

N° 12-001-X au catalogue
ISSN 1712-5685

Techniques d'enquête

Modélisation de la variation temporelle des taux de réponse aux enquêtes : approche bayésienne s'appliquant à l'enquête sur la santé réalisée aux Pays-Bas

par Shiya Wu, Harm-Jan Boonstra, Mirjam Moerbeek
et Barry Schouten

Date de diffusion : le 30 juin 2023



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 le ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par le ministre de l'Industrie 2023

Tous droits réservés. 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.

Modélisation de la variation temporelle des taux de réponse aux enquêtes : approche bayésienne s'appliquant à l'enquête sur la santé réalisée aux Pays-Bas

Shiya Wu, Harm-Jan Boonstra, Mirjam Moerbeek et Barry Schouten¹

Résumé

Des estimations précises et sans biais des propensions à répondre (PR) jouent un rôle décisif dans l'observation, l'analyse et l'adaptation d'une collecte de données. Dans un environnement d'enquête fixe, ces paramètres sont stables et leurs estimations finissent par converger lorsque suffisamment de données historiques sont recueillies. Dans les pratiques d'enquête, toutefois, les taux de réponse varient progressivement dans le temps. Comprendre la variation temporelle de la prédiction des taux de réponse est essentiel lors de l'adaptation d'un plan d'enquête. La présente étude met en lumière la variation temporelle des taux de réponse au moyen de modèles hiérarchiques (à plusieurs niveaux) de séries chronologiques. Il est possible de générer des prédictions fiables en apprenant à partir de séries chronologiques historiques et de mises à jour avec de nouvelles données dans un cadre bayésien. Pour illustrer une étude de cas, nous nous concentrons sur des taux de réponse en ligne dans le cadre de l'enquête sur la santé réalisée aux Pays-Bas de 2014 à 2019.

Mots-clés : Analyse bayésienne; modèle hiérarchique; propension à répondre; série chronologique.

1. Introduction

Au cours des deux dernières décennies, les plans d'enquête dynamiques et adaptatifs (Chun, Heeringa et Schouten, 2018) ont suscité un intérêt considérable en rassemblant des caractéristiques des plans d'enquête avant ou pendant la collecte des données, dans le but ultime d'optimiser le rapport coût-qualité des enquêtes aux fins d'attribution efficace des ressources. L'émergence des enquêtes en ligne, la disponibilité de données de processus et l'augmentation des coûts des enquêtes ont stimulé la recherche sur l'observation (Kreuter, 2013) et l'adaptation (Schouten, Peytchev et Wagner, 2017) de la collecte des données. Toutefois, une compréhension approfondie de l'incidence des caractéristiques des plans et des variations temporelles sur des paramètres importants des modèles de réponse et de coût est impérative pour permettre toute adaptation. Un facteur essentiel est, par exemple, la vraisemblance qu'un participant réponde à une enquête (c'est-à-dire sa propension à répondre) qui peut dépendre de facteurs à la fois dépendants et indépendants de la nature de l'enquête elle-même. De plus, le coût de l'enquête est un calcul complexe qui concerne toutes les étapes, de la planification de l'enquête jusqu'au bilan des données, en passant par l'exécution de l'enquête; il peut influencer directement sur le type d'enquête menée, qui peut à son tour influencer la propension à répondre (PR). Pour cette raison, l'élaboration de telles mesures de paramètres est nécessaire avant d'entreprendre toute activité de collecte de données.

Au cours de la dernière décennie, on a assisté à un regain d'intérêt à l'égard de la prévisibilité de la propension à répondre dans le cadre de plans d'enquête dynamiques et adaptatifs. En techniques d'enquête,

1. Shiya Wu, Utrecht University, Department of Methodology and Statistics. Courriel : s.wu@uu.nl; Harm-Jan Boonstra, Statistics Netherlands, Department of Statistical Methods; Mirjam Moerbeek, Utrecht University, Department of Methodology and Statistics; Barry Schouten, Statistics Netherlands, Department of Statistical Methods and Utrecht University, Department of Methodology and Statistics.

L'utilisation de scores de propension (Rosenbaum et Rubin, 1983) est la façon courante d'adapter des fonctionnalités différentielles à des cas échantillonnés pour atteindre les objectifs souhaités en matière de coût ou de qualité. Dans un environnement changeant de collecte de données, le rendement et la structure d'un plan d'enquête dépendent considérablement de modèles de propension pouvant mener à des décisions inefficaces. En s'appuyant uniquement sur des données sur les processus ou des données de réponse au cours des premières étapes d'une enquête à plan de sondage adaptatif, par exemple, les estimations de la propension à répondre peuvent produire des estimations biaisées de la propension à répondre finale à la fin de la collecte des données (Wagner et Hubbard, 2014). De plus, l'incertitude des estimations de propension à répondre devrait être intégrée aux modèles de propension, afin d'éviter les conceptions moins optimales (Burger, Perryck et Schouten, 2017).

Des estimations exactes de la propension à répondre sont par conséquent au cœur des activités d'enquête. C'est pourquoi les spécialistes de la recherche sur les enquêtes appliquent des données historiques afin d'estimer les coefficients d'un modèle de propension, puis utilisent ces coefficients estimés dans les cycles suivants de l'enquête. L'analyse bayésienne (Gelman, Carlin, Stern, Dunson, Vehtari et Rubin, 2013) est une approche naturelle d'utilisation à la fois de données historiques et nouvelles pour améliorer les prédictions. Les conclusions de lois *a priori* tirées de données historiques deviennent des lois *a posteriori*, qui servent de lois *a priori* dans les analyses suivantes, à mesure que les données s'accumulent. Schouten, Mushkudiani, Shlomo, Durrant, Lundquist et Wagner (2018) ont été les premiers à appliquer un modèle bayésien général pour analyser la propension à répondre et le coût dans le cadre de l'enquête sur la santé réalisée aux Pays-Bas. Ils affirment qu'une mauvaise définition des lois *a priori* peut affaiblir le rendement de la prédiction. Ainsi, la déduction d'une loi *a priori* devient une étape influente. L'intégration de conclusions d'expert est un préalable pour une telle déduction de loi *a priori*. Cela découle d'un long historique de documentation biométrique et médicale; toutefois, cette application n'est qu'à ses débuts dans le contexte des enquêtes. De récents exemples ont été donnés par : West, Wagner, Coffey et Elliott (2021), qui ont examiné des données empiriques en matière de prédiction de la propension dans le cadre d'enquêtes; Coffey, West, Wagner et Elliott (2020), qui ont consulté des questionnaires de la collecte de données en matière de coefficients estimés; et Wu, Schouten, Meijers et Moerbeek (2022), qui ont eu recours à du personnel de collecte de données à titre de spécialistes afin de tirer parti des données historiques pertinentes selon les critères d'une nouvelle enquête ou d'une enquête remaniée.

Pour l'instant, ces approches supposent que les propensions à répondre sont stables sur une période relativement courte. Dans un environnement d'enquête fixe, ces paramètres demeurent stables et leurs estimations convergent finalement vers l'accumulation de données historiques. Dans les pratiques d'enquête, cependant, ces paramètres varient progressivement au fil du temps ce qui signifie que les prévisions peuvent ne pas converger. On peut, par exemple, observer une variation saisonnière et des tendances à la baisse des taux de réponse. Ainsi, l'avantage d'une déduction de loi *a priori* pourrait éventuellement être annulé si l'on ne tient pas compte des variations temporelles. Des études récentes de Mushkudiani et Schouten (2019) ainsi que de Fang, Burger, Meijers et van Berkel (2020) décrivent quels

facteurs dépendant du temps affectent de manière significative l'exactitude de l'estimation de paramètres; l'incidence sur l'exactitude de la prévision est toutefois toujours inconnue, ce qui est le sujet de la présente étude.

La présente étude fournit de nouveaux renseignements, de façon structurelle, sur des modèles souples de séries chronologiques de propension à répondre dans le cadre de plans d'enquête adaptatifs. Nous tentons d'interpréter la variation temporelle dans les propensions à répondre aux enquêtes présentant une corrélation significative avec les biais de non-réponse lorsque cette dernière fait l'objet d'une variation dans le temps. Notre approche s'applique à des enquêtes transversales répétées comptant de multiples étapes de collecte de données.

Notre principal objectif est de faire des prédictions fiables de la propension à répondre pour les diverses strates de population pertinentes. (Il convient de mentionner que les strates de population au sein desquelles les propensions à répondre différent peuvent être des sous-populations d'intérêt; on les appelle strates dans la présente étude, même si elles ne correspondent pas nécessairement aux strates d'échantillonnage). Nous examinons également le rendement de la prédiction, afin de pouvoir mesurer dans quelle mesure le temps modifie la propension à répondre. Cette question générale peut se résumer à quatre aspects concrets :

- 1) Quelles composantes d'une série chronologique contribuent le plus à la variation des propensions à répondre ?
- 2) Quel degré d'exactitude de la prédiction de la propension à répondre peut être obtenu pour la période suivante ?
- 3) Dans quelle mesure l'exactitude de la prédiction varie-t-elle pour les strates de population ?
- 4) Dans quelle mesure l'exactitude de la prédiction dépend-elle de la longueur de la série chronologique d'enquête historique ?

Les nombreuses connaissances relatives aux séries chronologiques d'enquête historiques nous permettent d'étudier les effets de facteurs temporels sur la propension à répondre. Nous considérons deux niveaux (temps et strate) qui constituent plusieurs composantes intervenant dans un modèle de séries chronologiques. Ces composantes décrivent la variation au fil du temps ou parmi les strates ou les deux, et peuvent être analysées individuellement et collectivement. Plusieurs études se rapportant aux techniques d'enquête reposent sur une telle approche par modèle hiérarchique de séries chronologiques pour les statistiques officielles; Boonstra et van den Brakel (2019 et 2022), par exemple, estiment les taux de chômage régionaux mensuels et trimestriels à l'aide d'un modèle hiérarchique bayésien pour emprunter des informations dans le temps, l'espace et découlant de séries auxiliaires. Une telle utilisation provient de la documentation sur l'estimation sur petits domaines (Rao et Molina, 2015).

Dans la présente étude, nous utilisons l'enquête sur la santé réalisée aux Pays-Bas (GEZO), afin d'évaluer notre approche relativement aux quatre questions de recherche susmentionnées. Cette enquête présente une plan stable depuis 2011; nous nous concentrons sur la série chronologique de 2014 à 2019.

Pour optimiser les prédictions, nous comparons un ensemble de compositions de modèles selon divers critères d'information, afin d'atteindre un équilibre entre la qualité de l'ajustement et la complexité du modèle. Pour évaluer le modèle « optimal », nous évaluons sa performance prédictive et son exactitude en fonction de sa capacité à saisir correctement l'ampleur et la variation de la propension à répondre. Il convient de mentionner que nous nous concentrons sur l'obtention d'une inférence fiable au fil du temps, plutôt que sur l'atténuation de l'erreur de non-réponse, qui est l'un des objectifs des plans d'enquête adaptatifs.

Cette étude présente d'abord plusieurs facteurs temporels particulièrement pertinents en matière de variation et donne un exemple hypothétique à la section 2, puis elle traite des compositions différentielles de modèles sous la forme générale d'un modèle hiérarchique bayésien de séries chronologiques à la section 3. La section 4 présente l'optimisation du rendement du modèle à partir de l'analyse empirique de l'enquête GEZO. Nous discutons de nos constatations et concluons l'étude par un aperçu des travaux à venir à la section 5.

2. Composantes des séries chronologiques dans les taux de réponse aux enquêtes

Il est bien connu que la propension à répondre varie progressivement au fil du temps. Ne pas intégrer cette dépendance temporelle dans les décisions relatives au plan peut mener à des conceptions d'enquête inefficaces. Dans la présente section, nous avons recours à un exemple présentant des facteurs temporels associés à une variation considérable de la propension à répondre.

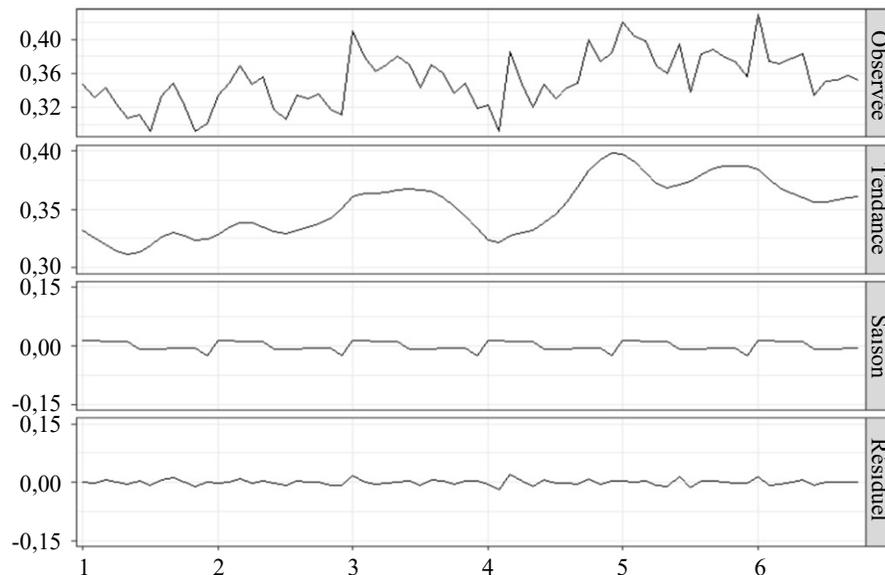
Nous nous concentrons sur les sous-groupes de population, ou strates, désignés par $g \in \{1, \dots, G\}$, puisque notre intention, en fin de compte, est que les modèles proposés servent à éclairer les décisions en matière de plan d'enquête adaptatif. Les strates sont formées à l'aide de variables auxiliaires liées à l'échantillon et sont donc disponibles pour toutes les unités de l'échantillon. Une propension à répondre de série chronologique $\rho_{g,t}$ pour la strate g et le temps t est une séquence de variables aléatoires. En supposant que les données historiques d'enquête sont disponibles jusqu'au moment t , nous souhaitons mesurer la variation qui découle de facteurs temporels dans les prédictions de la propension à répondre les plus à jour. Pour atteindre ce but, nous proposons d'abord des facteurs temporels potentiels. En donnant un exemple d'une série chronologique divisée selon les composantes suivantes : tendance, saisonnalité, etc., la figure 2.1 permet de comparer le taux de réponse général pour la variation temporelle suivante :

- *Tendance*. La tendance rend compte de la variation à long terme de la série chronologique observée sans la variation saisonnière. Elle montre la tendance générale des taux de réponse au niveau de la population au fil des années, qui peut être linéaire ou non linéaire. Ainsi, cette tendance permet d'étudier la croissance ou la diminution des prévisions à long terme. Comme le montre la figure 2.1, la direction à long terme ne se comporte pas comme une fluctuation cyclique. Il est de la plus haute importance pour l'élaboration d'un modèle de séparer la tendance

totale en une tendance globale partagée par toutes les strates et des tendances locales, c'est-à-dire propres aux strates.

- *Saisonnalité.* La variation saisonnière dans les réponses générales rend compte des variations périodiques à récurrence régulière qui n'influent pas sur les moyennes annuelles. Les fluctuations périodiques présentent une nature systématique et liée au calendrier qui peut être prédite et attribuée à une saison fixe dans une année. Le taux de réponse serait, par exemple, plus élevé en début d'année et relativement plus faible au milieu de l'année ou en décembre.
- *Variation résiduelle.* La variation résiduelle est la portion du signal obtenu après avoir exclu toutes les composantes susmentionnées. On modélise généralement cette portion en tant que bruit blanc, c'est-à-dire des variations indépendantes normalement distribuées.

Figure 2.1 Série observée de taux de réponse générale simulés sur plusieurs années et sa décomposition.



De plus, d'autres composantes dépendantes du temps que la figure 2.1 ne montre pas peuvent exister et avoir une incidence importante sur la fiabilité des prédictions de la propension à réponse pour les strates. Nous considérons donc des composantes temporelles supplémentaires liées aux strates :

- *Strate.* Différents sous-groupes ont des comportements de réponse différents; par exemple, des sous-groupes de jeunes sont plus susceptibles de répondre à l'enquête en ligne que les sous-groupes de personnes plus âgées, car ces derniers peuvent ne pas être aussi familiers avec Internet et ne pas y avoir autant accès. Cette variation selon les sous-groupes entraîne une tendance différentielle au niveau de la strate et peut également contribuer à une variation saisonnière différentielle.

- *Variation d'échantillonnage.* La variation d'échantillonnage complique l'estimation de la propension à répondre, en particulier pour les strates de taille d'échantillon réduite. La variation d'échantillonnage est prise en compte en adoptant une vraisemblance binomiale.
- *Événements inattendus.* Des événements inattendus, comme des serveurs Web temporairement inaccessibles, apparaissent comme des valeurs aberrantes et peuvent ne pas correspondre à la tendance existante. Ils correspondent à des variations irrégulières sur de courtes périodes. La variation obtenue ne suit pas de modèle particulier, est imprévisible et peut avoir une influence sur la prédiction de la propension à répondre future.
- *Intervention.* Une modification du plan de l'enquête, comme l'introduction d'un incitatif, est largement appliquée comme intervention volontaire en vue de stimuler des réponses et améliorer la qualité de la collecte des données, voire d'attribuer de façon efficace des ressources limitées afin de réduire le coût d'une enquête. Une intervention a une incidence permanente sur la propension à répondre. Cette influence peut être prédite, mais peut uniquement être étudiée au détriment de la valeur potentielle de riches données historiques et d'une longue période de collecte de données avant l'application de l'intervention. La variation obtenue est moins susceptible d'influer sur les tendances saisonnières, alors qu'elle peut avoir des répercussions similaires sur les taux de réponse de certaines strates.

Ensemble, toutes ces composantes, à l'exception de la variation d'échantillonnage, constituent le signal, c'est-à-dire les propensions à répondre latentes réelles, mais inconnues. Les formulations mathématiques correspondant à chaque composante sont présentées dans la section suivante où nous proposons le modèle structurel de séries chronologiques (voir Harvey, 1990 ainsi que Durbin et Koopman, 2012 pour obtenir des renseignements contextuels généraux sur ces composantes et ces modèles de séries chronologiques).

3. Méthodologie

Dans la présente section, nous traduisons les composantes de la série chronologique dont il est question dans la section 2 en modèles hiérarchiques de séries chronologiques et élaborons la stratégie d'estimation. Nous adoptons une approche bayésienne, afin de tenir compte de l'incertitude dans les données d'enquête historiques et de mettre à jour les prédictions de propension à répondre au fil du temps. Le recours à des modèles hiérarchiques est étendu dans le cadre d'estimations sur petits domaines, pour lesquelles on se concentre sur une estimation fiable pour des domaines tels que des zones géographiques, des périodes, des sous-groupes démographiques ou une combinaison de ceux-ci, dont les tailles d'échantillons sont souvent trop réduites pour fournir des estimations directes fiables (voir Rao et Molina, 2015 pour en avoir un aperçu). Parmi les références anciennes à la littérature portant sur les études de petits domaines à l'aide de modèles hiérarchiques de séries chronologiques, mentionnons les ouvrages de Pfeiffermann et Burck (1990), Rao et Yu (1994), Datta, Lahiri, Maiti et Lu (1999), You, Rao et Gambino (2003). Dans la plupart de ces études, y

compris celles de Boonstra et van den Brakel (2019), on suppose une distribution d'échantillon gaussienne, éventuellement après transformation adéquate des données. Une différence notable dans notre présente application à la propension à répondre est que nous utilisons une distribution d'échantillon binomiale, qui est une distribution naturelle permettant de décrire le processus de réponse en fonction du nombre de personnes échantillonnées dans chaque sous-groupe démographique et chaque période. Franco et Bell (2015) ont étudié de tels modèles binomiaux de séries chronologiques. Leur approche ressemble à notre stratégie, même si la nôtre fait intervenir davantage de types différents de composantes de séries chronologiques dans la spécification du modèle, comme le caractère saisonnier.

Nous commençons la discussion de notre méthode en présentant la notation utilisée tout au long de l'étude. Nous décrivons ensuite notre modèle et la stratégie utilisée pour estimer la propension à répondre et nous concluons en présentant les critères utilisés pour évaluer le rendement et l'applicabilité des modèles de prédiction de la propension à répondre dans le cadre bayésien.

3.1 Spécification du modèle hiérarchique de séries chronologiques

L'objectif est de prédire les propensions à répondre au niveau de la strate à un certain moment. Nous répartissons la population ou un échantillon en strates fondées sur plusieurs variables auxiliaires (c'est-à-dire l'application d'une stratification), ce qui équivaut à un classement recoupé de certaines variables. Ici, nous supposons que la stratification est précisée avant d'ajuster les modèles. Les catégories de chaque variable peuvent être fusionnées pour fournir des tailles d'échantillons suffisantes.

Soit $n_{g,t}$ la taille d'échantillon pour la strate g pour la vague t et $r_{g,t}$, le nombre de répondants, où $g \in \{1, \dots, G\}$ et $t \in \{1, \dots, T\}$. Le nombre de strates G est généralement de l'ordre de 10 à 20 et T désigne les vagues d'enquête, chacune étant une nouvelle réplique de l'enquête à partir d'un nouvel échantillon. Nous supposons que toutes les unités échantillonnées sont indépendantes quant à leur comportement de réponse au sein des strates et entre elles. Pour la strate g et le temps t , la réponse $r_{g,t}$ suit une distribution binomiale, si la propension à répondre (PR) est $\rho_{g,t}$ et la taille d'échantillon $n_{g,t}$, c'est-à-dire $r_{g,t} | n_{g,t}, \rho_{g,t} \sim \text{Binom}(n_{g,t}, \rho_{g,t})$. Puisque la propension à répondre est limitée de sorte qu'elle se situe entre 0 et 1, nous transformons l'échelle 0-1 en ligne réelle \mathbb{R} en utilisant une fonction de lien logit (d'autres fonctions de lien peuvent également être utilisées). La fonction fournit une transformation non linéaire et produit une variable latente $\theta_{g,t}$, qui suit la fonction de logarithme du risque :

$$\theta_{g,t} = \text{logit}(\rho_{g,t}) = \ln \left(\frac{\rho_{g,t}}{1 - \rho_{g,t}} \right).$$

Nous pouvons remanier la transformation pour calculer $\rho_{g,t}$:

$$\rho_{g,t} = \frac{\exp(\theta_{g,t})}{1 + \exp(\theta_{g,t})}.$$

Pour toute strate g et tout temps t , la variable explicative linéaire $\theta_{g,t}$ peut prendre la forme la plus générale, être linéaire, additive, multiniveau et constituée de plusieurs composantes de séries chronologiques. Comme cela est expliqué à la section 2.2, des variables démographiques définissent les strates, une tendance globale, une variation saisonnière, des tendances propres aux strates et une variation résiduelle. Le modèle hiérarchique devient ainsi :

$$\theta_{g,t} = \boldsymbol{\beta}' \mathbf{x}_g + \gamma t + \boldsymbol{\delta}' \mathbf{s}_t + v_g + u_t + z_{g,t} + w_{g,t}, \quad (3.1)$$

où le vecteur- p des effets de régression $\boldsymbol{\beta}$ est associé à des covariables indépendantes du temps \mathbf{x}_g . Dans l'application sur laquelle nous nous concentrons plus loin dans la présente étude, toutes les covariables sont binaires, car nous considérons uniquement des variables catégoriques. Toutefois, dans le cadre d'une utilisation plus générale, les entrées pourraient être des variables ordinales ou numériques, comme des tentatives de prise de contact et elles pourraient même varier au fil du temps.

Le scalaire γ est le paramètre de pente de la tendance temporelle linéaire générale. Le vecteur $\boldsymbol{\delta}$ comprend des effets saisonniers, le vecteur \mathbf{s}_t permettant de sélectionner la saison correspondant au mois t . Les effets saisonniers sont soit communs à toutes les strates, soit propres à une strate. Dans la présente étude, nous définissons les saisons comme une division de mois dans une année civile, c'est-à-dire des ensembles $\{1, 2\}$, $\{3, 4, 5\}$, $\{6, 7, 8\}$, $\{9, 10, 11\}$ et $\{12\}$ correspondant à l'hiver, au printemps, à l'été, à l'automne et à Noël.

Les trois premiers termes sont modélisés comme des effets fixes alors que les quatre derniers termes sont modélisés comme des effets aléatoires dans (3.1). Le premier de ces termes aléatoires correspond aux ordonnées à l'origine aléatoires pour la strate que l'on suppose être normalement distribuée avec une moyenne 0 et une variance σ_v^2 selon :

$$v_g \sim N(0, \sigma_v^2) \quad (3.2)$$

de façon identique et indépendante pour $g = 1, \dots, G$. Deuxièmement, une tendance temporelle globale est définie par un vecteur d'effet aléatoire $\mathbf{u} = (u_1, \dots, u_T)$ distribué selon :

$$\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{V}_u). \quad (3.3)$$

La matrice de covariance \mathbf{V}_u rend compte des structures de covariance entre tout u_i et tout u_j . On peut supposer une marche aléatoire de premier ordre (RW1, connu comme la tendance au niveau local) ou une marche aléatoire de deuxième ordre (RW2, c'est-à-dire la tendance dite lisse). Les structures dépendant du temps sont exprimées de façon plus pratique par la matrice de précision $\mathbf{Q}_u = \mathbf{V}_u^{-1}$. On préfère la matrice de précision à la matrice de covariance, car elle est parcimonieuse et permet le calcul efficace de l'inférence de la loi *a posteriori* hiérarchique dans une analyse bayésienne (voir par exemple Rue et Held, 2005). La matrice \mathbf{Q}_u pour RW1 et RW2 est une matrice tridiagonale et une matrice pentadiagonale (on suppose une matrice en bande $Q = (q_{i,j})$, 1 a une bande non nulle le long de la diagonale principale de telle sorte que

$q_{i,j} = 0$ si $|i - j| > 1$, alors que 2 présente deux bandes non nulles, de telle sorte que $q_{i,j} = 0$ si $|i - j| > 2$) (voir l'annexe C pour connaître leurs définitions). Il convient de mentionner que les matrices de précision \mathbf{Q}_u sont singulières, menant à une loi *a priori* inappropriée. Cela n'est pas un problème, puisque des contraintes peuvent être imposées sur ces effets aléatoires, afin de s'assurer que tous les coefficients de modèles demeurent identifiables. Sous RW1 et RW2 \mathbf{u} , la contrainte est $\sum_t u_t = 0$. Sous RW2 \mathbf{u} , la contrainte $\sum_t t u_t = 0$ est également imposée, de sorte que le niveau global correspondant et la pente linéaire soient saisis par l'ordonnée à l'origine du modèle et l'effet fixe γ .

Nous envisageons également des distributions autres que la distribution normale en (3.3). Nous envisageons, en particulier, la loi *a priori* de Laplace, la loi *a priori* de Student et la distribution *a priori* horseshoe comme autres options. De telles lois *a priori* peuvent être encadrées comme des combinaisons d'échelle de la distribution normale (voir West, 1987; Carvalho, Polson et Scott, 2010; Polson et Scott, 2010).

Le troisième terme d'effet aléatoire $\mathbf{z}_g = (z_{g,1}, \dots, z_{g,T})$ désigne des tendances propres aux strates réparties selon :

$$\mathbf{z}_g \sim N(0, \sigma_z^2 \mathbf{V}_z) \quad (3.4)$$

pour $g = 1, \dots, G$. La matrice de covariance \mathbf{V}_z décrit une marche aléatoire RW1 au cours des mois. La matrice de précision correspondante est la même que celle décrite ci-dessus et une contrainte à somme nulle est imposée à chaque vecteur de tendance \mathbf{z}_g , puisque les niveaux propres aux strates sont déjà saisis par les ordonnées à l'origine aléatoires v_g . Il est important de souligner que les tendances \mathbf{z}_g partagent un paramètre de covariance commun σ_z^2 . Il serait possible d'envisager un paramètre de variance distinct, mais nous avons conclu que cela entraînerait un surajustement.

Le dernier terme $w_{g,t}$ de (3.1) représente le bruit blanc et correspond à la variation non structurée restante dans les propensions à répondre au fil du temps et pour les strates, c'est-à-dire au niveau le plus détaillé. Pour toute strate g et tout temps t , ces composantes sont distribuées de façon indépendante et identique selon :

$$w_{g,t} \sim N(0, \sigma_w^2), \quad (3.5)$$

à l'aide du paramètre de variance unique σ_w^2 .

L'expression (3.1) expose le modèle le plus général envisagé en combinant toutes les composantes sous-jacentes. À la section 4, nous étudions ce modèle complet ainsi que des modèles reposant sur divers sous-ensembles des composantes décrites en (3.2)-(3.5).

3.2 Stratégie d'estimation

Dans la présente section, nous adoptons une approche bayésienne hiérarchique pour estimer les coefficients de modèle et prédire la propension à répondre (PR). Puisque les distributions *a posteriori* ne

sont pas disponibles sous forme fermée, un échantillonneur de Gibbs est utilisé selon la mise en œuvre du paquetage R *mcmcsm* (Boonstra, 2021). Nous commençons la sous-section en spécifiant les lois *a priori* attribuées aux paramètres du modèle.

Pour les effets fixes β , nous supposons une loi *a priori* faiblement informative :

$$\beta \sim N(0, 100\mathbf{I}_\beta),$$

avec la matrice d'identité \mathbf{I}_β . Des erreurs-types pour β de 10 sont prises en compte, ce qui est suffisamment élevé pour l'échelle des PR relative aux échelles de covariables. De la même façon, nous attribuons également à la tendance temporelle linéaire γ et aux effets saisonniers δ des lois *a priori* faiblement informatives, avec la même erreur-type.

Pour les composantes d'effets aléatoires, nous attribuons aux paramètres de variance de (3.2) à (3.5) des lois *a priori* inverses χ^2 , selon la condition de paramètres auxiliaires ξ , avec 1 degré de liberté et un paramètre d'échelle ξ^2 ; par exemple, $\sigma_v^2 | \xi_v \sim \text{Inv} - \chi^2(1, \xi_v^2)$. Les hyperparamètres ξ sont attribués aux lois *a priori* $N(0, 1)$. Combiner les variances normale ξ et de khi carré inverse conditionnelle donne des lois *a priori* marginales demi-Cauchy pour chaque paramètre d'écart type σ_v , σ_u , σ_z et σ_w . Comme le suggèrent Gelman (2006) ainsi que Polson et Scott (2010), les lois *a priori* demi-Cauchy pour les écarts types, ou la famille de lois *a priori* demi-t plus générales, sont généralement plus efficaces que les lois *a priori* gamma inverses couramment utilisées pour les paramètres de variance, qui peuvent être trop informatives.

Le vecteur des (hyper) paramètres, désigné par ψ :

$$\psi = (\beta, \gamma, \delta, v, u, z, w, \sigma_v^2, \sigma_u^2, \sigma_z^2, \sigma_w^2, \xi_v, \xi_u, \xi_z, \xi_w)$$

comprend tous les paramètres de (3.1), les paramètres de variance associés aux termes d'effet aléatoire ainsi que les paramètres auxiliaires introduits. La fonction de vraisemblance peut être exprimée comme suit :

$$p(r | n, \psi) \propto \prod_{g,t} \rho_{g,t}^{r_{g,t}} (1 - \rho_{g,t})^{n_{g,t} - r_{g,t}}, \quad (3.6)$$

où $\rho = \text{logit}^{-1}(\theta(\psi))$ et θ est la fonction linéaire de variable explicative du vecteur ψ comme l'exprime (3.1). Selon le théorème de Bayes, la loi *a posteriori* du vecteur ψ est proportionnelle au produit de la loi *a priori* et de la vraisemblance, c'est-à-dire $p(\psi | n, r) \propto p(\psi) p(r | n, \psi)$. L'échantillonneur de Gibbs génère alors des échantillons à partir de la loi *a posteriori* conjointe et les estimations de la loi *a posteriori* de PR $\rho_{g,t}$ sont obtenues comme sous-produits de ces échantillons; par échantillon, les PR peuvent être calculées à l'aide d'une transformation logit inversée. Des échantillons répétés sont tirés de la loi *a posteriori* conditionnelle complète de chaque (hyper) paramètre. (Voir l'annexe D pour obtenir de plus amples renseignements sur les distributions *a posteriori* conditionnelles complètes.)

L'échantillonneur de Gibbs produit trois chaînes de Markov à l'aide du paquetage *mcmcsm* (Boonstra, 2021) programmé dans R (R Core Team, 2020). Chaque chaîne consiste en 1 500 tirages générés

séquentiellement; toutefois seuls les 1 000 derniers tirages sont conservés pour l'algorithme d'estimation. La convergence de l'échantillon de la méthode de Monte Carlo par chaînes de Markov (MCMC) est évaluée à l'aide de tracés de trajectoire et d'autocorrélation. Le facteur potentiel de réduction d'échelle de Gelman-Rubin (Gelman et Rubin, 1992) est évalué pour diagnostiquer la combinaison de chaînes. On réduit, en particulier, l'autocorrélation des tirages séquentiels, puisque l'échantillonneur de Gibbs par blocs met simultanément à jour tous les coefficients fixes et aléatoires. De plus, cette approche inclut une nouvelle approche d'augmentation des données aux fins d'échantillonnage à partir de modèles logistiques binomiaux (Polson, Scott et Windle, 2013), laquelle est connue pour mener à un échantillonneur efficace et convergent relativement rapidement.

3.3 Critères de rendement

Pour guider la construction du modèle à l'aide des composantes de modèle et des lois *a priori* décrites aux sections 3.1 et 3.2 ainsi que pour évaluer la pertinence des modèles, nous utilisons trois critères pour évaluer les modèles et un pour évaluer le rendement de prédiction des modèles.

Les critères de sélection courants et populaires des environnements hiérarchiques bayésiens sont le critère d'information largement applicable WAIC (Watanabe, 2010, 2013) et le critère d'information de déviance (DIC) (Spiegelhalter, Best, Carlin et van der Linde, 2002). Ils sont choisis pour obtenir un équilibre raisonnable entre l'ajustement du modèle, la complexité du modèle et un calcul efficace (voir l'annexe C pour connaître les définitions). Nous préférons les modèles présentant des critères DIC/WAIC inférieurs. Nous utilisons ensuite des valeurs p prédictives *a posteriori* pour vérifier l'ajustement du modèle, c'est-à-dire en simulant des tirages à partir d'une distribution prédictive *a posteriori* et en les comparant avec les données observées (voir par exemple Gelman, Meng et Stern, 1996). Cela permet d'évaluer si le modèle hiérarchique peut reproduire des données similaires aux observations. Les valeurs p sont définies selon :

$$p = \Pr\left(S(r^{\text{rép}}) \geq S(r) \mid r\right), \quad (3.7)$$

où S est une variable à tester et $r^{\text{rép}}$ désigne un ensemble de données répété généré à partir d'une distribution prédictive *a posteriori* fondée sur le modèle ajusté, $p(r^{\text{rép}} \mid r) = \int p(r^{\text{rép}} \mid \rho, n) p(\rho \mid r, n) d\rho$. Les valeurs p sont estimées à partir du résultat de la méthode MCMC; les valeurs proches de 0 ou de 1 indiquent un piètre ajustement pour des statistiques S . Nous considérons deux variables à tester :

1. $S(r) = \bar{r}$, la moyenne non pondérée du vecteur de données répliqué;
2. $S(r) = \frac{1}{GT-1} \sum_{g,t} (r_{g,t} - \bar{r})^2$, la variance non pondérée du vecteur de données répliqué et \bar{r} est la moyenne de $r_{g,t}$.

Pour évaluer le rendement de prédiction des modèles, nous définissons une mesure prédictive : la racine de l'erreur quadratique moyenne (REQM ou RMSE) de la strate g pour le mois t , sous la forme de la racine carrée de la somme de deux termes : 1) les différences quadratiques entre les moyennes *a posteriori*

de $\rho_{g,t}$ et le taux de réponse (TR) observé, et 2) les variances *a posteriori* de $\rho_{g,t}$. La forme générale de l'expression de la strate g pour le mois t est :

$$\text{REQM}(g, t) = \sqrt{\left(E_{\pi_t}(\rho_{g,t}) - \hat{\rho}_{g,t}\right)^2 + \text{var}_{\pi_t}(\rho_{g,t})}, \quad (3.8)$$

où $\hat{\rho}_{g,t}$ est la valeur réalisée de la PR et est estimée par le TR observé et π_t est la distribution prédictive *a posteriori* de la PR, lorsque nous utilisons des données historiques jusqu'à $t-1$ inclus et de nouvelles données dans t pour la prédiction de PR. Pour faciliter la notation, les deux termes sous la racine carrée de (3.8) sont désignés comme le terme de biais ($B(g, t)$) et l'écart type ($\text{ET}(g, t)$). Le terme de biais de (3.8) est, en général, supérieur à zéro du fait de la variation aléatoire dans l'échantillonnage des strates et dans la réponse des unités d'échantillonnage. Nous comparons la REQM par rapport à la limite empirique inférieure désignée par REQM_{\min} . L'estimation de la limite inférieure s'appelle l'approximation de Monte Carlo de la moyenne *a posteriori* des écarts types binomiaux, qui est une fonction de l'itération k^e à partir des tirages *a posteriori* de $\rho_{g,t}$:

$$\text{REQM}_{\min}(g, t) = \frac{1}{K} \sum_{k=1}^K \sqrt{\frac{\rho_{g,t}^{(k)}(1 - \rho_{g,t}^{(k)})}{n_{g,t}}}, \quad (3.9)$$

où k désigne les passages sur les tirages MCMC et $n_{g,t}$ est la taille de l'échantillon de la strate g du mois t . (3.8) et (3.9) fournissent des évaluations sur un mois par strate g . Elles doivent être agrégées pour toutes les strates et tous les temps, afin d'obtenir des évaluations générales utiles.

Pour un mois en particulier, une strate de taille d'échantillon supérieure devrait mettre davantage de poids sur les prédictions fiables. Les poids $d_{g,t}$ sont définis comme la proportion de l'échantillon, c'est-à-dire :

$$d_{g,t} = \frac{n_{g,t}}{\sum_g n_{g,t}} \quad \text{en fonction de} \quad \sum_g d_{g,t} = 1.$$

Par conséquent, les sous-termes :

$$B(t) = \sqrt{\sum_g d_{g,t} \left(E_{\pi_t}(\rho_{g,t}) - \hat{\rho}_{g,t}\right)^2}$$

et

$$\text{ET}(t) = \sqrt{\sum_g d_{g,t} \text{var}_{\pi_t}(\rho_{g,t})}$$

pour le mois t devraient être la racine carrée de la somme des mesures individuelles pondérées $B(g, t)$ et $\text{ET}(g, t)$ de $d_{g,t}$ pour la strate, alors que la limite inférieure pour la strate au temps t devient :

$$\text{REQM}_{\min}(t) = \frac{1}{K} \sum_{k=1}^K \sqrt{\sum_g d_{g,t} \frac{\rho_{g,t}^{(k)}(1 - \rho_{g,t}^{(k)})}{n_{g,t}}}.$$

De plus, les sous-termes propres aux strates :

$$B(g, T) = \frac{1}{T} \sum_t \sqrt{\left(E_{\pi_t}(\rho_{g,t}) - \hat{\rho}_{g,t}\right)^2}$$

et

$$ET(g, T) = \frac{1}{T} \sum_t \sqrt{\text{var}_{\pi_t}(\rho_{g,t})}$$

pour une période $T = \{t | t_1, \dots, t_T\}$ sont la moyenne des mesures individuelles $B(g, t)$ et $ET(g, t)$ pour les mois où t indique un mois, alors que la limite inférieure propre à la strate pour la période T devient la moyenne des mesures individuelles $\text{REQM}_{\min}(g, t)$, soit :

$$\text{REQM}_{\min}(g, T) = \frac{1}{T} \sum_t \frac{1}{K} \sum_{k=1}^K \sqrt{\frac{\rho_{g,t}^{(k)}(1 - \rho_{g,t}^{(k)})}{n_{g,t}}}$$

De plus, les sous-termes ou termes généraux pour la période T deviennent la moyenne des sous-termes pondérés $B(t)$, $ET(t)$ et $\text{REQM}_{\min}(t)$ pour les mois, soit :

$$B(T) = \frac{1}{T} \sum_t \sqrt{\sum_g d_{g,t} \left(E_{\pi_t}(\rho_{g,t}) - \hat{\rho}_{g,t}\right)^2},$$

$$ET(T) = \frac{1}{T} \sqrt{\sum_g d_{g,t} \text{var}_{\pi_t}(\rho_{g,t})}$$

et

$$\text{REQM}_{\min}(T) = \frac{1}{T} \sum_t \frac{1}{K} \sum_{k=1}^K \sqrt{\sum_g d_{g,t} \frac{\rho_{g,t}^{(k)}(1 - \rho_{g,t}^{(k)})}{n_{g,t}}}$$

4. Analyse des résultats

Dans la section qui suit, nous présentons l'enquête sur la santé réalisée aux Pays-Bas (GEZO) comme étude de cas pour montrer la façon de construire les modèles hiérarchiques de séries chronologiques et de mettre à jour la PR au fil du temps. Nous étudions quatre questions de recherche dans les sous-sections correspondantes.

4.1 Enquête sur la santé des Pays-Bas

L'enquête GEZO des Pays-Bas est menée tous les ans depuis 1981 par le Bureau central de la statistique des Pays-Bas à titre d'enquête transversale répétée dans le cadre de laquelle un échantillon de ménages est

interviewé dans le but de fournir un aperçu de l'évolution de la santé, de la consommation de soins de santé, du style de vie et des comportements préventifs de la population néerlandaise. La base de sondage est formée en tirant d'abord un échantillon de municipalités, puis un échantillon obtenu auprès des personnes vivant dans les municipalités sélectionnées. À compter de 2010, l'enquête est passée à un mode mixte faisant intervenir une réponse initiale en ligne, puis un suivi téléphonique (ou en personne); on communique par téléphone avec les non-répondants à l'enquête en ligne si leur numéro de téléphone est connu au registre, sinon on organise une interview en personne. Au cours de ces années, la taille de l'échantillon a augmenté pour s'établir à 15 000 personnes et le taux de réponse général a augmenté de 25 %. À partir de 2014, on est passé d'un mode de suivi mixte à une interview en personne. En 2018, toutefois, une partie des non-réponses en ligne a été traitée à l'aide d'une interview en personne de manière plus efficace. La propension à répondre à des interviews personnelles dans le cadre d'une série chronologique dépend fortement du résultat de la réponse en ligne; ainsi, la modélisation de la propension du suivi dépend du modèle de propension à répondre en ligne. Cet enjeu nécessite d'être pris en compte plus que l'interprétation de la variation dans le temps de la PR en ligne et dépasse l'objectif principal de la présente étude. Pour simplifier, nous nous concentrons, dans la présente étude, sur la modélisation de la propension à répondre en ligne comme étape initiale fondamentale; ainsi, la modélisation de la PR de suivi dans une série chronologique est plus adaptée à des recherches futures. Il est important de mentionner ici que seules les données de l'enquête GEZO en ligne de 2014-01 à 2019-10 inclusivement sont analysées dans la présente étude. Nous utilisons trois variables auxiliaires qui découlent de bases de sondage ou de registres administratifs. Les variables auxiliaires obligatoires sont l'âge, le sexe et l'ethnicité, qui répartissent la population ou son échantillon en 20 strates disjointes (voir l'annexe A pour obtenir de plus amples renseignements).

L'enquête GEZO menée sur de nombreuses années présente un plan d'enquête relativement constant. Cette caractéristique permet de valider l'étude de la dépendance de la PR dans le temps grâce aux nombreuses séries chronologiques. Nous nous intéressons aux données de réponse mensuelles, c'est-à-dire à la taille d'échantillon et au nombre de répondants pour chaque strate. Des prédictions sont faites tous les mois, mais peuvent également être agrégées par trimestre ou par année.

4.2 Quelles composantes d'une série chronologique contribuent le plus à la variation des propensions à répondre ?

Nous traitons cette question en deux étapes : tout d'abord, nous exécutons les combinaisons de modèles, puis nous comparons leurs rendements. La comparaison de plusieurs modèles se fait selon deux perspectives : 1) « Quelle combinaison est le mieux ajustée aux données de réponse ? » et 2) « Quelle combinaison génère les prédictions les plus fiables ? ». Nous utilisons des critères d'information et des valeurs p prédictives *a posteriori* pour mesurer le rendement de chaque modèle et ainsi rechercher le

modèle « optimal ». Le modèle présentant les critères d'information les plus bas et les valeurs p prédictives les plus proches de 0,5 est privilégié.

Puisque l'essai de toutes les combinaisons de composantes de (3.1) entraîne de lourds calculs, il est important de procéder à une recherche efficace du modèle « optimal ». Pour ce faire, nous ajustons les modèles aux données de réponse à l'aide de la stratégie suivante :

1. Commencer par le modèle de base (uniquement les variables auxiliaires).
2. Ajouter au modèle de base les effets fixes en séquence, la tendance temporelle linéaire et les tendances saisonnières.
3. Étudier si le modèle présenté au point 2 continue à s'améliorer avec les effets temporels généraux ou la saisonnalité générale.
4. Étudier si les tendances temporelles propres aux strates ou les effets saisonniers continuent d'améliorer le modèle.
5. Déterminer si un terme de bruit blanc représentant une variation inexplicée est nécessaire.
6. Étudier la robustesse pour des valeurs aberrantes à l'aide de différentes spécifications de loi *a priori* ou d'une structure dépendante du temps d'ordonnées à l'origine aléatoires générales au fil du temps.
7. Évaluer le modèle à l'aide de plusieurs diagnostics.

Le tableau 4.1 présente les résultats de la sélection. Les modèles à effets fixes (M1 à M3) se comportent moins bien que les modèles à effets mixtes en matière de compromis entre l'ajustement et la complexité, puisque ces derniers produisent des critères d'information (DIC, WAIC) inférieurs. La comparaison de M2/M3 et M1 indique que la pente du temps λ ou la saisonnalité δ entraîne une diminution des critères d'information (CI). Toutefois, le modèle continue de s'améliorer en introduisant la tendance générale u_t , car on observe une diminution importante des CI dans le M4 par rapport au M3. Comme le montrent les M5 et M6, l'amélioration se poursuit en ajoutant des ordonnées à l'origine aléatoires pour la strate v_g et les tendances temporelles propres aux strates $z_{g,t}$. Même si le bruit blanc $w_{g,t}$ semble apporter très peu au M7 globalement, les valeurs p prédictives *a posteriori* pour les variances indiquent qu'il est utile d'inclure ce bruit blanc. De plus, nous avons constaté qu'utiliser une tendance au niveau local (RW1) ou une tendance lisse (RW2) comme tendance générale u_t ne change pratiquement rien en ce qui a trait aux CI des modèles M6 à M11.

Enfin, la quatrième colonne du tableau 4.1 présente la loi *a priori* utilisée pour les coefficients de tendance générale u_t . Les lois *a priori* non normales mises à l'essai n'améliorent pas plus les CI, mais, du fait de queues de distribution plus lourdes, elles aident à combattre une valeur aberrante dans les données (problème exceptionnel survenu en février 2017).

Tableau 4.1
Résumé des modèles hiérarchiques de séries chronologiques envisagés

Modèle	Fixe	Aléatoire	Loi <i>a priori</i>	DIC	pDIC	WAIC	pWAIC	PPP	
								Moyenne	Variance
M1	β	-	-	7 511	7	7 518	13	0,501	0,006
M2	β, λ	-	-	7 415	8	7 421	15	0,503	0,051
M3	β, λ, δ	-	-	7 368	12	7 378	22	0,498	0,092
M4	β, δ	u_i	Normale	7 255	43	7 280	68	0,484	0,168
M5	β, δ	u_i, v_g	Normale	6 916	56	6 925	65	0,491	0,172
M6	β, δ	$u_i, v_g, z_{g,t}$	Normale	6 790	98	6 781	90	0,494	0,356
M7	β, δ	$u_i, v_g, z_{g,t}, w_{g,t}$	Normale	6 790	131	6 769	110	0,517	0,425
M8	β, δ	$u_i, v_g, z_{g,t}, w_{g,t}$	Laplace	6 790	133	6 768	111	0,503	0,397
M9	β, δ	$u_i, v_g, z_{g,t}, w_{g,t}$	Distribution T	6 790	130	6 769	110	0,492	0,391
M10	β, δ	$u_i, v_g, z_{g,t}, w_{g,t}$	Horseshoe	6 790	131	6 769	110	0,518	0,411
M11	β, λ, δ	$u_i, v_g, z_{g,t}, w_{g,t}$	Laplace	6 806	150	6 779	122	0,519	0,413

Notes : « - » indique aucun effet aléatoire ou aucune loi *a priori*.

DIC signifie Deviance Information Criterion (critère d'information de déviance); WAIC signifie Widely Applicable Information Criterion (critère d'information largement applicable); PPP signifie valeurs *p* prédictives *a posteriori*.

Les lois *a priori* de distribution T et horseshoe s'adapteront probablement mieux et seront probablement plus robustes par rapport aux valeurs aberrantes que les lois *a priori* normales et Laplace, comme le montre la comparaison de leur moyenne *a posteriori* de la tendance générale u_i qui figure à l'annexe B. De plus, la tendance au niveau local du M8 semble compenser légèrement les tendances de lissage du M11. Les valeurs *p* de la moyenne de M8 fournissent une valeur plus proche de 0,5 que les M9 et M11.

Pour déterminer le modèle étant souple par rapport aux valeurs aberrantes et celui générant une estimation fiable pour toute la série, nous examinons plus précisément l'écart entre les observations et les estimations fondées sur les modèles, en particulier M8, M9 et M11. Cette comparaison démontre que ces trois modèles présentent une capacité limitée à tenir compte de valeurs aberrantes. Les quantiles inférieurs tentent d'atteindre la valeur aberrante, mais ne parviennent pas à la traiter. De plus, la loi *a priori* de Laplace présente une incertitude légèrement inférieure à la loi de distribution T en matière d'estimations *a posteriori*, mais une incertitude de taille similaire au modèle de tendance lissée (voir l'annexe B).

4.3 Quel degré d'exactitude de prédiction de la propension à répondre peut être obtenu pour la nouvelle période suivante ?

Pour répondre à la deuxième question de recherche, nous estimons le niveau et la variation de la prédiction de la réponse générale pour la vague de collecte de données à venir. Le niveau estimé est l'écart entre la prédiction de la propension *a posteriori* attendue et le taux de réponse atteint, alors que la variation estimée désigne l'exactitude de la prédiction de la propension à répondre générale. Nous mesurons également l'équilibre entre le niveau et la variation et le comparons avec la référence de (3.9). Cette

évaluation nous permet de vérifier s'il est possible d'améliorer notre méthode. Il est possible de prendre des mesures pour adapter ou maintenir la collecte de données de la vague suivante, une fois l'amélioration connue à partir de la série historique.

Nous soulignons que l'analyse est effectuée en fonction du modèle « optimal », M8. Pour toutes les strates d'un nouvel échantillon par mois, les mois à partir de janvier 2014 jusqu'au mois présent (mais celui-ci n'étant pas compris) sont considérés comme la série chronologique historique, qui est utilisée pour entraîner M8. Nous utilisons ensuite le nouvel échantillon du mois actuel pour le critère prédictif estimé. La série chronologique historique est cumulée et des critères prédictifs sont mis à jour avec la nouvelle vague. L'évaluation continue se termine avec 2019-09; en effet, un mois doit demeurer pour l'exercice de prédiction, car 2019-10 correspond au dernier mois de données disponibles. Pour conférer de la robustesse à l'incidence de la taille historique sur le rendement prédictif, nous laissons la série chronologique historique commencer avec 60 mois (de 2014-01 à 2018-12) comme essai initial par défaut.

Le tableau 4.2 montre que l'incertitude *a posteriori* des prédictions de la propension à répondre générale diminue régulièrement, mais lentement et converge vers 0,027 environ. Du fait de la variation d'échantillonnage inhérente au terme de biais, la tendance du biais est erratique et présente une légère diminution au mieux. Par rapport aux taux de réponse réalisés, l'écart le plus important des moyennes *a posteriori* est d'environ 0,07 en janvier et juin et l'écart le plus petit est d'environ 0,04 en mars, mai, août et octobre. Les résultats de la racine carrée de l'erreur quadratique moyenne (REQM) varient avec le terme de biais tout au long des mois, car les écarts types estimés sont bien inférieurs aux biais estimés. La REQM enregistre une valeur maximale de 0,084 en janvier, ce qui est probablement attribuable aux mois comportant des valeurs aberrantes du début de 2017. Même si le modèle réagit à cette perturbation, cette dernière a une incidence négative sur le rendement des prédictions obtenues au cours de ce mois. Outre janvier, pour certains mois, la REQM estimée est proche de la référence $REQM_{\min}$. On peut en conclure que l'exactitude estimée est relativement proche de l'exactitude possible maximale.

Tableau 4.2

Prédiction un mois avant de trois mesures de la propension à répondre (PR) pour les strates : biais, écart type (ET) et racine carrée de l'erreur quadratique moyenne (REQM) par rapport à la référence ($REQM_{\min}$)

	2019									
	Janv.	Févr.	Mars	Avr.	Mai	Juin	Juill.	Août	Sept.	Oct.
Biais	0,078	0,064	0,045	0,062	0,046	0,077	0,063	0,049	0,058	0,048
ET	0,030	0,031	0,031	0,029	0,028	0,029	0,028	0,027	0,028	0,027
REQM	0,084	0,071	0,055	0,069	0,054	0,082	0,068	0,056	0,065	0,055
$REQM_{\min}$	0,055	0,056	0,055	0,055	0,055	0,055	0,048	0,048	0,049	0,049

Note : La colonne indique le mois actuel d'évaluation du rendement de la prédiction.

4.4 Dans quelle mesure l'exactitude de la prédiction varie-t-elle pour les strates de population ?

Cette question de recherche porte sur les différentes strates et le rendement du modèle à prédire la PR par strate. À cette fin, nous considérons la REQM au niveau de la strate ainsi que ses deux composantes,

c'est-à-dire le biais et l'écart type. Les mesures d'évaluation sont la moyenne des prédictions sur les 10 mois précédents. Le mois 2019-10 est le dernier disponible. Cette anticipation de presque un an permet au personnel de la collecte des données de planifier des plans adaptatifs bien longtemps à l'avance.

De façon similaire à la section 4.3, nous limitons l'analyse aux hypothèses. Le modèle retenu est sélectionné à partir de la section 4.1 et la série chronologique historique est fixée à 60 mois (de 2014 à 2018). Pour chaque strate, le modèle est entièrement entraîné par les données historiques fixes et permet des déductions pour les mois restants en 2019.

Le tableau 4.3 présente les critères de prédiction pour chaque strate ainsi que la référence. Les termes de biais estimés varient largement entre les strates. Le plus important écart de l'attente *a posteriori* par rapport aux taux de réponse réalisés a lieu pour les strates 8, 10, 12 et 18, qui présentent toutes des biais supérieurs à 0,1. Par rapport aux biais, une variation relativement lisse des écarts types estimés d'environ 0,03 est observée, lorsque la strate 4 présente l'incertitude la moins importante par rapport aux estimations *a posteriori* (0,018).

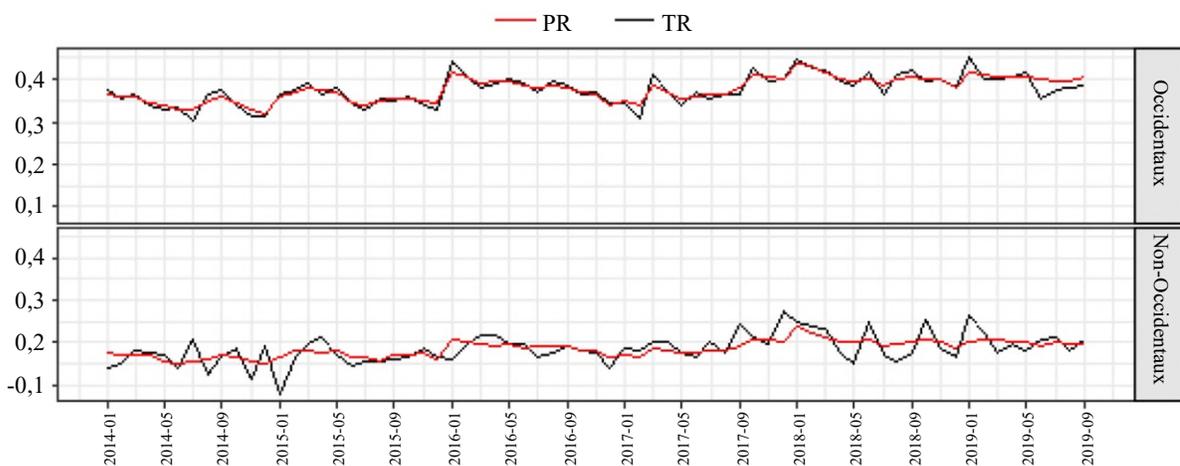
Certaines strates ayant des biais plus importants peuvent présenter une moindre exactitude des estimations *a posteriori* de la PR que les strates présentant des propensions moins biaisées. De façon similaire, plus la prédiction est biaisée, plus on estime une REQM élevée. Les biais estimés sont en effet bien supérieurs aux écarts types estimés. Il n'est pas étonnant que la strate 10 enregistre la valeur de REQM la plus importante, pour laquelle la prédiction est la plus biaisée et la précision la moins bonne. Les résultats de REQM peuvent rattraper voire être comparables ou supérieurs à la référence. Lorsque le modèle génère une prédiction pour la strate 20, par exemple, des améliorations plus importantes peuvent être obtenues que pour d'autres strates.

Tableau 4.3
Moyenne de la prédiction 10 mois d'avance de trois mesures pour chaque strate : biais, écart type (ET) et racine carrée de l'erreur quadratique moyenne (REQM) par rapport à la référence (REQM_{min})

	Biais	ET	REQM	REQM _{min}
1	0,045	0,030	0,060	0,046
2	0,066	0,030	0,077	0,094
3	0,028	0,025	0,039	0,039
4	0,049	0,018	0,053	0,061
5	0,035	0,026	0,045	0,035
6	0,047	0,021	0,053	0,064
7	0,062	0,032	0,073	0,054
8	0,105	0,030	0,111	0,154
9	0,047	0,031	0,060	0,045
10	0,165	0,035	0,173	0,160
11	0,044	0,031	0,057	0,048
12	0,134	0,030	0,138	0,092
13	0,030	0,027	0,044	0,042
14	0,081	0,022	0,086	0,074
15	0,044	0,028	0,056	0,038
16	0,067	0,022	0,072	0,072
17	0,030	0,031	0,048	0,053
18	0,114	0,029	0,120	0,146
19	0,031	0,029	0,046	0,041
20	0,095	0,030	0,105	0,172

Le rendement prédictif affiche une différence significative entre les strates, lorsqu'il n'y a qu'une caractéristique différente. La REQM de la strate 20, par exemple, est inférieure de 0,06 à celle de la strate 10. Cela semble sous-entendre que les groupes de femmes pourraient présenter un biais ou une variance inférieurs aux groupes d'hommes lorsque les personnes non occidentales âgées de plus de 64 ans sont prises en compte. Étant donné l'âge et l'ethnicité des groupes et comparativement aux groupes non occidentaux (lignes paires), les résultats de la REQM sont bien inférieurs pour les groupes occidentaux (lignes impaires). Pour valider cette supposition, certaines strates sont combinées en sous-groupes présentant des caractéristiques moins détaillées. Comme le montre la figure 4.1, le modèle fournit de meilleures prédictions pour le groupe occidental que pour les groupes non occidentaux, puisque les estimations *a posteriori* attendues atteignent principalement la réponse observée par mois. Le rendement comparatif pour les groupes d'âge ou de sexe est présenté en annexe B.

Figure 4.1 Moyennes *a posteriori* mensuelles de la propension à répondre (PR) agrégées pour les groupes ethniques par rapport aux taux de réponse (TR) observés des groupes ethniques.



Note : Mois de 2014-01 à 2018-12 pour le modèle estimé et mois de 2019-01 à 2019-10 pour les prédictions de PR.

4.5 Dans quelle mesure l'exactitude de la prédiction dépend-elle de la longueur de la série chronologique d'enquête historique ?

La principale préoccupation pour cette question est de déterminer la robustesse du rendement de la prédiction pour la quantité de séries chronologiques historiques utilisées aux fins d'entraînement du modèle et de prédiction. À cette fin, nous poursuivons avec la moyenne de prédictions de REQM sur trois mois et ses deux termes (le biais et l'écart type) au niveau général à un moment donné quelconque. Nous appelons cette moyenne fondée sur cette longueur la moyenne trimestrielle. Pour étudier l'incidence de la taille des données historiques, nous procédons à une validation croisée des séries chronologiques divisées par trois sur l'ensemble de données, c'est-à-dire un ajout successif de trois mois de nouvelles données à l'ensemble

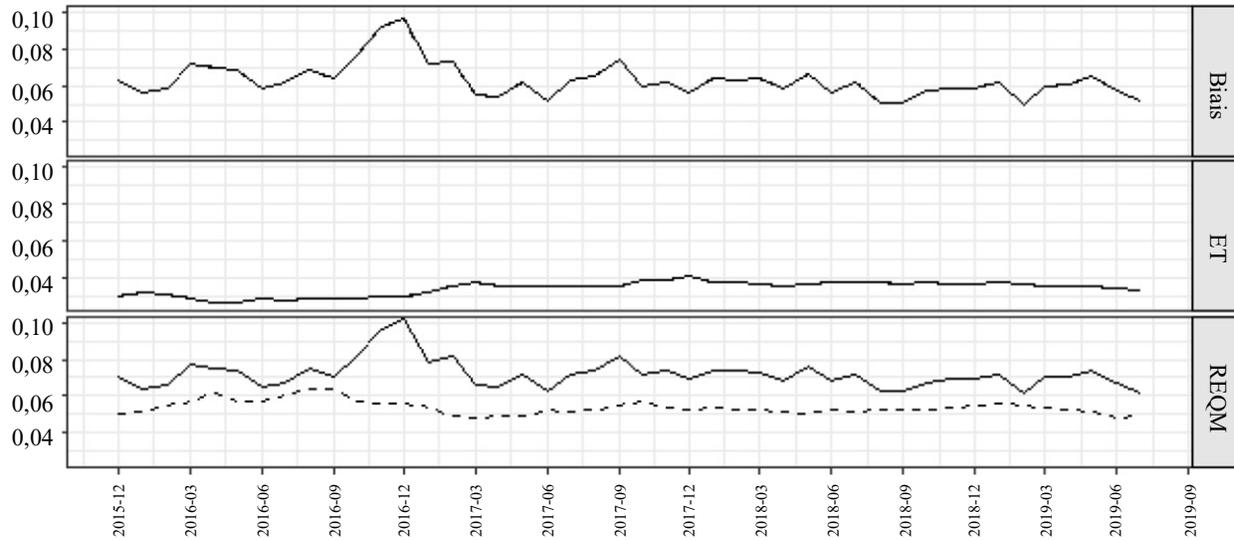
de données d'entraînement utilisé pour les prédictions fondées sur le modèle. Cette analyse est répétée en continu et la stratégie étape par étape est établie comme suit :

1. Sélectionner les composantes du modèle basées sur la série chronologique dans son ensemble.
2. Sélectionner l'ensemble de référence des périodes historiques de longueur t . Diviser la fenêtre comme suit : l'ensemble d'entraînement D_o des premières périodes $t - 3$ et l'ensemble d'essai D_t des trois dernières périodes.
3. Utiliser les données D_o pour entraîner le modèle sélectionné, au moyen d'une simulation à partir de la distribution *a posteriori* de tous les paramètres du modèle, en connaissant D_o .
4. En fonction du modèle simulé en 3, les moyennes et variances prédictives *a posteriori* sont calculées pour les PR de chaque strate et de chaque moment donné de l'ensemble d'essai D_t .
5. En fonction des prédictions de PR individuelles mentionnées au point 4, calculer le biais général et l'écart type à l'aide de la proportion de l'échantillon $d_{g,t}$ pour la strate g au temps t comme poids; alors l'équation (3.8) permet de calculer la moyenne trimestrielle de la REQM.
6. Étendre la fenêtre à $t + 1$, avec une prévision sur une période supplémentaire. Répéter les étapes 3 à 5 pour mettre à jour les prédictions de la REQM et ses deux termes.
7. Répéter 6 jusqu'à ce que la longueur t soit la dernière période disponible.

Nous soulignons qu'il est nécessaire d'utiliser au moins deux ans comme longueur d'entraînement initiale, lorsque des composantes saisonnières sont incluses. Dans notre cas, les périodes sont des mois et la série chronologique s'étend jusqu'au 2019-07.

À la figure 4.2, les estimations de la REQM, des biais et de l'écart type dépendent du temps et sont calculées pour les strates trois mois à l'avance. La fenêtre de référence des données d'entraînement pour ajuster le modèle optimal va de 2014-01 à 2015-12. Pendant la fenêtre de la série chronologique, les résultats des biais sont supérieurs à environ deux fois les écarts types. Par conséquent, les résultats de REQM sont dominés par les biais et présentent la même volatilité. Au bout du compte, leurs estimations sont d'environ 0,06, alors que les estimations des écarts types enregistrent une légère augmentation. Ce dernier point est un peu étonnant, car l'on s'attendrait normalement à ce qu'utiliser une série chronologique plus longue pour estimer le modèle diminue les erreurs-types de prédiction *a posteriori*. Il s'avère, toutefois, que deux événements ont eu une importante incidence sur le rendement de la prédiction. Tout d'abord, au début de 2017, la collecte des données a été interrompue du fait de problèmes techniques liés au serveur Web. Cet incident a eu d'importantes répercussions immédiates sur la PR et, par conséquent, sur le rendement de prédiction du modèle. Ensuite, en 2018, des incitatifs conditionnels ont été introduits et le questionnaire de l'enquête a été remanié pour pouvoir être utilisé à partir d'un téléphone intelligent. Cette intervention sur le plan a eu un effet plus progressif et durable.

Figure 4.2 Moyenne mobile en avant d'un pas du biais trimestriel (volet supérieur), écart type trimestriel (volet du milieu) et racine carrée de l'erreur quadratique moyenne (REQM) trimestrielle (volet inférieur). REQM (ligne pleine) par rapport à la REQM de référence (ligne pointillée) lorsque la longueur de l'ensemble de données d'entraînement se déplace le long de l'axe des x.



Les résultats des biais et de la REQM enregistrent une importante hausse entre le 2016-10 et le 2017-02. Lorsque la fenêtre d'entraînement arrive au point 2016-10, la fenêtre d'essai commence à inclure les données de 2017-01 pour lesquelles les PR ont accusé une baisse. Leurs courbes continuent de croître et atteignent des maximums autour de 0,1, lorsque la fenêtre d'entraînement passe à 2016-12 et la fenêtre d'essai passe d'abord au mois « stable » 2017-03, lorsque les courbes des biais et de la REQM redescendent vers environ 0,05. Au cours de ces mois, on peut également observer une légère augmentation progressive de l'écart type.

L'inclusion de ces mois aberrants influe sur l'exactitude des prédictions en 2017. Entre les mois 2017-03 et 2017-12, les courbes des biais et de REQM sont plus volatiles et diminuent uniquement après 2018-01. Les écarts types présentent une tendance à la hausse entre 2017-03 et 2017-12 et diminuent à peine.

En 2018, l'incidence sur la propension à répondre de l'intervention sur le plan a été bien plus modeste qu'en 2017, mais comme elle est structurelle, elle a en fait une incidence sur l'exactitude de la prédiction en 2018 et en 2019. Ces deux événements, l'un technique et l'autre relevant d'un changement de plan, sont réalistes dans les pratiques d'enquête et peuvent avoir des répercussions désastreuses lorsqu'ils ne sont pas pris en considération. Nous présentons un exemple de la façon de les gérer. Les efforts supplémentaires pour cela sont les suivants :

- *Méthode de rejet.* Pour les données initiales, supprimer les chiffres de réponse $r_{g,t}$ dans 2017-01 et 2017-02, puis les traiter comme des données manquantes. Imputer ces données manquantes $r_{g,t}$ par les moyennes *a posteriori* des réponses simulées à partir de la distribution prédictive *a posteriori*. Il convient de mentionner que dans la section 4.2, nous avons expliqué qu'utiliser des lois *a priori* particulières non normales pour les composantes de la série chronologique peut

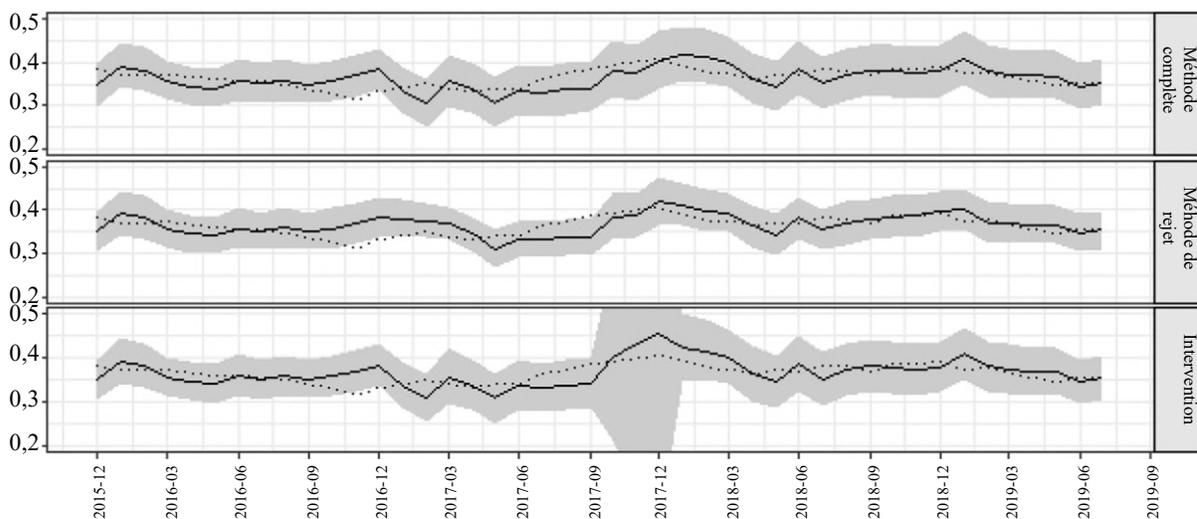
également limiter l'effet de valeurs aberrantes. Si une valeur aberrante est assez extrême et qu'il est connu qu'elle a lieu à un moment particulier, il peut toutefois être préférable de la rejeter.

- *Méthode d'intervention.* Inclure un terme d'intervention dans le modèle saisissant la potentielle modification structurelle. Ajouter des variables d'intervention binaires à la série de données initiales et les définir à 0-1, où, dans notre cas, elles prendraient la valeur 1 et s'activeraient à partir de 2018-01. Les effets potentiels relatifs à l'intervention pourraient être un effet fixe unique, un effet aléatoire propre à la strate ou les deux.

Les résultats de l'application de ces deux méthodes séparément sont présentés à la figure 4.3.

Ces deux méthodes ont un effet évident sur les prédictions. Dans la période allant de 2016-12 à 2017-05, où s'arrête la fenêtre de la série chronologique d'entraînement, les moyennes *a posteriori* de la méthode de rejet présentent une tendance à la baisse, par rapport aux moyennes *a posteriori* du modèle initial (« complet » de la figure 4.3). Toutefois, entre 2017-10 et 2018-03, la différence dans les moyennes fondées sur le mode et dans les observations diminue dans la méthode de rejet. De plus, la méthode de rejet fait diminuer l'incertitude relative aux moyennes *a posteriori*, à mesure que l'intervalle de crédibilité devient plus étroit à partir de 2016-12. L'incidence de l'intervention sur la propension à répondre générale ne peut pas être bien estimée au moyen de quelques nouveaux mois seulement.

Figure 4.3 Moyenne mobile en avant d'un pas des estimations *a posteriori* trimestrielles de la propension à répondre générale selon trois scénarios : 1) série chronologique initiale (volet supérieur); 2) nouvelle série chronologique en rejetant les données du début de 2017 (volet du milieu) et 3) nouvelle série chronologique en ajoutant les effets d'intervention (volet inférieur).



Note : Par rapport à la moyenne mobile trimestrielle des taux de réponse observés (ligne pointillée), les estimations fondées sur le modèle sont résumées par les moyennes *a posteriori* (ligne pleine) avec un intervalle de crédibilité à 95 % (bande grise). L'axe des X désigne la longueur des données d'entraînement de 2014 pour ce moment donné.

Même si nous n'avons pas l'intention de fournir une explication détaillée des options de modélisation de changements accidentels et structuraux, le modèle de séries chronologiques que nous proposons peut être modifié d'une manière relativement simple et souple. Une réplique avec une longue série chronologique d'enquête s'impose pour se faire une idée des meilleures options.

5. Discussion

Une prédiction exacte et fiable des propensions à répondre (PR) est la clé pour améliorer et optimiser des plans d'enquête adaptatifs. Une telle inférence peut être compliquée par une variation saisonnière et des tendances temporelles potentiellement propres à des strates de population. Dans la présente étude, nous introduisons un modèle bayésien hiérarchique de séries chronologiques pour fournir des prédictions de PR au niveau de la strate. Ce modèle est suffisamment souple pour inclure une variation saisonnière, diverses formes de tendances, des modifications du plan d'enquête et une dépendance aux strates, de sorte qu'il peut faciliter la préparation du plan d'enquête adaptatif dans un environnement d'enquête changeant. Ces modèles sont obtenus à partir de données d'enquête historiques et mis à jour avec de nouvelles données d'enquête.

Dans la présente étude, nous appliquons cette méthode à une enquête répétée sur une population générale, l'enquête sur la santé menée aux Pays-Bas du Bureau central de la statistique des Pays-Bas, afin de fournir un soutien empirique pour un cas réaliste. Nous nous concentrons principalement sur l'amélioration des prédictions de PR au niveau de la strate soumises à des facteurs temporels. Selon diverses combinaisons de modèles comprenant ces facteurs, l'un de nos objectifs concrets est de rechercher le modèle le plus efficace offrant un compromis entre l'ajustement du modèle et la facilité du calcul. Le modèle optimal est sélectionné en fonction de critères permettant d'évaluer le rendement (CI élevés, valeurs $p \approx 0,5$) et une capacité prédictive ($REQM_{PR}$ basse). Ces mesures fournissent une information précieuse sur l'amélioration relative obtenue en ajoutant de nouveaux facteurs. Cette approche souple permet à d'autres spécialistes de la recherche sur les enquêtes de tenir compte de différents facteurs temporels et de choisir finalement le modèle privilégié selon leurs paramètres.

Les autres objectifs de la présente étude consistent à évaluer le rendement de la prédiction des PR dans leur ensemble et au niveau de la strate selon le modèle retenu. Nous utilisons des mesures prédictives, en particulier la racine carrée de l'erreur quadratique moyenne (REQM), pour évaluer l'incertitude des prédictions. Cela nous permet de comparer directement : 1) la réponse prévue générale avec les premières données à venir; 2) des moyennes annuelles de la réponse prévue pour chaque strate; et 3) des moyennes trimestrielles de la réponse prévue générale. Nous évaluons finalement le rôle de la longueur de la série chronologique d'enquête historique pour le modèle privilégié et dans un modèle réoptimisé lorsque des données sont reçues. Ce faisant, nous pouvons déterminer le moment propice pour commencer à mettre en œuvre un plan d'enquête adaptatif. Il convient de mentionner que lorsque le plan de l'enquête devient adaptatif, il devient moins évident de savoir comment connaître le changement temporel des paramètres du modèle. De plus, le modèle de séries chronologiques lui-même peut nécessiter une mise à jour selon le type d'adaptation d'enquête.

Alors que notre tentative constitue un premier pas vers des plans d'enquête adaptatifs, il existe toutefois divers points à prendre en compte sur les plans méthodologique et pratique. Tout d'abord, notre approche s'applique à une enquête transversale fréquemment répétée. Les données historiques de telles enquêtes représentent une importante source d'estimations relativement robustes de coefficients de modèles et de prédictions fiables. Lorsqu'il s'agit d'une nouvelle enquête ou d'une enquête peu fréquente d'un bureau de la statistique, notre approche peut être moins efficace. Deuxièmement, nous supposons que la stratification

est effectuée sur un modèle entièrement saturé, c'est-à-dire que les strates sont prédéfinies par des variables auxiliaires qui sont des prédicteurs robustes de réponse en ligne. Comment varie le rendement de la prédiction lors de l'ajout de variables auxiliaires moins influentes ? Il est important d'évaluer la sensibilité des prédictions fiables en fonction du choix de variables auxiliaires. De plus, nous supposons que les strates sont fixes tout au long de la série chronologique. Dans les pratiques d'enquête, les données auxiliaires sélectionnées peuvent varier progressivement dans le temps, et donc, la pertinence de certaines strates également. Par conséquent, il est essentiel de tenir compte d'une variation de la stratification liée aux données auxiliaires lors de la prédiction de la réponse. Troisièmement, nous supposons que le plan de l'enquête ne change pas au fil du temps, c'est-à-dire que les hypothèses du modèle doivent être valides tout au long de la série chronologique. En cas d'intervention ou de l'introduction d'un autre mode d'autodéclaration (par exemple l'utilisation de téléphones intelligents), il convient d'inclure explicitement la variation des réponses que cela entraîne. Le recours à une méthodologie avancée est nécessaire, car il n'existe alors pas de connaissances historiques antérieures relatives à une modification du plan. L'inclusion d'un tel changement dans le modèle peut entraîner une importante variation et, avant que le modèle ne puisse fournir des renseignements sur les effets du changement sur les PR, un échantillonnage historique pour une période suffisamment longue doit d'abord être obtenu.

Nous constatons également des limites dans la méthodologie proposée. Pour une année en particulier des données de l'enquête sur la santé réalisée aux Pays-Bas, nous observons une augmentation soudaine des écarts types de la propension à répondre prévue et des indicateurs de la qualité généraux. Cette hausse était attribuable à une intervention (les téléphones intelligents ont été introduits comme dispositifs de réponse ainsi que des incitatifs conditionnels). Les résultats montrent que le modèle peut être sensible à un changement de plan. Par conséquent, tenir compte des changements apportés au plan est nécessaire et réduira temporairement la performance prédictive.

De futurs travaux de recherche doivent aborder les prédictions de réponse conditionnelles dans le cadre de plans d'enquête en mode mixte. Dans la présente étude, nous nous sommes uniquement concentrés sur des prédictions de réponse en mode unique. De telles considérations sont utiles pour optimiser les décisions en matière de plan d'enquête adaptatif, comme passer, par exemple, à un mode plus ou moins coûteux en fonction du budget. Notre méthode ouvre la voie à l'élaboration de tels modèles conditionnels.

Actuellement, le modèle proposé est conçu pour les enquêtes transversales répétées, mais il peut être étendu à d'autres plans d'enquête et plans de sondage, comme le renouvellement de panel. Une telle extension signifierait qu'il faudrait ajouter au vecteur de modèle des propensions des panels à répondre ou des propensions à l'attrition et revoir la structure de corrélation entre les propensions.

Remerciements

Ce projet est généreusement soutenu par la bourse du conseil de Chine, le Bureau central de la statistique des Pays-Bas et l'université d'Utrecht. Nous aimerions remercier Anouk Roberts pour les échanges fructueux.

Annexe A

Tableau A.1

Les variables auxiliaires forment 20 strates et la saison est jugée être un facteur qui influe sur la prédiction de la propension à répondre

		Catégorie
Variable auxiliaire	Sexe	Hommes Femmes
	Âge	Enfants (≤ 17 ans) Jeunes (18 à 34 ans) Personnes d'âge moyen (35 à 54 ans) Personnes âgées (55 à 64 ans) Retraités (≥ 65 ans)
	Ethnicité	Occidentaux (y compris, de souche et première et deuxième générations d'Occidentaux) Non-Occidentaux (y compris première et deuxième générations de non-Occidentaux)
Variable	Saison	Hiver (janvier et février) Printemps (mars à mai) Été (juin à août) Automne (septembre à novembre) Noël (décembre)

Annexe B

Figure B.1 Moyennes *a posteriori* des tendances temporelles générales u_t , selon M7 à M10.

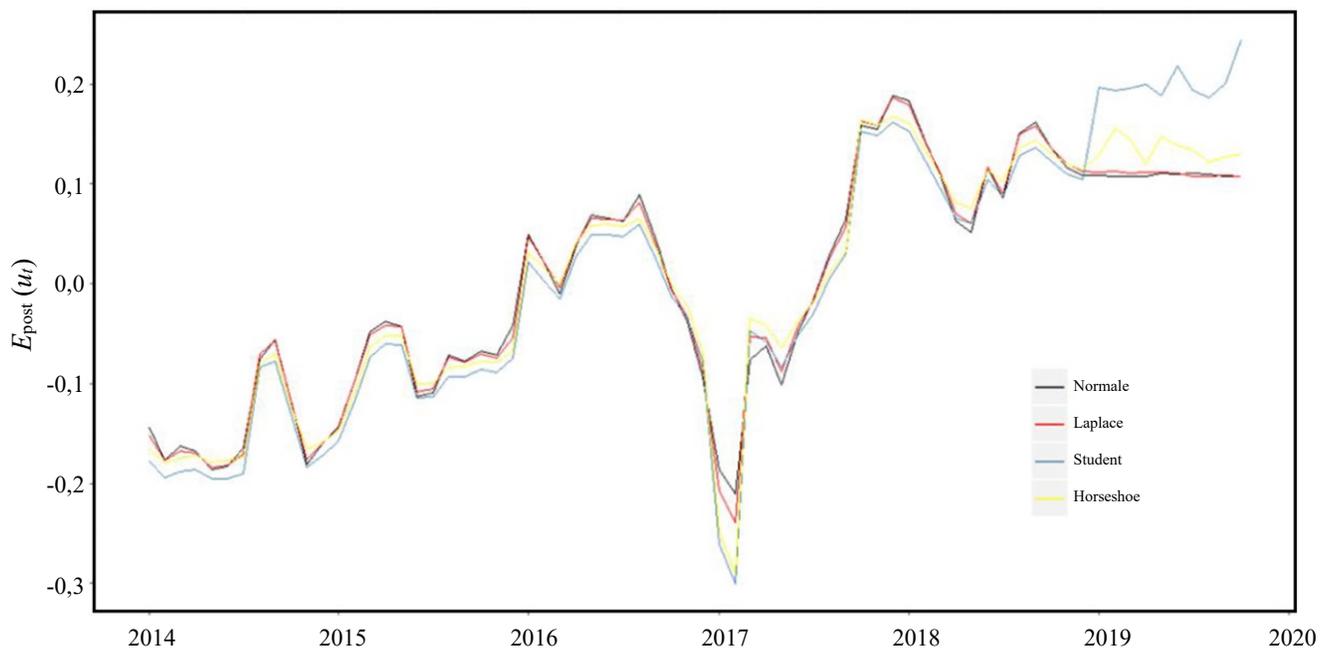
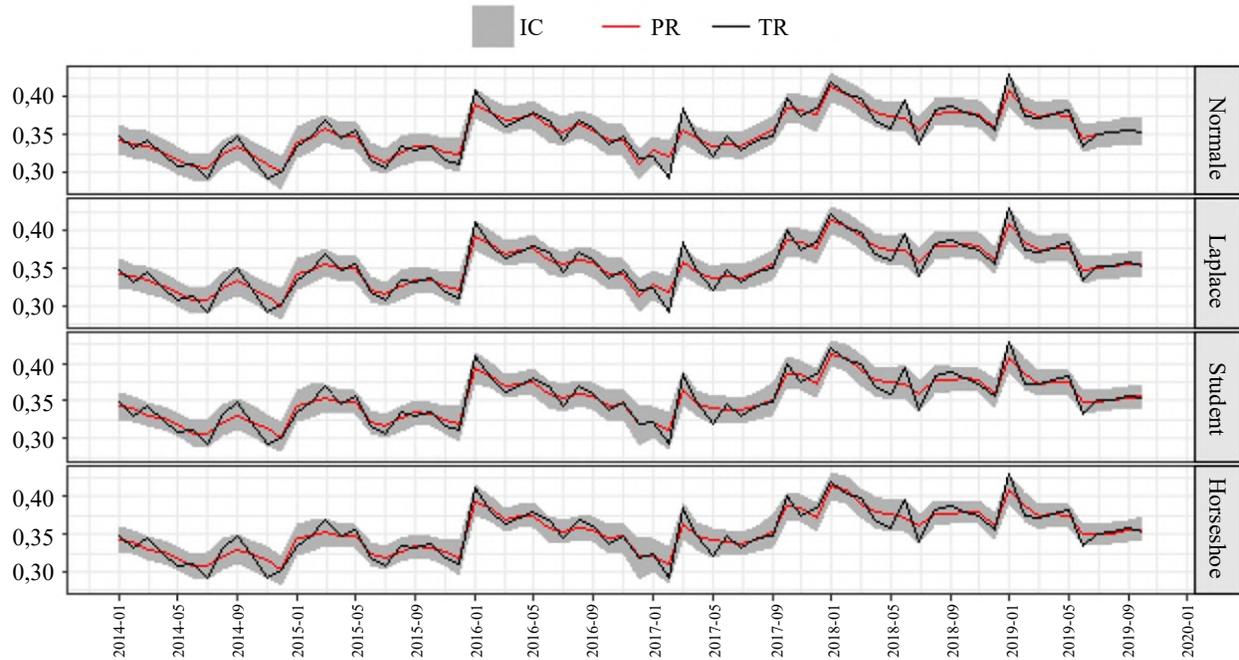
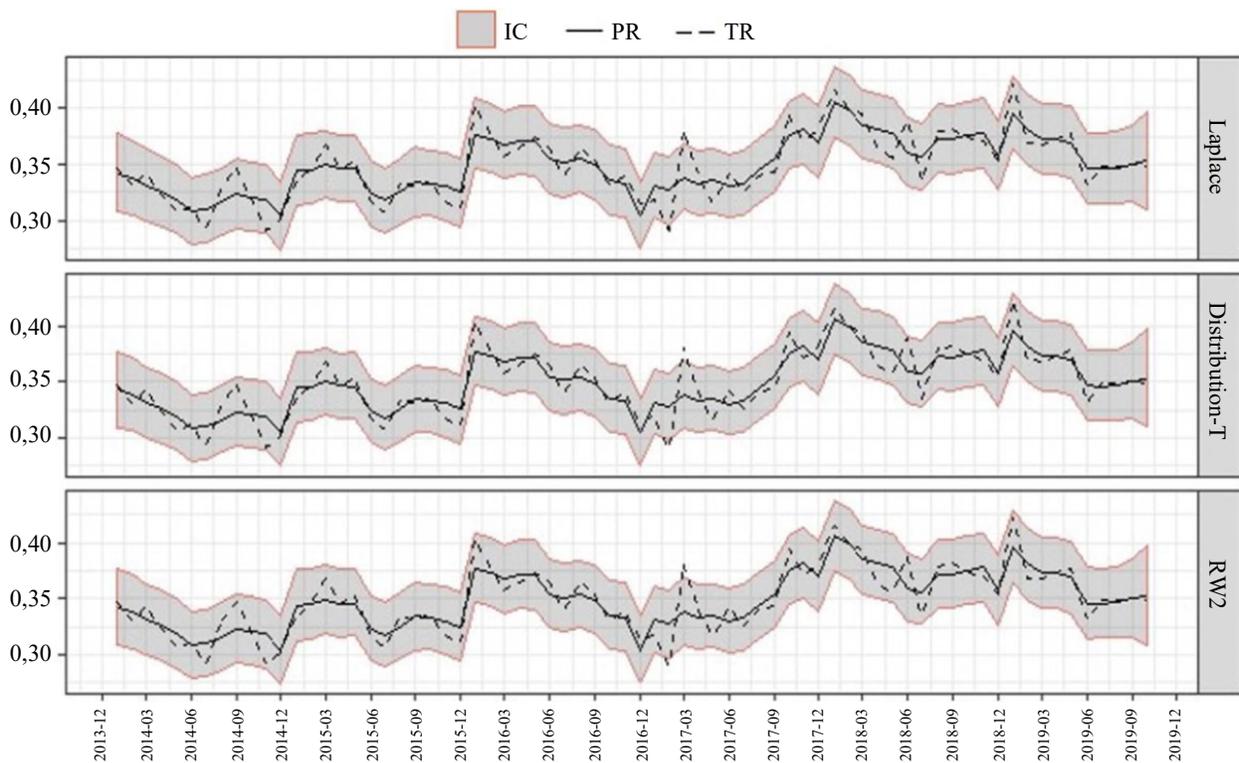


Figure B.2 Comparaison des prédictions *a posteriori* de la propension à répondre (PR) pour des strates selon quatre modèles (M7 à M10) et des TR observés et sélection du modèle le plus compatible avec les valeurs aberrantes observées.



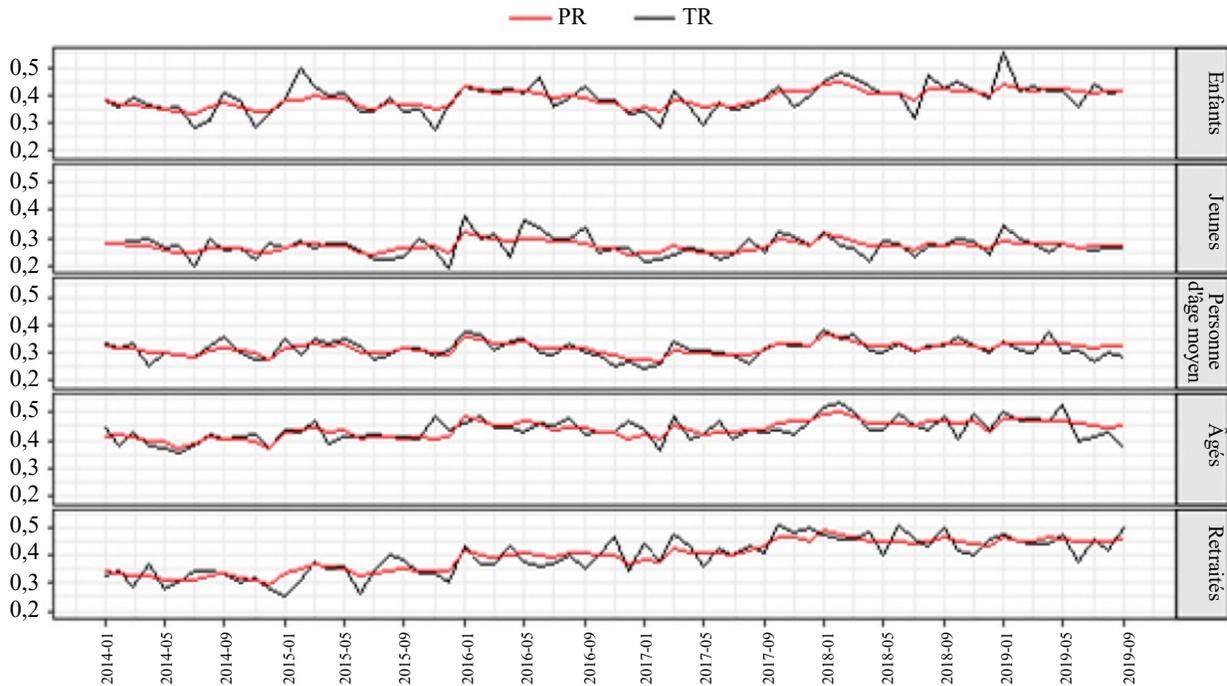
Note : Les prédictions de la PR générale sont résumées comme les moyennes *a posteriori* (PR) et l'intervalle de crédibilité à 95 % (IC).

Figure B.3 Prédictions *a posteriori* de la propension à répondre (PR) pour des strates par rapport aux TR observés selon M8, M9 et M11.



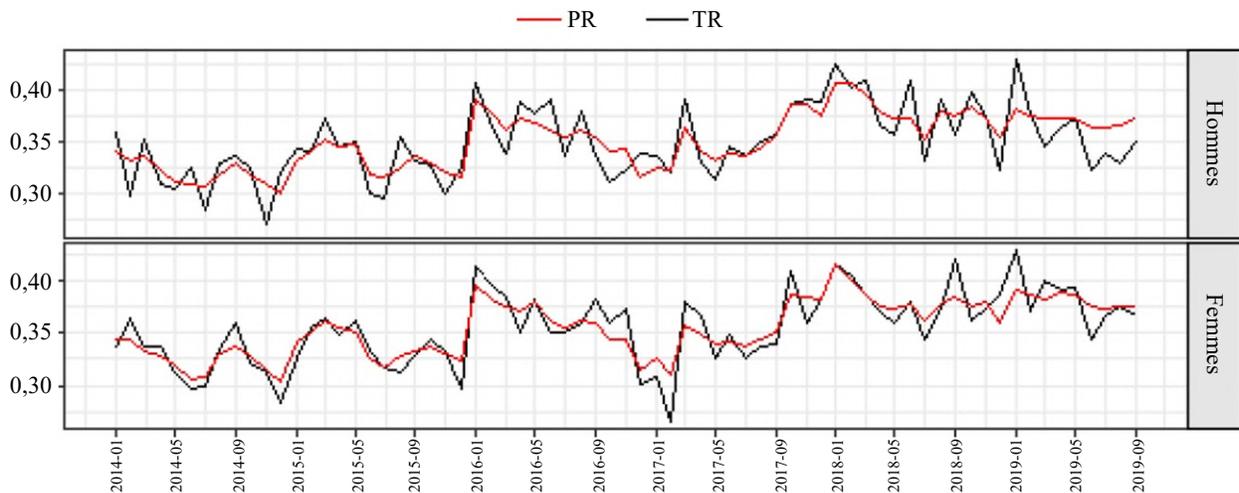
Note : Les prédictions de la PR générale sont résumées comme les moyennes *a posteriori* (PR) et l'intervalle de crédibilité à 95 % (IC).

Figure B.4 Moyennes *a posteriori* mensuelles de la propension à répondre (PR) pour les groupes d'âge par rapport aux taux de réponse (TR) observés des groupes d'âge.



Note : Mois de 2014-01 à 2018-12 pour un ajustement du modèle et mois de 2019-01 à 2019-10 pour des prédictions de la PR.

Figure B.5 Moyennes *a posteriori* mensuelles de la propension à répondre (PR) pour les hommes et les femmes par rapport aux taux de réponse (TR) observés pour chaque sexe.



Note : Mois de 2014-01 à 2018-12 pour un ajustement du modèle et mois de 2019-01 à 2019-10 pour des prédictions de la PR.

Annexe D

Les modèles hiérarchiques binomiaux de séries chronologiques sont ajustés au moyen d'un échantillonneur de Gibbs. Pour la dérivation de l'ensemble de distributions conditionnelles complètes, nous nous reportons à (l'annexe de) la version du rapport technique de Boonstra et van den Brakel (2022). L'échantillonneur de Gibbs y est obtenu pour une catégorie générale de modèles hiérarchiques, qui comprend l'ensemble de modèles dont il est question dans la présente étude, à l'exception du fait que nous employons ici un modèle binomial plutôt qu'une distribution de données gaussienne. Heureusement, l'utilisation de l'approche d'augmentation de données à combinaison d'échelles pour des modèles logistiques binomiaux (Polson, Scott et Windle, 2013) permet de s'assurer que les mêmes distributions conditionnelles complètes de forme fermée comme dans le cas gaussien puissent être utilisées avec seulement des modifications minimales de leurs paramètres, ainsi qu'une distribution conditionnelle complète supplémentaire des facteurs d'échelle latents auxiliaires. Pour commencer avec cette dernière, la distribution conditionnelle complète pour le facteur d'échelle ω_i est fournie par :

$$p(\omega_i | r, \cdot) = \text{PG}(\omega_i | n_i, \theta_i)$$

indépendamment de toutes les valeurs de i . Pour simplifier la notation, nous utilisons l'indice i plutôt que le double indice g, t utilisé dans le texte principal, et r désigne le vecteur de réponse observée complet. Ici θ_i est la variable explicative linéaire et $\text{PG}(\omega_i | n_i, \theta_i)$ désigne la distribution de Pólya-Gamma avec les paramètres n_i et θ_i (voir Polson, Scott et Windle, 2013). Les distributions conditionnelles complètes des coefficients présentent des changements uniquement dans leurs paramètres. Par exemple, dans la distribution conditionnelle complète pour une composante d'effets aléatoires généraux, l'équation (A.28) du rapport technique, la matrice de précision Σ^{-1} , devient $\Sigma^{-1} = \text{diag}(\omega)$ et le vecteur de réponse y est remplacé par la « réponse de travail » $\frac{r-n/2}{\omega}$. Il en est de même pour les distributions conditionnelles complètes des effets fixes et des paramètres auxiliaires ξ . Toutes les autres distributions conditionnelles complètes demeurent inchangées.

Bibliographie

- Boonstra, H. (2021). mcmcsae: MCMC small area estimation. *R Package Version 0.6.0*. <https://cran.r-project.org/web/packages/mcmcsae/index.html>.
- 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](#). *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>.
- Boonstra, H.J., et van den Brakel, J.A. (2022). Multilevel time-series models for small area estimation at different frequencies and domain levels. Accepté pour publication dans *l'Annals of Applied Statistics*. Rapport technique 2018-12, <https://www.cbs.nl/en-gb/background/2018/50/models-for-estimation-at-various-aggregation-levels>, Statistics Netherlands.

- Burger, J., Perryck, K. et Schouten, J.G. (2017). Robustness of adaptive survey designs to inaccuracy of design parameters. *Journal of Official Statistics*, 33(3), 687-708. <https://doi.org/10.1515/jos-2017-0032>.
- Carvalho, C.M., Polson, N.G. et Scott, J.G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2), 465-480. <https://doi.org/10.1093/biomet/asq017>.
- Chun, A.Y., Heeringa, S. et Schouten, J.G. (2018). Responsive and adaptive design for survey optimization. *Journal of Official Statistics*, 34(3), 581-597. <https://doi.org/10.2478/jos-2018-0028>.
- Coffey, S., West, B.T., Wagner, J. et Elliott, M.R. (2020). What do you think? Using expert opinion to improve predictions of response propensity under a Bayesian framework. *Methods Data Analyses*, 14(2), 159-194. <https://doi.org/10.12758/mda.2020.05>.
- Datta, G.S., Lahiri, P., Maiti, T. et Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the US. *Journal of the American statistical Association*, 94(448), 1074-1082.
- Durbin, J., et Koopman, S.J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Fang, Q., Burger, J., Meijers, R. et van Berkel, K. (2020). The role of time, weather and google trends in understanding and predicting Web survey response. *Survey Research Methods*, 15(1), 1-25. <https://doi.org/10.18148/srm/2021.v15i1.7633>.
- Franco, C., et Bell, W.R. (2015). Borrowing information over time in binomial/logit normal models for small area estimation. *Statistics in Transition New Series*, 16(4), 563-584. <https://sciendo.com/article/10.21307/stattrans-2015-033>.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515-533. <https://doi.org/10.1214/06-ba117a>.
- Gelman, A., et Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457-472. <https://doi.org/10.1214/ss/1177011136>.
- Gelman, A., Meng, X.-L. et Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733-760.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. et Rubin, D.B. (2013). *Bayesian Data Analysis*. CRC press.

- Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Kreuter, F. (2013). Facing the nonresponse challenge. *Annals of the American Academy of Political and Social Science*, 645(1), 23-35. <https://doi.org/10.1177/0002716212456815>.
- Mushkudiani, N., et Schouten, B. (2019). Time-Dependent Survey Design Parameters: Choosing the Length of Historic Survey Data in a Bayesian Analysis, Application to the Dutch Health Survey. Document d'atelier pour l'avances in adaptive and responsive survey design.
- Pfeffermann, D., et Burck, L. (1990). [Estimation robuste pour petits domaines par la combinaison de données chronologiques et transversales](#). *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>.
- Polson, N.G., et Scott, J.G. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian statistics*, 9, 501-538. <https://doi.org/10.1093/acprof:oso/9780199694587.003.0017>.
- Polson, N.G., Scott, J.G. et Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339-1349. <https://doi.org/10.1080/01621459.2013.829001>.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienne, Autriche. URL: <https://www.R-project.org/>.
- Rao, J.N.K., et Molina, I. (2015). *Small Area Estimation*. Wiley StatsRef: Statistics Reference Online, 1-8.
- Rao, J.N.K., et Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics*, 22(4), 511-528. <https://doi.org/10.2307/3315407>.
- Rosenbaum, P.R., et Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55. <https://doi.org/10.1093/biomet/70.1.41>.
- Rue, H., et Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC press.
- Schouten, B., Peytchev, A. et Wagner, J. (2017). *Adaptive Survey Design*. CRC press.

- Schouten, B., Mushkudiani, N., Shlomo, N., Durrant, G., Lundquist, P. et Wagner, J. (2018). A Bayesian analysis of design parameters in survey data collection. *Journal of Survey Statistics and Methodology*, 6(4), 431-464. <https://doi.org/10.1093/jssam/smy012>.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.R. et van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 64, 583-616. <https://doi.org/10.1111/1467-9868.00353>.
- Wagner, J., et Hubbard, F. (2014). Producing unbiased estimates of propensity models during data collection. *Journal of Survey Statistics and Methodology*, 2(3), 323-342. <https://doi.org/10.1093/jssam/smu009>.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571-3594. <http://jmlr.org/papers/v11/watanabe10a.html>.
- Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, 14, 867-897.
- West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74(3), 646-648. <https://doi.org/10.1093/biomet/74.3.646>.
- West, B.T., Wagner, J., Coffey, S. et Elliott, M.R. (2021). Deriving priors for Bayesian prediction of daily response propensity in responsive survey design: Historical data analysis vs. literature review. *Journal of Survey Statistics and Methodology*. <https://doi.org/10.1093/jssam/smab036>.
- Wu, S., Schouten, B., Meijers, R. et Moerbeek, M. (2022). Data collection expert prior elicitation in survey design: Two case studies. *Journal of Official Statistics*, 38(2), 637-662. <https://sciendo.com/es/article/10.2478/jos-2022-0028>.
- You, Y., Rao, J.N.K. et Gambino, J. (2003). [Estimation du taux de chômage fondée sur un modèle pour l'Enquête sur la population active du Canada : Une approche bayésienne hiérarchique](#). *Techniques d'enquête*, 29, 1, 27-36. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2003001/article/6602-fra.pdf>.