervention.

Les variables L_t^j et S_t^j selon $j = \{+, 0, -\}$ sont des modèles de tendances lisses et des modèles saisonniers trigonométriques décrits à la section 3.1. Pour les termes de perturbation de la pente dans la tendance (3.5) et les termes de perturbation de la composante saisonnière (3.7), on suppose qu'ils sont mutuellement indépendants et qu'ils suivent une distribution normale d'espérance nulle comportant des composantes de variance indépendantes du temps. À partir de (4.2), il s'ensuit que les coefficients des discontinuités, β^+ , β^0 et β^- , doivent obéir à la restriction selon laquelle ils totalisent zéro. Cela est renforcé par les équations de transition suivantes dans le modèle d'espace d'états :

$$\begin{aligned} \beta_t^+ &= \beta_{t-1}^+ \\ \beta_t^- &= \beta_{t-1}^- \\ \beta_t^0 &= -\beta_{t-1}^+ - \beta_{t-1}^- \end{aligned} \quad (4.5)$$

L'indice inférieur t désigne la notation des équations de transition. Comme il n'y a pas de terme de perturbation, β est toujours indépendant du temps. Les termes irréguliers de population (I_t) sont modélisés comme des variables aléatoires normalement et indépendamment distribuées, c'est-à-dire $I_t^j \sim N(0, \sigma_{I,j}^2)$, $j = \{+, 0, -\}$. Les erreurs d'échantillonnage e_t sont modélisées comme des variables normalement distribuées $e_t \sim N(0, \Sigma_t)$ ayant une variance et des covariances dérivées des données d'entrée :

$$\text{Cov}(e_t^j, e_{t'}^{j'}) = \text{Cov}(\hat{p}_t^j, \hat{p}_{t'}^{j'}) = \begin{cases} \hat{p}_t^j (100 - \hat{p}_t^j) / n_t & \text{si } j = j' \text{ et } t = t' \\ -\hat{p}_t^j \hat{p}_{t'}^{j'} / n_t & \text{si } j \neq j' \text{ et } t = t'. \end{cases} \quad (4.6)$$

D'après la section 3, le fait de ne pas tenir compte de la stratification et de l'effet de l'estimateur par la régression généralisée sur l'approximation de la variance a un effet mineur sur les erreurs-types estimées. Par conséquent, les approximations de (co)variance pragmatiques ci-dessus peuvent être commodément utilisées dans l'application dont il est question.

Le modèle (4.4) à (4.6) étant utilisé pour estimer les discontinuités, seule la restriction (4.5) est appliquée aux coefficients modélisant les discontinuités. Il est toutefois entendu que la somme sur les trois séries de données d'entrée est également limitée à 100 %. Cela se traduit par des restrictions supplémentaires sur les composantes de tendance et les composantes saisonnières; par exemple, une restriction sur la somme sur les signaux $L_t^+ + L_t^0 + L_t^- + S_t^+ + S_t^0 + S_t^- = 100$ ou une restriction distincte sur la tendance $L_t^+ + L_t^0 + L_t^- = 100$ combinée à une restriction nulle sur les composantes saisonnières $L_t^+ + S_t^+ + S_t^0 + S_t^- = 0$. Différentes méthodes sont proposées dans la documentation pour imposer des restrictions sur les variables d'état variant dans le temps. Doran (1992) a proposé d'élargir l'équation de mesure du modèle d'espace d'états au moyen de séries supplémentaires imposant ces deux restrictions. Harvey et Koopman (1993) ainsi que Hungnes, Skjerpen, Hamre, Jansen, Pham et Sandvik (2024) ont montré la façon dont la contrainte selon laquelle la somme sur des variables d'état totalise zéro peut être imposée à l'aide d'une matrice de covariance à rang réduit pour les termes de perturbation d'état.

Il convient de noter que, la catégorie neutre n'étant pas utilisée dans les séries de référence $\hat{y}_{1,t}, \dots, \hat{y}_{5,t}$ ni dans les indices combinés $\hat{y}_{6,t}$, $\hat{y}_{7,t}$ et $\hat{y}_{8,t}$, une solution de rechange pragmatique est de modéliser les séries de pourcentages positifs et négatifs dans un modèle bivarié sans restriction sur les coefficients des discontinuités. Les différences entre les estimations des discontinuités fondées sur cette approche et le modèle (4.4) à (4.6) étaient inférieures à 1 point de pourcentage et également inférieures à l'erreur-type des estimations des discontinuités. La représentation de l'espace d'états du modèle (4.4) à (4.6) est fournie en annexe.

4.2 Méthodes de correction pour tenir compte des discontinuités

Comme cela est expliqué au début de la section 4, les séries de pourcentages observées avant le remaniement de janvier 2017 sont corrigées pour refléter le niveau des séries de pourcentages observées dans le cadre du nouveau plan. Ces séries de pourcentages corrigées servent à compiler les séries de données d'entrée corrigées pour le modèle de séries chronologiques (3.4) à (3.9). Ces pourcentages ne peuvent avoir que des valeurs admissibles dans la fourchette $[0, 100]$. Par conséquent, les méthodes sont considérées

comme produisant des séries corrigées dont les valeurs se situent dans cette fourchette admissible. Dans ce cas, les variables cibles corrigées $\hat{y}_1, \dots, \hat{y}_5$ auront également des valeurs dans la fourchette admissible $[-100, +100]$. Soit $\tilde{p}_{i,t}^{j,N}$, les séries corrigées de $\hat{p}_{i,t}^{j,O}$. La correction suivante est proposée pour les pourcentages :

$$\tilde{p}_{i,t}^{j,N} = \hat{p}_{i,t}^{j,O} + \hat{\beta}_i^j \frac{\hat{p}_{i,t}^{j,O} (100 - \hat{p}_{i,t}^{j,O})}{\hat{p}_{i,\tau}^{j,O} (100 - \hat{p}_{i,\tau}^{j,O})} \cdot t = 1, \dots, T-1, \quad (4.7)$$

où T est le mois du passage au nouveau plan de sondage, c'est-à-dire janvier 2017. De plus, $\hat{p}_{i,\tau}^{j,O}$ désigne une estimation selon l'ancien plan obtenue pendant toute la période de l'essai parallèle (désignée par τ). La discontinuité estimée $\hat{\beta}_i^j$ est multipliée par un facteur proportionnel à la variance du pourcentage, que l'on estime par $\hat{p}_{i,t}^{j,O} (100 - \hat{p}_{i,t}^{j,O})$. De cette façon, la taille de la correction pour tenir compte de la discontinuité varie en fonction du temps et dépend de la valeur de $\hat{p}_{i,t}^{j,O}$. La correction est nulle lorsque $\hat{p}_{i,t}^{j,O} = 0$ ou $\hat{p}_{i,t}^{j,O} = 100$, et est maximale lorsque $\hat{p}_{i,t}^{j,O} = 50$. Le dénominateur du facteur est la variance de la population du pourcentage selon l'ancien plan pendant les trois mois de l'essai parallèle et assure que l'estimation du pourcentage corrigé obtenue pour ces trois mois est proche des valeurs observées selon le nouveau plan pendant l'essai parallèle.

Lorsque les trois pourcentages $\hat{p}_i^j$ pour $j = \{+, 0, -\}$ sont corrigés avec (4.7), la somme des pourcentages corrigés n'est plus 100. Le pourcentage neutre $\hat{p}_i^0$ n'étant pas utilisé dans le calcul des indicateurs y_i , ce pourcentage est corrigé selon $\tilde{p}_{i,t}^{0,N} = 100 - \tilde{p}_{i,t}^{+,N} - \tilde{p}_{i,t}^{-,N}$. La taille de la correction dans (4.7) diminue lorsque le pourcentage $\hat{p}_{i,t}^{j,O}$ est proche de 0 ou de 100. Il n'est néanmoins pas garanti que les valeurs de $\tilde{p}_{i,t}^{j,N}$ se trouvent dans la fourchette admissible $[0, 100]$. Elles peuvent être hors de cette fourchette lorsque les pourcentages pendant l'essai parallèle $\hat{p}_{i,\tau}^{j,O}$ sont proches de 0 ou de 100 et que la discontinuité $\hat{\beta}_i^j$ est importante. Il convient de noter que l'hypothèse sous-jacente de la méthode proposée est que la discontinuité est faible lorsque la variance est faible. Cette hypothèse ne serait donc pas respectée dans ce cas.

Comme autre approche, une correction additive après l'application d'une transformation logarithmique du ratio est envisagée. Cela force la série corrigée à obéir à la restriction de ne pas prendre de valeurs hors de la fourchette admissible $[0, 100]$ et force la somme sur les trois catégories à être égale à 100. Les pourcentages neutres $\hat{p}_{i,t}^{0,d}$ pour $d = \{O, N\}$ sont utilisés comme catégorie de référence dans le dénominateur de la transformation logarithmique du ratio, que l'on définit comme suit (n'oublions pas que « O » désigne l'« ancien plan » et « N », le « nouveau plan ») :

$$z_{i,t}^{+,d} = \ln \left(\frac{\hat{p}_{i,t}^{+,d}}{\hat{p}_{i,t}^{0,d}} \right), \quad z_{i,t}^{-,d} = \ln \left(\frac{\hat{p}_{i,t}^{-,d}}{\hat{p}_{i,t}^{0,d}} \right), \quad \text{pour } d = \{O, N\}. \quad (4.8)$$

À l'étape suivante, les séries observées selon l'ancien plan sont corrigées pour refléter le niveau selon le nouveau plan à l'aide de la correction additive suivante de la série transformée par le logarithme du ratio :

$$\tilde{z}_{i,t}^{+,N} = \ln \left(\frac{\hat{p}_{i,t}^{+,O}}{\hat{p}_{i,t}^{0,O}} \right) + \ln \left(\frac{\hat{\beta}_i^+}{\hat{\beta}_i^0} \right), \quad \tilde{z}_{i,t}^{-,N} = \ln \left(\frac{\hat{p}_{i,t}^{-,O}}{\hat{p}_{i,t}^{0,O}} \right) + \ln \left(\frac{\hat{\beta}_i^-}{\hat{\beta}_i^0} \right), \quad \text{pour } t = 1, \dots, T-1. \quad (4.9)$$

Dans (4.9), $\hat{\beta}_i^+$, $\hat{\beta}_i^0$ et $\hat{\beta}_i^-$ désignent les estimations du filtre de Kalman pour les discontinuités obtenues avec le modèle (4.4) à (4.6) à l'aide des renseignements tirés de l'essai parallèle par l'initialisation informative du filtre de Kalman, comme cela est décrit à la section 4.1. Par la suite, la transformation inverse du logarithme du ratio

$$\tilde{p}_{i,t}^{+,N} = 100 \frac{e^{\tilde{z}_{i,t}^{+,N}}}{e^{\tilde{z}_{i,t}^+} + e^{\tilde{z}_{i,t}^-} + 1}, \quad \tilde{p}_{i,t}^{-,N} = 100 \frac{e^{\tilde{z}_{i,t}^{-,N}}}{e^{\tilde{z}_{i,t}^+} + e^{\tilde{z}_{i,t}^-} + 1}, \quad \tilde{p}_{i,t}^0 = 100 \frac{1}{e^{\tilde{z}_{i,t}^+} + e^{\tilde{z}_{i,t}^-} + 1}, \quad (4.10)$$

pour $t=1, \dots, T-1$, est appliquée pour obtenir trois séries corrigées pour les pourcentages respectant la restriction de prendre des valeurs dans la fourchette admissible $[0, 100]$ et de totaliser 100 %. L'un des inconvénients de la transformation logarithmique du ratio est que l'effet de la correction peut devenir très important lorsque les ratios $\hat{p}_{i,t}^{+,0} / \hat{p}_{i,t}^{0,0}$ ou $\hat{p}_{i,t}^{-,0} / \hat{p}_{i,t}^{0,0}$ sont inférieurs à 1. Cela sera démontré à la section 4.4.

4.3 Résultats de l'estimation des discontinuités

Le tableau 4.1 présente les estimations des discontinuités obtenues avec l'essai parallèle pour les cinq variables. Au cours de la période, la plupart des répondants étaient positifs quant à la situation économique. Pour les variables « Econ. 12P », « Econ. 12S », « Fin. 12P » et « Fin. 12S », le pourcentage lié à l'option de réponse « un peu mieux » a augmenté de manière substantielle selon le nouveau plan. De plus, le pourcentage lié à l'option de réponse « un peu moins bonne » a augmenté, mais dans une moindre mesure (données non présentées dans le tableau 4.1). Les pourcentages liés aux autres options de réponse ont tous diminué. Ces changements peuvent s'expliquer en partie par les modifications apportées au questionnaire. Selon l'ancien plan, le répondant pouvait uniquement choisir entre « mieux » et « pire », en plus des options neutres. En cas d'une situation ne changeant que légèrement, les options « mieux » ou « pire » ne semblaient probablement pas appropriées au répondant, qui ne savait pas que les réponses « un peu moins bonne » ou « un peu mieux » étaient également possibles. Le répondant a alors probablement choisi la catégorie de réponses neutres plus souvent. Selon le nouveau plan, le répondant choisit « un peu moins bonne » ou « un peu mieux » et, pendant l'essai parallèle, il choisit principalement « un peu mieux ». La question sur les achats importants (seule question pour laquelle le questionnaire n'a pas été modifié) est la seule question pour laquelle la discontinuité des réponses positives est négative et inférieure à celle des autres questions.

Tableau 4.1

Estimations des discontinuités obtenues avec l'essai parallèle, erreurs-types entre parenthèses

	Réponses positives ($\hat{\Delta}(p_i^+)$)		Réponses négatives ($\hat{\Delta}(p_i^-)$)		Indicateur ($\hat{\Delta}(p_i^+) - \hat{\Delta}(p_i^-)$)
Econ. 12P	11,4	(1,3)	-0,1	(0,9)	11,5
Econ. 12S	16,8	(1,2)	0,1	(0,8)	16,7
Fin. 12P	8,7	(1,0)	3,2	(1,1)	5,5
Fin. 12S	11,1	(1,0)	3,1	(0,9)	8,0
Achat imp.	-4,7	(1,2)	0,1	(0,8)	-4,8

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

Le modèle (4.4) à (4.6) est appliqué pour estimer les discontinuités. Les séries commencent en avril 1986 et se terminent en février 2020. Jusqu'à décembre 2016, inclusivement, les estimations sont fondées sur l'ancien plan et à compter de janvier 2017, elles sont fondées sur le nouveau plan. Les estimations des discontinuités sont présentées au tableau 4.2.

Tableau 4.2

Estimations lissées des discontinuités fondées sur le modèle de séries chronologiques structurel selon le modèle (4.4) à (4.6), erreurs-types entre parenthèses

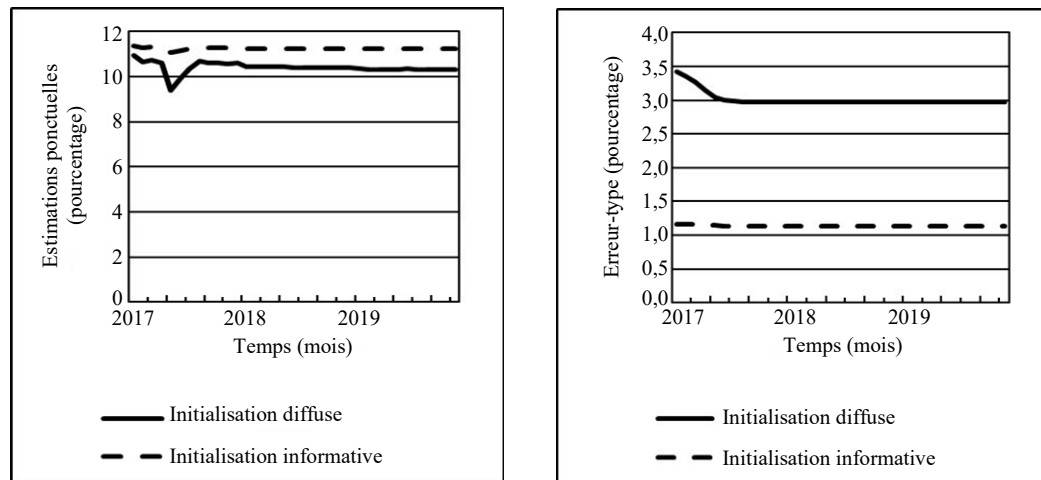
	Initialisation diffuse du filtre de Kalman				Initialisation informative du filtre de Kalman			
	réponses positives		réponses négatives		réponses positives		réponses négatives	
Econ. 12P	10,7	(3,0)	0,1	(2,5)	11,2	(1,1)	-0,1	(0,8)
Econ. 12S	20,1	(3,3)	-0,8	(3,1)	17,2	(1,1)	0,2	(0,7)
Fin. 12P	9,7	(1,2)	2,1	(1,5)	9,1	(0,8)	3,0	(0,9)
Fin. 12S	12,2	(1,2)	4,6	(1,5)	11,7	(0,7)	3,5	(0,8)
Achat imp.	-6,5	(1,8)	1,3	(1,5)	-5,1	(1,0)	0,3	(0,7)

Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents. La variable « Econ. 12S » se rapporte aux attentes de changements de la situation économique générale du pays au cours des 12 mois suivants. La variable « Fin. 12P » se rapporte à l'opinion sur les changements de la situation financière du ménage au cours des 12 mois précédents. La variable « Fin. 12S » se rapporte aux attentes de changements de la situation financière du ménage au cours des 12 mois suivants. La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

Les estimations obtenues au moyen du lisseur de Kalman dans le tableau 4.2 sont comparables aux estimations directes fondées sur l'essai parallèle du tableau 4.1. Il est particulièrement agréable de constater que les estimations obtenues au moyen de l'essai parallèle correspondent aux estimations du filtre de Kalman à initialisation diffuse, qui ne repose pas sur les résultats de l'essai parallèle. Par conséquent, les estimations ponctuelles obtenues à l'aide d'une distribution *a priori* diffuse et une distribution *a priori* informée s'équivalent bien. Si les renseignements supplémentaires de l'essai parallèle sont compris dans le filtre de Kalman par une initialisation informée, les erreurs-types des estimations des discontinuités diminuent clairement. Dans le cadre de l'initialisation diffuse, les erreurs-types des estimations des discontinuités sont substantiellement supérieures aux erreurs-types des estimations directes obtenues lors de l'essai parallèle. Dans cette situation, il n'y a pas de contrôle sur la précision de la discontinuité, puisque la précision dépend de l'instabilité des séries chronologiques disponibles. Cela est contraire à l'essai parallèle, pour lequel la précision peut être contrôlée en calculant la taille d'échantillon minimale requise pour observer une différence prédéterminée à un niveau de puissance et de signification prédéterminé. Lorsque les résultats de l'essai parallèle sont combinés à l'approche de modélisation de séries chronologiques au moyen d'une initialisation informative du filtre de Kalman, les estimations les plus précises des discontinuités sont obtenues, puisque tous les renseignements disponibles provenant des séries chronologiques observées avant et après le changement et l'essai parallèle sont combinés. L'amélioration de la précision par rapport au modèle de séries chronologiques à initialisation diffuse est substantielle. En ce qui concerne les estimations directes de l'essai parallèle, il n'y a qu'une légère amélioration de la précision des estimations des discontinuités dans l'application dont il est question. Cela pourrait toutefois être différent dans d'autres applications.

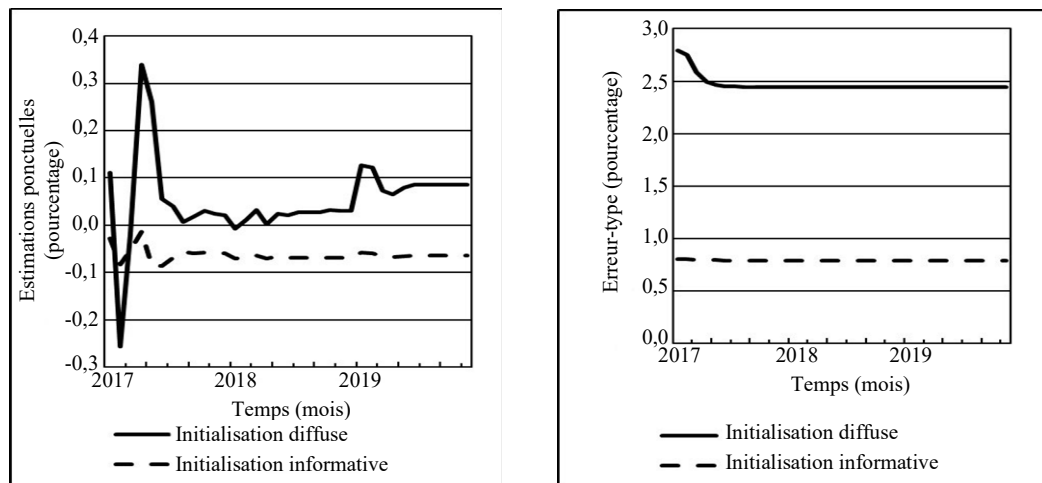
Les estimations des discontinuités obtenues à l'aide du modèle de séries chronologiques sont instables si seules quelques observations après le changement sont disponibles. Elles s'améliorent lorsque plus de données dans le cadre du nouveau plan deviennent disponibles. Les figures 4.1 et 4.2 présentent les estimations en temps réel des discontinuités fondées sur les données de janvier 2017 à février 2020 pour les pourcentages des réponses respectivement positives et négatives pour la variable « Econ. 12P ». Les chiffres montrent l'évolution des estimations des discontinuités lorsque plus de données deviennent disponibles. Les estimations initiales, directement après le passage au nouveau plan, ont le bon ordre d'ampleur. Néanmoins, certains changements sont visibles au cours des six premiers mois. Dans le cas de l'initialisation informative, les changements sont beaucoup plus petits. Au cours des six premiers mois, les erreurs-types des estimations diminuent, lorsque la diminution pour l'initialisation diffuse est substantielle. Après cette période, les estimations ponctuelles et les erreurs-types se stabilisent à des valeurs constantes. Il est difficile de déterminer en général le nombre d'observations requises selon le nouveau plan avant d'obtenir une estimation stable des discontinuités, puisque cela dépend de la souplesse de la tendance, c'est-à-dire la mesure dans laquelle la tendance varie au fil du temps. À mesure que la souplesse de la tendance augmente, les estimations des discontinuités deviennent plus dépendantes des observations locales proches de la période de passage au nouveau plan d'enquête. À mesure que la tendance devient invariante dans le temps, l'influence des observations plus éloignées de la période de changement augmente.

Figure 4.1 Évolution des estimations ponctuelles (partie de gauche) et des erreurs-types (partie de droite) des discontinuités des pourcentages de réponses positives, variable « Econ. 12P »



Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents.

Figure 4.2 Évolution des estimations ponctuelles (partie de gauche) et des erreurs-types (partie de droite) des discontinuités des pourcentages de réponses négatives, variable « Econ. 12P »



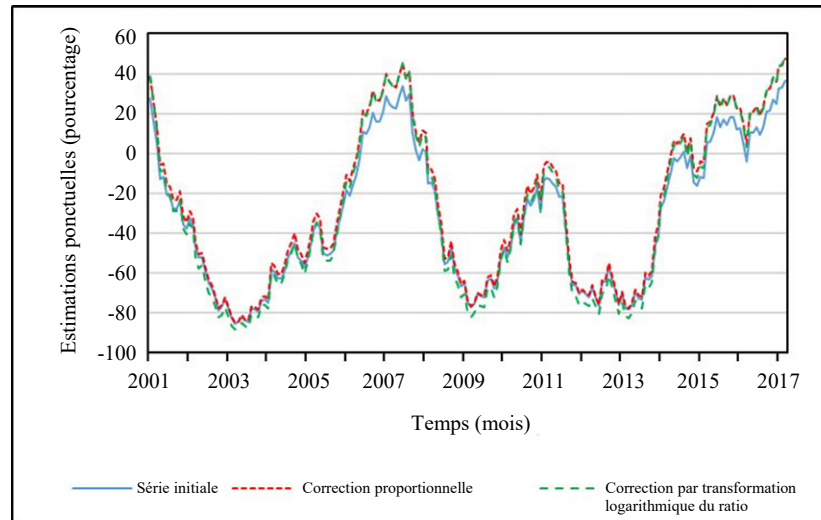
Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents.

4.4 Résultats pour les séries corrigés pour tenir compte des discontinuités

Dans la présente section, les effets des méthodes de correction des discontinuités décrites à la section 4.2 sur les séries d'indicateurs de l'Enquête sur la confiance des consommateurs sont étudiés pour les variables « Econ. 12P » et « Achat imp. ». Les estimations des discontinuités sont fondées sur le modèle de séries chronologiques (4.4) à (4.6) et une initialisation informative du filtre de Kalman. Ces estimations des discontinuités sont utilisées en vue de calculer des séries corrigées pour refléter le niveau du nouveau plan pour les pourcentages $p_{i,t}^+$, $p_{i,t}^0$, $p_{i,t}^-$ à l'aide de la correction proportionnelle définie par (4.7) et de la correction additive en combinaison avec la transformation logarithmique du ratio définie par (4.9) et (4.10). Ces séries corrigées servent à calculer les séries ininterrompues pour $y_{1,t}, \dots, y_{5,t}$, qui sont à leur tour les données d'entrée du modèle 1 de (3.4) à (3.9).

La figure 4.3 présente la série d'indicateur selon l'ancien plan pour la variable « Econ. 12P » de janvier 2001 à mars 2017 ainsi que les séries corrigées à l'aide des deux méthodes pour tenir compte des discontinuités. Les discontinuités de cette variable sont $\hat{\Delta}(p_1^+) = 11,2$ et $\hat{\Delta}(p_1^-) = -0,1$ (tableau 4.2) au cours d'une période où la confiance des consommateurs est positive. La figure 4.3 montre que pendant les périodes de valeurs positives pour la variable « Econ. 12P », les deux méthodes permettent de corriger la série à peu près de la même façon, c'est-à-dire que les séries corrigées deviennent plus positives et que les corrections sont égales. Pour les périodes de valeurs négatives pour la variable « Econ. 12P », la transformation logarithmique du ratio rend la série corrigée plus négative, tandis que la série corrigée proportionnellement demeure proche de la série initiale. Cette correction plus importante selon la transformation logarithmique du ratio dans la direction opposée que celle observée pendant l'essai parallèle semble moins plausible que la correction proportionnelle légère.

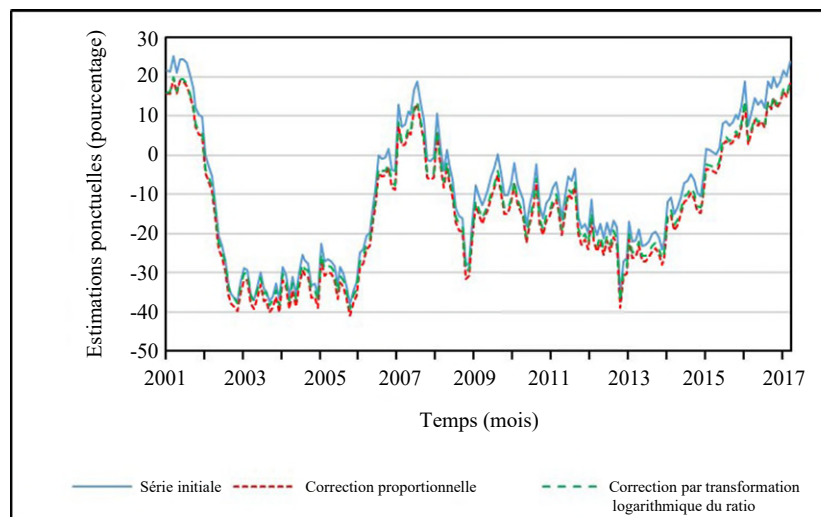
Figure 4.3 Comparaison des méthodes de correction pour l'indicateur « Econ. 12P », estimations des discontinuités fondées sur un modèle de séries chronologiques structurel à initialisation informative



Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents.

La figure 4.4 présente la série initiale et les séries corrigées pour « Achat imp. ». Pour cette variable, les discontinuités sont plutôt faibles, mais comparativement à la variable « Econ. 12P », elles sont de signes opposés : $\hat{\Delta}(p_5^-) = -5,1$ et $\hat{\Delta}(p_5^+) = 0,3$. Pour « Achat imp. », les deux méthodes donnent des résultats semblables : les valeurs des séries corrigées sont inférieures à celles de la série initiale et l'effet des deux corrections est le même. Pour cette variable, il n'y a pas de préférence pour l'une des deux méthodes.

Figure 4.4 Comparaison des méthodes de correction pour l'indicateur « Achat imp. », estimations des discontinuités fondées sur un modèle de séries chronologiques structurel à initialisation informative

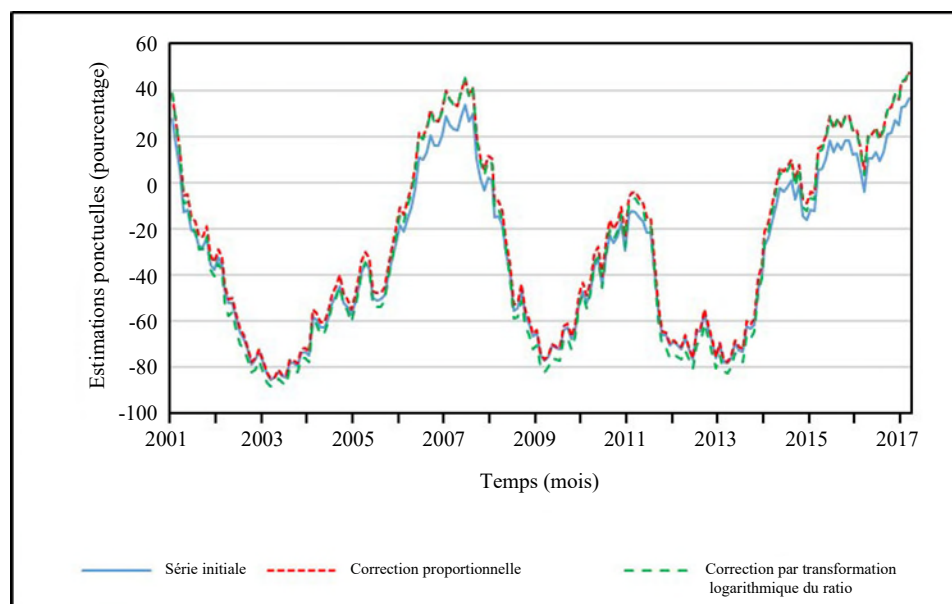


Note : La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

À partir de ces résultats, une légère préférence est accordée à la correction proportionnelle, puisque la transformation logarithmique du ratio entraîne parfois des corrections plus importantes dans une direction différente que celle observée pendant l'essai parallèle. Toutefois, un inconvénient de la méthode proportionnelle est que les valeurs corrigées pourraient se situer hors de la fourchette $[-100, +100]$, mais cela ne se produit que pour de grandes valeurs irréalistes des discontinuités. Dans la figure 4.5, cela est simulé à l'aide de valeurs choisies arbitrairement pour les discontinuités de la variable « Econ. 12P », à savoir $\hat{\Delta}(p_5^+) = -10$ et $\hat{\Delta}(p_1^-) = 20$, pour que les corrections proportionnelles mènent à des résultats inférieurs à -100 pour les périodes les plus négatives. Dans ce cas-ci, la transformation logarithmique du ratio pour corriger la série serait préférable.

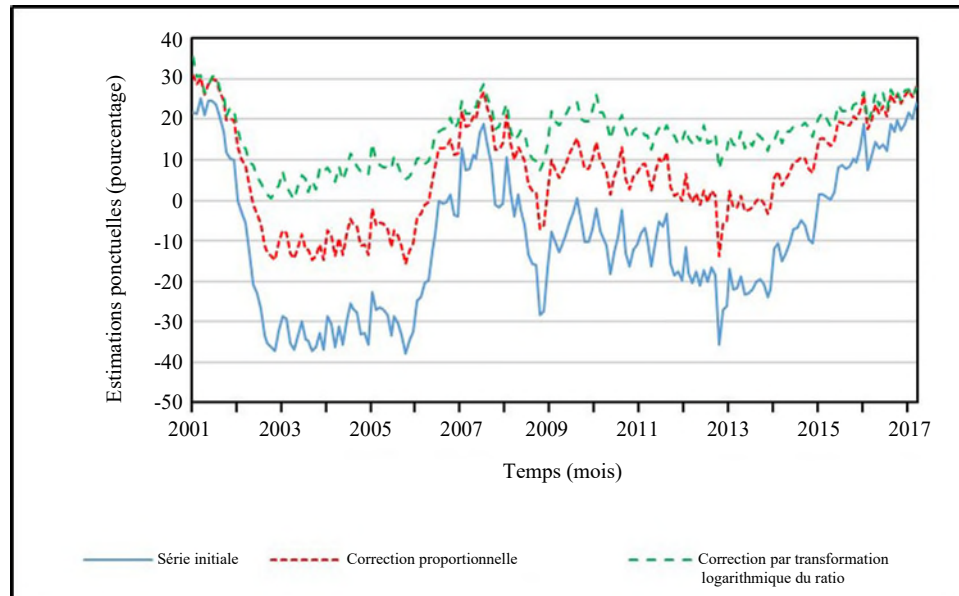
La figure 4.6 montre qu'en présence également de valeurs réalistes des discontinuités, la transformation logarithmique du ratio pourrait entraîner d'importantes corrections. Dans ce cas-ci, les deux méthodes sont appliquées à la série d'indicateur de la variable « Achat imp. » avec des valeurs choisies arbitrairement, mais réalistes, pour des discontinuités de $\hat{\Delta}(p_5^+) = -5$ et de $\hat{\Delta}(p_5^-) = -10$. Pour certaines périodes, la correction fondée sur la transformation logarithmique du ratio est faible (par exemple vers 2007), mais extrêmement importante pour d'autres périodes (par exemple 2003 à 2006 et 2009 à 2015). De plus, la série corrigée par transformation logarithmique du ratio est toujours positive, même pour les périodes pour lesquelles la série initiale est négative. Cela illustre l'inconvénient de la transformation logarithmique du ratio, à savoir que la correction peut devenir très importante lorsque le ratio de la valeur initiale est beaucoup plus petit que un. Cet effet est également visible pour la correction proportionnelle, mais dans une moindre mesure. À partir de ces résultats, on conclut d'utiliser les corrections proportionnelles pour les séries de données d'entrée $\hat{y}_1, \dots, \hat{y}_5$.

Figure 4.5 Comparaison des méthodes de correction pour l'indicateur « Econ. 12P » et des valeurs de discontinuités choisies arbitrairement, à savoir $\hat{\Delta}(p_1^+) = -10$ et $\hat{\Delta}(p_1^-) = 20$



Note : La variable « Econ. 12P » se rapporte à l'opinion sur les changements de la situation économique générale du pays au cours des 12 mois précédents.

Figure 4.6 Comparaison des méthodes de correction pour l'indicateur « Achat imp. » et des valeurs de discontinuités choisies arbitrairement, à savoir $\hat{\Delta}(p_s^+) = -5$ et $\hat{\Delta}(p_s^-) = -10$



Note : La variable « Achat imp. » se rapporte à la question de savoir s'il s'agit du bon moment pour les gens de faire des achats importants.

5. Publications officielles fondées sur la modélisation de séries chronologiques structurelle

Depuis 2017, le modèle de séries chronologiques (3.4) à (3.9) est mis en œuvre dans le processus de production de l'Enquête sur la confiance des consommateurs des Pays-Bas. Les séries de référence $\hat{y}_{1,t}, \dots, \hat{y}_{5,t}$ sont les séries de données d'entrée du modèle. Le modèle utilisé pour la production équivaut au modèle 1 en supposant des corrélations nulles pour les erreurs d'échantillonnage. (La prise en compte de la corrélation entre les erreurs d'échantillonnage est fondée sur les commentaires des examinateurs, après la mise en œuvre de la méthodologie en production.) Ces séries observées pour les années jusqu'à décembre 2016 sont corrigées pour refléter le niveau du nouveau plan en suivant l'approche décrite à la section 4.2. Les estimations des tendances filtrées sont publiées et remplacent les anciennes données désaisonnalisées. Ces séries d'estimations filtrées, corrigées pour tenir compte de la discontinuité, commencent en 1986.

À la figure 5.1, les estimations directes initiales et les estimations directes corrigées pour tenir compte de la discontinuité pour la confiance des consommateurs sont comparées aux estimations des tendances filtrées et lissées obtenues à l'aide du modèle de séries chronologiques (3.4) à (3.9). Les estimations directes observées selon l'ancien plan (ligne pointillée) sont représentées jusqu'en décembre 2016. La série corrigée de la confiance des consommateurs (ligne rouge continue) se poursuit avec les estimations directes observées selon le nouveau plan après janvier 2017. La série corrigée pour la période jusqu'à décembre 2016 est obtenue comme la moyenne sur les cinq séries de référence corrigées $\hat{y}_{1,t}, \dots, \hat{y}_{5,t}$. Ainsi, une série ininterrompue d'estimations directes de la confiance des consommateurs sur une base mensuelle est obtenue. La tendance filtrée (ligne bleue continue) remplace les chiffres désaisonnalisés figurant dans les

publications officielles de l'Enquête sur la confiance des consommateurs des Pays-Bas et peut être interprétée comme une composante tendance-cycle obtenue en supprimant les fluctuations saisonnières, le terme irrégulier de population et l'erreur d'échantillonnage de la série d'estimations directes. Pour la période précédant janvier 2017, la tendance est également corrigée pour refléter le niveau observé selon le nouveau plan, puisque le modèle de séries chronologiques (3.4) à (3.9) repose sur les cinq indices de référence corrigés comme séries de données d'entrée. Par conséquent, des séries de tendances ininterrompues sont obtenues pour la période commençant en 1986. La tendance lissée de la partie inférieure présente les estimations plus stables et optimales de la confiance des consommateurs, puisque les estimations des tendances pour chaque période sont fondées sur toutes les observations disponibles.

Les résultats de la figure 5.1 sont présentés jusqu'en décembre 2019. La pandémie de COVID-19, qui a commencé au premier trimestre de 2020, a entraîné une baisse spectaculaire de la confiance des consommateurs. Pour éviter une mauvaise spécification du modèle temporel, il a été nécessaire d'accroître la flexibilité de la tendance en faisant varier la variance des termes de perturbation de la pente dans le temps dans (3.6) selon une approche pareille à celle décrite dans van den Brakel, Souren et Krieg (2022). Les renseignements sur cette approche et d'autres méthodes envisagées pour tenir compte, dans le modèle, des changements soudains de la dynamique des séries sous-jacentes se trouvent dans van den Brakel, Krieg et Smeets (2023).

Figure 5.1 Comparaison des estimations directes (initiales et corrigées) avec les estimations de tendances filtrées publiées pour la confiance des consommateurs (partie supérieure) et les estimations de tendances lisses (partie inférieure)

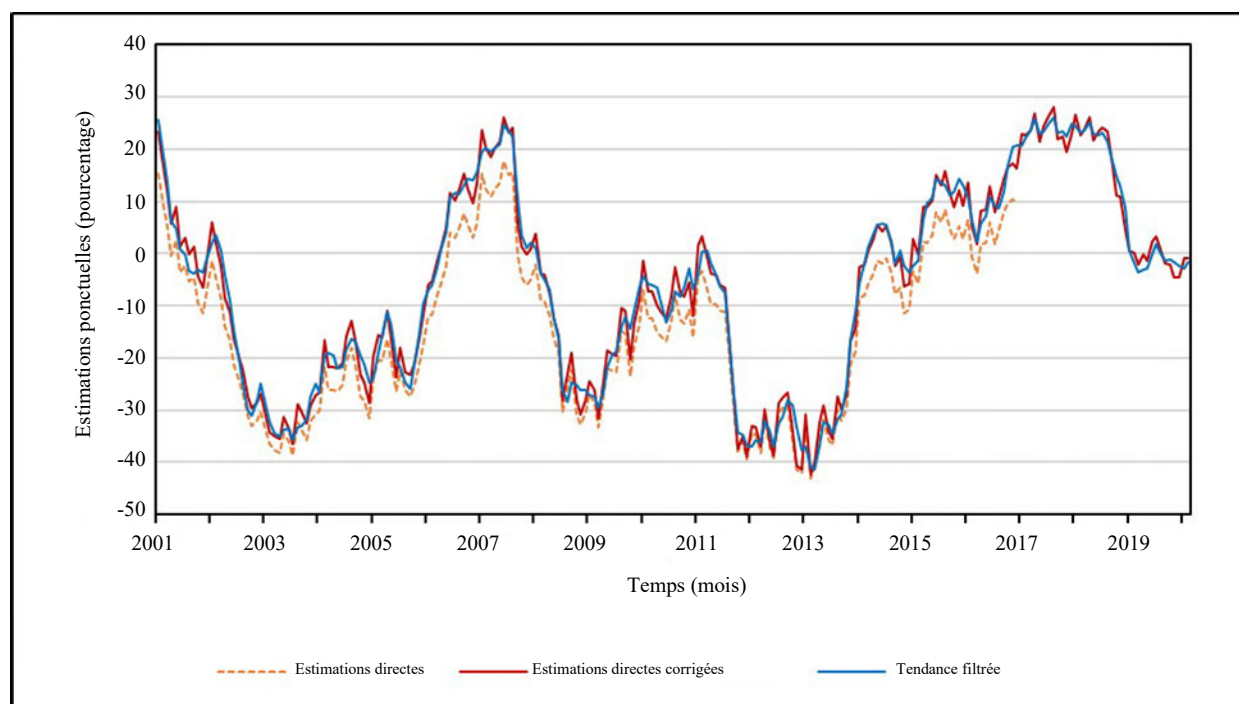
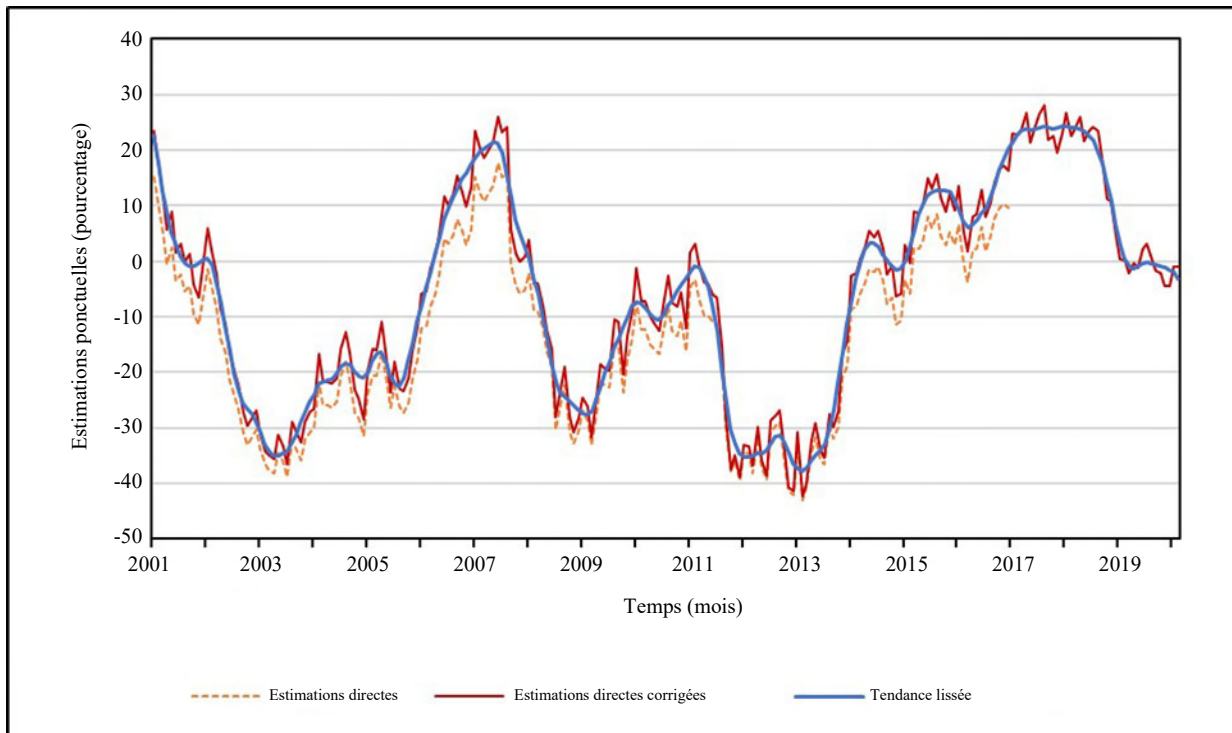


Figure 5.1 (suite) Comparaison des estimations directes (initiales et corrigées) avec les estimations de tendances filtrées publiées pour la confiance des consommateurs (partie supérieure) et les estimations de tendances lisses (partie inférieure)



6. Discussion

Dans la présente étude, une méthode d'inférence basée sur un modèle pour l'Enquête sur la confiance des consommateurs des Pays-Bas est élaborée. Cette méthode est fondée sur un modèle d'équation de séries chronologiques apparemment non liées dans le cadre duquel les estimations directes mensuelles de cinq indices de référence pour la confiance des consommateurs sont utilisées comme données d'entrée. Ces indices sont obtenus à partir de la différence entre le pourcentage de répondants ayant une opinion positive et une opinion négative sur cinq sujets relatifs à la situation économique et financière. À partir des cinq indices de référence, trois indices composites sont calculés. L'un d'eux est la confiance des consommateurs, qui est la moyenne sur les cinq indices de référence. Le modèle d'équation de séries chronologiques apparemment non liées est élaboré pour produire des indicateurs mensuels de la confiance des consommateurs plus précis ainsi que pour estimer et corriger les discontinuités causées par le passage à un nouveau plan d'enquête en janvier 2017.

Les séries chronologiques du modèle d'équation de séries chronologiques apparemment non liées sélectionné sont corrélées simultanément par les termes de perturbation de la pente de la tendance, les termes irréguliers de la population et les erreurs d'échantillonnage. Les estimations basées sur un modèle, y compris les erreurs-types pour les trois indicateurs composites, sont obtenues en calculant la moyenne sur la tendance

et le signal (tendance plus saisonnalité) des indicateurs de référence pertinents. Les tendances et les signaux filtrés sont plutôt instables dans l'application en question. Des estimations plus précises et stables sont obtenues à l'aide du lisseur de Kalman. Toutefois, ces dernières ne sont pas publiées par Statistique Pays-Bas, car cela exige une révision continue des chiffres récemment publiés.

Le remaniement de l'enquête en 2017 a entraîné des discontinuités dans la série de l'Enquête sur la confiance des consommateurs des Pays-Bas. Pour séparer les variations réelles d'un mois à l'autre des différences soudaines de mesure et de biais de sélection attribuables au remaniement, on estime les discontinuités en recueillant en parallèle des données selon l'ancien et le nouveau plan au cours du premier trimestre de 2017. La précision de ces estimations peut être améliorée au moyen d'un modèle de séries chronologiques dans le cadre duquel la discontinuité est modélisée avec un changement de niveau. Une initialisation informative est appliquée au coefficient de régression de ce changement de niveau dans le filtre de Kalman à l'aide des estimations directes et de leurs erreurs-types obtenues au cours de l'essai parallèle. De cette façon, les renseignements observés dans les séries chronologiques avant et après le changement sont utilisés pour améliorer encore les estimations directes des discontinuités.

Des discontinuités figurent dans la répartition en pourcentage entre les catégories de réponses positives, neutres et négatives obtenues pour les questions sur la situation économique, la situation financière et les achats importants. La façon dont les discontinuités dans les pourcentages mènent à des discontinuités dans les cinq indices de référence dépend de la répartition en pourcentage entre ces trois catégories. Comme ces répartitions varient considérablement au fil du temps, une méthode de correction variant dans le temps pour tenir compte des discontinuités est proposée. Pour ce faire, on modélise les discontinuités dans les séries chronologiques des pourcentages des catégories de réponses positives, neutres et négatives. Pour conserver des séries ininterrompues, les séries chronologiques de ces pourcentages observés dans le cadre de l'ancien plan sont corrigées pour refléter le niveau du nouveau plan. Une méthode de correction proposée tient compte du fait que les valeurs des proportions corrigées se situent dans la fourchette [0, 100]. Selon cette méthode, la taille de la correction pour chaque période est rendue proportionnelle à la variance de la proportion estimée. Une transformation logarithmique du ratio ou une transformation logit est également envisagée. L'inconvénient de la transformation logarithmique du ratio est que les corrections peuvent être extrêmement importantes au cours de périodes où la valeur du ratio initial est inférieure à un. Une autre solution serait la fonction de lien probit, qui reste à approfondir.

Les séries corrigées des cinq indices de référence sont dérivées des séries corrigées des pourcentages des catégories de réponses positives, neutres et négatives. Les corrections apportées aux cinq séries de référence varient dans le temps, puisque la répartition des répondants entre les catégories de réponses positives, neutres et négatives varie au fil du temps et que les corrections apportées aux séries de ces pourcentages varient dans le temps également. Ces séries corrigées sont prolongées au moyen des estimations directes observées après avril 2017 et sont utilisées comme séries de données d'entrée dans le modèle de séries chronologiques multivarié afin d'estimer les tendances mensuelles pour l'Enquête sur la confiance des consommateurs. Cette méthode a été mise en œuvre en avril 2017 pour la publication de chiffres et de

tendances mensuels ininterrompus de l'Enquête sur la confiance des consommateurs, qui a commencé en 1986. Avec l'introduction du nouveau plan, les séries ne sont révisées qu'une seule fois, ce qui permet aux utilisateurs de données d'avoir une longue série chronologique ininterrompue. Les estimations des tendances remplacent les chiffres désaisonnalisés et peuvent être interprétées comme une tendance-cycle découlant des estimations directes dont la composante saisonnière, la composante irrégulière de population et l'erreur d'échantillonnage sont supprimées.

Remerciements

Les auteurs tiennent à remercier deux examinateurs anonymes et un réviseur associé d'avoir lu attentivement une précédente ébauche de la présente étude et d'avoir fourni des commentaires utiles.

Annexe

Estimateur par la régression généralisée pour l'Enquête sur la confiance des consommateurs

Les paramètres de l'Enquête sur la confiance des consommateurs sont définis comme des moyennes. Soit y , la moyenne de population inconnue d'un paramètre d'intérêt. Soit y_k , l'observation obtenue du répondant k pour le paramètre y , $\mathbf{x}_k = (x_{k,1}, \dots, x_{k,l})'$, un vecteur de dimension l comportant des variables auxiliaires pour le répondant k et n , le nombre de répondants observés dans l'échantillon. De plus, $\mathbf{x}$ désigne un vecteur de dimension l avec les moyennes de population réelles des variables auxiliaires $\mathbf{x}_k$. L'estimateur par la régression généralisée est fondé sur le modèle de régression linéaire suivant : $y_k = \mathbf{b}'\mathbf{x}_k + \varepsilon_k$, où $\mathbf{b}$ est un vecteur de dimension l avec des coefficients de régression et ε_k est des résidus distribués indépendamment avec $E(\varepsilon_k) = 0$ et $\text{Var}(\varepsilon_k) = \sigma^2$. L'estimateur par la régression généralisée pour y est alors défini comme

$$\hat{y}^r = \hat{y}^d + \hat{\mathbf{b}}'(\mathbf{x} - \hat{\mathbf{x}}^d),$$

où $\hat{y}^d$ et $\hat{\mathbf{x}}^d$ sont les estimations d'échantillon non pondérées pour les moyennes de population de y et les variables auxiliaires $\mathbf{x}$, c'est-à-dire $\hat{y}^d = \frac{1}{n} \sum_{k=1}^n y_k$ et $\hat{\mathbf{x}}^d = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k$. Par ailleurs, un estimateur de $\mathbf{b}$ est donné par $\hat{\mathbf{b}} = \left(\sum_{k=1}^n \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_{k=1}^n \mathbf{x}_k y_k$. La variable auxiliaire utilisée dans l'estimateur par la régression généralisée pour l'Enquête sur la confiance des consommateurs est définie par le système de pondération suivant :

Revenu(5) * Comp.Ménage(2) + Urbanisation(5) * Comp.Ménage(2) + TailleMénage(5) +
PropriétéLogement(2) * Comp.Ménage(2) + Âge(7) + NiveauScolarité(4) + TypeMénage(4) + Sexe(2) *
Comp.Ménage(2)

où

- *Revenu(5)* : revenu disponible du ménage auquel la personne appartient classé dans cinq quantiles, c'est-à-dire tranche inférieure de 20 %, 20 % à 40 %, 40 % à 60 %, 60 % à 80 % ou tranche supérieure de 20 %.
- *Comp.Ménage(2)* : chef du ménage célibataire ou en couple.
- *Urbanisation(5)* : niveau d'urbanisation de la municipalité classé dans cinq catégories, c'est-à-dire très urbaine, urbaine, modérément urbaine, peu urbaine ou rurale.
- *TailleMénage(5)* : ménage comptant une, deux, trois, quatre ou cinq personnes ou plus.
- *PropriétéLogement(2)* : personne vivant dans un logement loué ou un logement dont il est propriétaire.
- *Âge(7)* : âge selon sept catégories, c'est-à-dire 15 à 24 ans, 25 à 34 ans, 35 à 44 ans, 45 à 54 ans, 55 à 64 ans, 65 à 74 ans ou 75 ans et plus.
- *NiveauScolarité(4)* : niveau de scolarité selon quatre catégories, c'est-à-dire faible, moyen, élevé ou inconnu.
- *TypeMénage(4)* : type de ménage auquel une personne appartient selon quatre catégories, c'est-à-dire couple sans enfants, couple ayant des enfants, ménage monoparental ou autre ménage y compris les célibataires.
- *Sexe(2)* : sexe selon deux catégories, c'est-à-dire hommes ou femmes.

Calcul de (3.10)

Le calcul est fourni pour une fraction. Il est entendu que cela s'applique également aux paramètres définis comme des pourcentages. Pour simplifier la notation, l'indice inférieur t est supprimé dans les calculs suivants. Puisque $\hat{y}_i = \hat{p}_i^+ - \hat{p}_i^-$, il s'ensuit que

$$\text{Var}(\hat{y}_i) = \text{Var}(\hat{p}_i^+) + \text{Var}(\hat{p}_i^-) - 2\text{Cov}(\hat{p}_i^+, \hat{p}_i^-). \quad (\text{A.1})$$

Un estimateur de la variance selon un échantillonnage aléatoire simple pour une moyenne d'échantillon est obtenu par

$$\text{Var}(\hat{p}_i^+) = \frac{1}{n} \frac{1}{n-1} \sum_{k=1}^n (y_{i,k}^+ - \bar{y}_i^+)^2.$$

Puisque $y_{i,k}^+ = 1$ si le répondant k répond positivement à la question i et $y_{i,k}^+ = 0$ sinon, il s'ensuit que $y_{i,k}^{+2} = y_{i,k}^+$, $\bar{y}_i^+ = \hat{p}_i^+$ et

$$\text{Var}(\hat{p}_i^+) = \frac{1}{n} \frac{1}{n-1} \sum_{k=1}^n (y_{i,k}^+ - 2y_{i,k}^+ \hat{p}_i^+ + \hat{p}_i^{+2}) = \frac{1}{n} \frac{1}{n-1} n \hat{p}_i^+ (1 - \hat{p}_i^+) \approx \frac{1}{n} \hat{p}_i^+ (1 - \hat{p}_i^+).$$

Dans un cas semblable, il s'ensuit que $\text{Var}(\hat{p}_i^-) \approx \frac{1}{n} \hat{p}_i^- (1 - \hat{p}_i^-)$. Un estimateur de la covariance selon l'échantillonnage aléatoire simple de deux moyennes d'échantillon est obtenu par

$$\text{Cov}(\hat{p}_i^+, \hat{p}_i^-) = \frac{1}{n} \frac{1}{n-1} \sum_{k=1}^n (y_{i,k}^+ - \bar{y}_i^+) (y_{i,k}^- - \bar{y}_i^-).$$

Si le répondant k répond positivement à la question i , alors $y_{i,k}^+ = 1$ et $y_{i,k}^- = 0$. Si le répondant k répond négativement à la question i , alors $y_{i,k}^+ = 0$ et $y_{i,k}^- = 1$. Si le répondant k répond de manière neutre à la question i , alors $y_{i,k}^+ = 0$ et $y_{i,k}^- = 0$. Par conséquent, $y_{i,k}^+ y_{i,k}^- = 0$ pour tous les k et

$$\begin{aligned} \text{Cov}(\hat{p}_i^+, \hat{p}_i^-) &= \frac{1}{n} \frac{1}{n-1} \sum_{k=1}^n (y_{i,k}^+ y_{i,k}^- - y_{i,k}^+ \bar{y}_i^- - y_{i,k}^- \bar{y}_i^+ + \bar{y}_i^+ \bar{y}_i^-) \\ &= \frac{1}{n} \frac{1}{n-1} (-2n \bar{y}_i^+ \bar{y}_i^- + n \bar{y}_i^+ \bar{y}_i^-) \\ &= -\frac{1}{n} \frac{1}{n-1} n \bar{y}_i^+ \bar{y}_i^- \approx -\frac{1}{n} \hat{p}_i^+ \hat{p}_i^-. \end{aligned}$$

Insérer les résultats pour $\text{Var}(\hat{p}_i^-)$, $\text{Var}(\hat{p}_i^+)$ et $\text{Cov}(\hat{p}_i^+, \hat{p}_i^-)$ donne (3.10).

Représentations d'espace d'états

Les représentations d'espace d'états des modèles d'équation de séries chronologiques apparemment non liées, à la section 3, sont définies par une équation de mesure et une équation de transition. L'équation de mesure définit la façon dont la série observée $\hat{\mathbf{y}}_t = (\hat{y}_{1,t}, \hat{y}_{2,t}, \hat{y}_{3,t}, \hat{y}_{4,t}, \hat{y}_{5,t})'$ dépend d'un vecteur $\mathbf{a}_t$ de dimension p contenant les variables d'état non observées :

$$\hat{\mathbf{y}}_t = \mathbf{Z}_t \mathbf{a}_t, \quad (\text{A.2})$$

où $\mathbf{Z}_t$ est une matrice de plan de sondage. (A.2) ne contient pas d'erreur de mesure, puisque dans la présente application, les erreurs de mesure sont comprises dans le vecteur d'état. L'équation de transition décrit la façon dont les variables d'état évoluent de la période $t-1$ à t et est définie comme suit :

$$\mathbf{a}_t = \mathbf{T} \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \quad (\text{A.3})$$

$$\begin{aligned} \text{E}(\boldsymbol{\eta}_t) &= \mathbf{0}_p, \\ \text{Cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t'}) &= \begin{cases} \mathbf{Q}_t & \text{si } t = t' \\ \mathbf{O}_{p \times p} & \text{si } t \neq t', \end{cases} \end{aligned}$$

où $\mathbf{0}_p$ est un vecteur de dimension p dont chaque élément est égal à zéro et $\mathbf{O}_{p \times p}$ est une matrice de dimension $p \times p$ dont chaque élément est égal à zéro. De même, $\mathbf{I}_p$ désigne la matrice d'identité $p \times p$ et $\mathbf{1}_p$, un vecteur de dimension p dont chaque élément est égal à un. Dans ce qui suit, il est entendu que $\text{diag}(\mathbf{x})$ est un opérateur transformant le vecteur $\mathbf{x}$ de dimension p en une matrice diagonale $p \times p$ et que les éléments de $\mathbf{x}$ sont sur la diagonale. De la même manière, $\text{diagblocs}(\mathbf{X}, \dots, \mathbf{Y})$ est un opérateur

transformant une liste de matrices déterminées $\mathbf{X}, \dots, \mathbf{Y}$ en une grande matrice diagonale par blocs dont les matrices déterminées sont des éléments diagonaux par blocs.

Pour le modèle (3.4) à (3.9), il s'ensuit que la représentation de l'espace d'états est obtenue au moyen de :

$$\begin{aligned}\mathbf{a}_t &= \left(\mathbf{a}_t^{L1}, \mathbf{a}_t^{L2}, \mathbf{a}_t^{L3}, \mathbf{a}_t^{L4}, \mathbf{a}_t^{L5}, \mathbf{a}_t^{S1}, \mathbf{a}_t^{S2}, \mathbf{a}_t^{S3}, \mathbf{a}_t^{S4}, \mathbf{a}_t^{S5}, \mathbf{a}_t^I, \mathbf{a}_t^e \right)', \\ \mathbf{a}_t^{Li} &= \left(L_{i,t}, R_{i,t} \right), i=1, \dots, 5, \\ \mathbf{a}_t^{Si} &= \left(S_{i,t,1}, S_{i,t,1}^*, S_{i,t,2}, S_{i,t,2}^*, S_{i,t,3}, S_{i,t,3}^*, S_{i,t,4}, S_{i,t,4}^*, S_{i,t,5}, S_{i,t,5}^*, S_{i,t,6} \right), i=1, \dots, 5, \\ \mathbf{a}_t^I &= \left(I_{1,t}, I_{2,t}, I_{3,t}, I_{4,t}, I_{5,t} \right), \\ \mathbf{a}_t^e &= \left(e_{1,t}, e_{2,t}, e_{3,t}, e_{4,t}, e_{5,t} \right),\end{aligned}$$

$$\begin{aligned}\mathbf{Z}_t &= \left(\mathbf{Z}_t^L, \mathbf{Z}_t^S, \mathbf{Z}_t^I, \mathbf{Z}_t^e \right), \text{ où} \\ \mathbf{Z}_t^L &= \mathbf{I}_5 \otimes (1, 0), \\ \mathbf{Z}_t^S &= \mathbf{I}_5 \otimes (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1), \\ \mathbf{Z}_t^I &= \mathbf{Z}_t^e = \mathbf{I}_5,\end{aligned}$$

$$\begin{aligned}\mathbf{T} &= \text{diagblocs} \left(\mathbf{T}^L, \mathbf{T}^S, \mathbf{T}^I, \mathbf{T}^e \right) \text{ où} \\ \mathbf{T}^L &= \mathbf{I}_5 \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \\ \mathbf{T}^S &= \mathbf{I}_5 \otimes \text{diagblocs} \left(\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3, \mathbf{C}_4, \mathbf{C}_5, 1 \right) \text{ avec (A.4)} \\ \mathbf{C}_i &= \begin{pmatrix} \cos(h_i) & \sin(h_i) \\ -\sin(h_i) & \cos(h_i) \end{pmatrix}, h_i = \frac{\pi l}{6}, l=1, \dots, 5, \\ \mathbf{T}^I &= \mathbf{T}^e = \mathbf{O}_{5 \times 5},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\eta}_t &= \left(\boldsymbol{\eta}_t^{L1}, \boldsymbol{\eta}_t^{L2}, \boldsymbol{\eta}_t^{L3}, \boldsymbol{\eta}_t^{L4}, \boldsymbol{\eta}_t^{L5}, \boldsymbol{\eta}_t^{S1}, \boldsymbol{\eta}_t^{S2}, \boldsymbol{\eta}_t^{S3}, \boldsymbol{\eta}_t^{S4}, \boldsymbol{\eta}_t^{S5}, \boldsymbol{\eta}_t^I, \boldsymbol{\eta}_t^e \right) \text{ avec} \\ \boldsymbol{\eta}_t^{Li} &= \left(0, \eta_{R,i,t} \right), i=1, \dots, 5, \\ \boldsymbol{\eta}_t^{Si} &= \left(\eta_{S,i,t,1}, \eta_{S,i,t,1}^*, \dots, \eta_{S,i,t,5}, \eta_{S,i,t,5}^*, \eta_{S,i,t,6} \right), i=1, \dots, 5, \\ \boldsymbol{\eta}_t^I &= \left(I_{1,t}, I_{2,t}, I_{3,t}, I_{4,t}, I_{5,t} \right), \\ \boldsymbol{\eta}_t^e &= \left(e_{1,t}, e_{2,t}, e_{3,t}, e_{4,t}, e_{5,t} \right),\end{aligned}$$

$$\mathbf{Q}_t = \text{diagblocs} \left(\mathbf{Q}^L, \mathbf{Q}^S, \mathbf{Q}^I, \mathbf{Q}^e \right) \text{ avec}$$

$\mathbf{Q}^L$ = matrice de taille 10×10 avec éléments de diagonale $(0, \sigma_{R,1}^2, 0, \sigma_{R,2}^2, 0, \sigma_{R,3}^2, 0, \sigma_{R,4}^2, 0, \sigma_{R,5}^2)$ et éléments hors diagonale $q_{2i,2i'} = \varsigma_{R,i,i'}^2$ pour $i, i' = 1, \dots, 5$, et tous les autres éléments hors diagonale sont égaux à zéro,

$$\mathbf{Q}^S = \text{diag} \left(1_{11} \otimes \sigma_{S,1}^2, 1_{11} \otimes \sigma_{S,2}^2, 1_{11} \otimes \sigma_{S,3}^2, 1_{11} \otimes \sigma_{S,4}^2, 1_{11} \otimes \sigma_{S,5}^2 \right),$$

$\mathbf{Q}^I$ = matrice de taille 5×5 avec éléments de diagonale $(\sigma_{I,1}^2, \sigma_{I,2}^2, \sigma_{I,3}^2, \sigma_{I,4}^2, \sigma_{I,5}^2)$ et éléments hors diagonale $\varsigma_{I,i,i'}$ pour $i, i' = 1, \dots, 5$,

$\mathbf{Q}_t^e$ = matrice de taille 5×5 avec éléments définis par (3.9).

La représentation de l'espace d'états du modèle (4.4) à (4.6) à la section 4 est obtenue à l'aide de l'équation de mesure (A.2), dans laquelle la série observée $\hat{\mathbf{y}}_t$ est remplacée par $\hat{\mathbf{p}}_t = (\hat{p}_t^+, \hat{p}_t^0, \hat{p}_t^-)'$, et l'équation de transition (A.3), où :

$$\begin{aligned}\mathbf{a}_t &= (\mathbf{a}_t^{L+}, \mathbf{a}_t^{L0}, \mathbf{a}_t^{L-}, \mathbf{a}_t^{S+}, \mathbf{a}_t^{S0}, \mathbf{a}_t^{S-}, \mathbf{a}_t^\beta, \mathbf{a}_t^I, \mathbf{a}_t^e)', \\ \mathbf{a}_t^{Li} &= (L_{i,t}, R_{i,t}), i = +, 0, -, \\ \mathbf{a}_t^{Si} &= (S_{i,t,1}, S_{i,t,1}^*, S_{i,t,2}, S_{i,t,2}^*, S_{i,t,3}, S_{i,t,3}^*, S_{i,t,4}, S_{i,t,4}^*, S_{i,t,5}, S_{i,t,5}^*, S_{i,t,6}), i = +, 0, -, \\ \mathbf{a}_t^\beta &= (\beta_t^+, \beta_t^0, \beta_t^-), \quad \mathbf{a}_t^I = (I_t^+, I_t^0, I_t^-), \quad \mathbf{a}_t^e = (e_t^+, e_t^0, e_t^-),\end{aligned}$$

$$\mathbf{Z}_t = (\mathbf{Z}_t^L, \mathbf{Z}_t^S, \mathbf{Z}_t^\beta, \mathbf{Z}_t^I, \mathbf{Z}_t^e) \text{ où}$$

$$\mathbf{Z}_t^L = \mathbf{I}_3 \otimes (1, 0),$$

$$\mathbf{Z}_t^S = \mathbf{I}_3 \otimes (1, 0, 1, 0, 1, 0, 1, 0, 1),$$

$$\mathbf{Z}_t^\beta = x_t \mathbf{I}_3, \quad \mathbf{Z}_t^I = \mathbf{Z}_t^e = \mathbf{I}_3,$$

$$\mathbf{T} = \text{diagblocs}(\mathbf{T}^L, \mathbf{T}^S, \mathbf{T}^\beta, \mathbf{T}^I, \mathbf{T}^e) \text{ où}$$

$$\mathbf{T}^L = \mathbf{I}_3 \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

$$\mathbf{T}^S = \mathbf{I}_3 \otimes \text{diagblocs}(\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3, \mathbf{C}_4, \mathbf{C}_5, 1), \text{ selon la définition de } \mathbf{C}_i \text{ dans (A.4)}$$

$$\mathbf{T}^\beta = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & -1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{T}^I = \mathbf{T}^e = \mathbf{O}_{3 \times 3},$$

$$\boldsymbol{\eta}_t = (\boldsymbol{\eta}_t^{L+}, \boldsymbol{\eta}_t^{L0}, \boldsymbol{\eta}_t^{L-}, \boldsymbol{\eta}_t^{S+}, \boldsymbol{\eta}_t^{S0}, \boldsymbol{\eta}_t^{S-}, \boldsymbol{\eta}_t^\beta, \boldsymbol{\eta}_t^I, \boldsymbol{\eta}_t^e) \text{ avec}$$

$$\boldsymbol{\eta}_t^{Li} = (0, \eta_{R,i,t}), i = +, 0, -,$$

$$\boldsymbol{\eta}_t^{Si} = (\eta_{S,i,t,1}, \eta_{S,i,t,1}^*, \dots, \eta_{S,i,t,5}, \eta_{S,i,t,5}^*, \eta_{S,i,t,6}), i = +, 0, -,$$

$$\boldsymbol{\eta}_t^\beta = \mathbf{0}_3, \quad \boldsymbol{\eta}_t^I = (I_t^+, I_t^0, I_t^-), \quad \boldsymbol{\eta}_t^e = (e_t^+, e_t^0, e_t^-),$$

$$\mathbf{Q}_t = \text{diagblocs}(\mathbf{Q}_t^L, \mathbf{Q}_t^S, \mathbf{Q}_t^\beta, \mathbf{Q}_t^I, \mathbf{Q}_t^e) \text{ avec}$$

$$\mathbf{Q}_t^L = \text{diag}(0, \sigma_{R,+}^2, 0, \sigma_{R,0}^2, 0, \sigma_{R,-}^2),$$

$$\mathbf{Q}_t^S = \text{diag}(1_{11} \otimes \sigma_{S,+}^2, 1_{11} \otimes \sigma_{S,0}^2, 1_{11} \otimes \sigma_{S,-}^2),$$

$$\mathbf{Q}_t^\beta = \mathbf{O}_{3 \times 3},$$

$$\mathbf{Q}_t^I = \text{diag}(\sigma_{I,+}^2, \sigma_{I,0}^2, \sigma_{I,-}^2),$$

$$\mathbf{Q}_t^e = \text{matrice de taille } 3 \times 3 \text{ dont les éléments sont définis par (4.6).}$$

Bibliographie

- Bell, W.R. et Hillmer, S.C. (1990). [Estimation dans les enquêtes à passages répétés au moyen de séries chronologiques](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1990002/article/14535-fra.pdf). *Techniques d'enquête*, 16(2), 205-227. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1990002/article/14535-fra.pdf>.
- Binder, D.A. et Dick, J.P. (1989). [Enquêtes répétées – Modélisation et estimation](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1989001/article/14579-fra.pdf). *Techniques d'enquête*, 15(1), 31-48. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1989001/article/14579-fra.pdf>.
- Binder, D.A. et Dick, J.P. (1990). [Méthode pour l'analyse des modèles ARMMI](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1990002/article/14533-fra.pdf). *Techniques d'enquête*, 16(2), 251-265. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1990002/article/14533-fra.pdf>.
- Blight, B.J.N. et Scott, A.J. (1973). A stochastic model for repeated surveys. *Journal of the Royal Statistical Society, Series B*, 35, 61-66.
- Bollineni-Balabay, O., van den Brakel, J.A. et Palm, F. (2017). [La modélisation espace-état appliquée aux séries chronologiques de l'Enquête sur la population active des Pays-Bas : sélection de modèles et estimation de l'erreur quadratique moyenne](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2017001/article/14819-fra.pdf). *Techniques d'enquête*, 43(1), 47-75. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2017001/article/14819-fra.pdf>.
- Boonstra, H.J. et van den Brakel, J.A. (2019). [Estimation du niveau et de la variation du chômage au moyen de modèles de séries chronologiques structurels](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2019003/article/00005-fra.pdf). *Techniques d'enquête*, 45(3), 421-454. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2019003/article/00005-fra.pdf>.
- Doornik, J.A. (2007). *An Object-oriented Matrix Programming Language Ox 5*. Londres: Timberlake Consultants Press.
- Doran, H.E. (1992). Constraining Kalman filter and smoothing estimates to satisfy time varying restrictions. *Review of Economics and Statistics*, 74 568-572.
- Durbin, J. et Koopman, S.J. (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Elliot, D.J. et Zong, P. (2019). Improving timeliness and accuracy from the UK labour force survey. *Statistical Theory and Related Fields*, 3, 186-198.
- Feder, M. (2001). Time series analysis of repeated surveys: The state-space approach. *Statistica Neerlandica*, 55, 182-199.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

- Harvey, A.C. et Chung, C.H. (2000). Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society, Series A*, 163, 303-339.
- Harvey, A.C. et Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society, Series A*, 149, 187-227.
- Harvey, A. et Koopman, S.J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88(424), 1228-1236.
- Hungnes, H., Skjerpen, T., Hamre, J.I., Jansen, X.C., Pham, D.Q. et Sandvik, O. (2024). Structural break in the Norwegian Labor Force Survey due to a redesign during a pandemic. *Journal of Official Statistics*, 40, 122-160.
- Koopman, S.J., Shephard, N. et Doornik, J.A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, Royal Economic Society, 2(1), 107-160.
- Koopman, S.J., Shephard, N. et Doornik, J.A. (2008). *Statistical Algorithms for Models in State Space using SsfPack 3.0*. Londres: Timberlake Consultants Press.
- Krieg, S. et van den Brakel, J.A. (2012). Estimation of the monthly unemployment rate for six domains through structural time series modelling with cointegrated trends. *Computational Statistics and Data Analysis*, 56, 2918-2933.
- Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business & Economic Statistics*, 9, 163-175.
- Pfeffermann, D. et Bleuer, S.R. (1993). [Modélisation conjointe robuste de séries de données sur l'activité pour de petites régions](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1993002/article/14458-fra.pdf). *Techniques d'enquête*, 19(2), 159-174. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1993002/article/14458-fra.pdf>.
- Pfeffermann, D. et Burck, L. (1990). [Estimation robuste pour petits domaines par la combinaison de données chronologiques et transversales](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1990002/article/14534-fra.pdf). *Techniques d'enquête*, 16(2), 229-249. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1990002/article/14534-fra.pdf>.
- Pfeffermann, D. et Sikov, A. (2011). Imputation and estimation under nonignorable nonresponse in household surveys with missing covariate information. *Journal of Official Statistics*, 27, 181-209.
- Pfeffermann, D. et Tiller, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, 26, 893-916.
- Pfeffermann, D. et Tiller, R. (2006). Small area estimation with state space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101, 1387-1397.

- Rao, J.N.K. et Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics*, 22, 511-528.
- Särndal, C.E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.
- Sverchkov, M. (2008). A new approach to estimation of response probabilities when missing data are not missing at random. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 867-874.
- Tam, S.M. (1987). Analysis of repeated surveys using a dynamic linear model. *Revue Internationale de Statistique*, 55, 63-73.
- Tiller, R.B. (1992). Time series modelling of sample survey data from the U.S. current population survey. *Journal of Official Statistics*, 8, 149-166.
- van den Brakel, J.A. et Krieg, S. (2015). [Remédier aux petites tailles d'échantillon, au biais de groupe de renouvellement et aux discontinuités dans les plans de sondage avec renouvellement de panel](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2015002/article/14231-fra.pdf). *Techniques d'enquête*, 41(2), 281-312. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2015002/article/14231-fra.pdf>.
- van den Brakel, J.A. et Krieg, S. (2016). Small area estimation with state-space common factor models for rotating panels. *Journal of the Royal Statistical Society, Series A*, 179, 763-791.
- van den Brakel, J.A., Krieg, S. et Smeets, M. (2023). *The Methodology of the Dutch Consumer Confidence Survey During the Corona Crisis*. Document de travail, décembre 2023. Statistics Netherlands, Heerlen.
- van den Brakel, J.A. et Roels, J. (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics*, 4, 1105-1138.
- van den Brakel, J.A., Söhler, E., Daas, P. et Buelens, B. (2017). [Les médias sociaux comme source de données pour les statistiques officielles; l'Indice de confiance des consommateurs des Pays-Bas](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2017002/article/54871-fra.pdf). *Techniques d'enquête*, 43(2), 201-230. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2017002/article/54871-fra.pdf>.
- van den Brakel, J.A., Souren, M. et Krieg, S. (2022). Estimating monthly Labour Force Figures during the COVID-19 pandemic in the Netherlands. *Journal of the Royal Statistical Society, Series A*, 185, 1560-1583.
- van den Brakel, J.A., Zhang, X. et Tam, S.M. (2020). Measuring discontinuities in time series obtained with repeated sample surveys. *Revue Internationale de Statistique*, 88, 155-175.