

# Diminution des coûts de santé : le rôle des groupes communautaires

## Annexe méthodologique

RAPHAËL LANGEVIN, chercheur

### Table des matières

p. 03	1. Caractéristiques et avantages de l'estimateur en seconde différence
p. 10	2. Somme des effets annuels et ratios bénéfiques-coûts
p. 12	3. Résultats supplémentaires
p. 13	4. Simulations
p. 16	Références

# 1. Caractéristiques et avantages de l'estimateur en seconde différence

## 1.1 BIAIS D'ESTIMATION

Le modèle statistique utilisé dans la note est similaire à celui employé par [Langevin \(2021\)](#) et peut être représenté par l'équation suivante :

$$\ln(Y_{it}) = \ln(X_{it})^\top \beta + \sum_{k=0}^5 \gamma_k^j \ln(W_{it-k}^j) + \alpha_i \times t + \delta_i + \delta_t + \epsilon_{it}, \quad (1)$$

où  $Y_{it}$  correspond aux dépenses réelles de santé par habitant pour la région  $i$  à l'année  $t$  avec  $i = 1, \dots, N$  et  $t = 1, \dots, T$ ,  $X_{it}$  est un vecteur-colonne contenant les 10 covariables mentionnées dans la section 2.1 de la note socioéconomique,  $\beta$  est le vecteur-colonne des élasticités associées à chacun des éléments dans  $X_{it}$ ,  $W_{it}^j$  est le montant d'argent réel par habitant versé aux organismes communautaires de la catégorie  $j$  au sein de la région  $i$  à l'année  $t$ ,  $\gamma_k^j$  est l'élasticité des dépenses communautaires pour les organismes dans la  $j^{\text{ième}}$  catégorie après  $k$  années, et où  $\alpha_i$ ,  $\delta_i$  et  $\delta_t$  désignent respectivement une tendance linéaire pour la région  $i$ , un effet fixe régional pour la région  $i$  et un effet fixe temporel pour l'année  $t$ . Notez que ces trois derniers termes ainsi que le vecteur  $\beta$  peuvent changer selon la catégorie considérée, mais nous omettons volontairement l'exposant  $j$  afin d'alléger la notation et aussi parce que ces coefficients ne sont pas des coefficients d'intérêt. Finalement, il est supposé, pour l'instant, que le terme d'erreur  $\epsilon_{it}$  est indépendamment et identiquement distribué (i.i.d.) avec moyenne nulle et variance constante et finie ( $\sigma_\epsilon^2 < \infty$ ).

La condition permettant l'absence de biais dans les modèles de ce genre est que le terme d'erreur  $\epsilon_{it}$  ne doit pas être corrélé avec un quelconque régresseur présent dans l'équation de régression, ce qui implique généralement que le modèle doit être correctement spécifié. Par contre, dans le modèle décrit par l'équation (1), la convergence lorsque  $T \rightarrow \infty$  est impossible en raison de la présence des tendances linéaires  $\alpha_i \times t$ <sup>1</sup>. En appliquant l'opérateur de première différence  $\Delta$  à l'équation (1) (où  $\Delta z_{it} = z_{it} - z_{it-1}$  pour n'importe quelle variable  $z_{it}$ ), nous obtenons

$$\Delta y_{it} = \Delta x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta w_{it-k}^j + \alpha_i + \Delta \delta_t + \Delta \epsilon_{it}, \quad (2)$$

où  $z_{it} = \ln(Z_{it})$  pour n'importe quelle variable  $Z_{it}$  qui n'est pas déjà exprimée en logarithme et où le coefficient de tendance linéaire pour la région  $i$ ,  $\alpha_i$ , correspond maintenant à un effet fixe régional dans le modèle en première différence.

Comme les tendances linéaires n'apparaissent plus dans ce modèle, nous pouvons estimer de manière convergente et non biaisée les coefficients  $\gamma_k^j$  sous l'hypothèse que l'équation (1) est une

---

1. Cela provient du fait que la matrice  $\mathbb{E} \left[ \frac{\mathbf{x}^\top \mathbf{x}}{NT} \right]$  ne peut pas converger vers une matrice finie avec  $N, T \rightarrow \infty$  lorsque des tendances temporelles sont incluses dans le modèle, où  $\mathbf{X}$  est la matrice contenant l'ensemble des régresseurs, incluant les effets fixes et les tendances linéaires. Cette matrice doit être finie asymptotiquement afin d'identifier précisément les coefficients du modèle.

représentation fidèle de la réalité. D'abord, notons que

$$\overline{\overline{\Delta y_{it}}} = \{\overline{\overline{\Delta x_{it}}}\}^\top \beta + \sum_{k=0}^5 \gamma_k^j \overline{\overline{\Delta w_{it-k}^j}} + \overline{\overline{\Delta \epsilon_{it}}}, \quad (3)$$

où  $\overline{\overline{\Delta z_{it}}} := \Delta z_{it} - \overline{\overline{\Delta z_i}} - \overline{\overline{\Delta z_t}} + \overline{\overline{\Delta z}}$  pour n'importe quelle variable  $\Delta z_{it}$  qui est exprimée en première différence, avec  $\overline{\overline{\Delta z_i}} = (T-1)^{-1} \sum_{t=2}^T \Delta z_{it}$ ,  $\overline{\overline{\Delta z_t}} = N^{-1} \sum_{i=1}^N \Delta z_{it}$ , et où  $\overline{\overline{\Delta z}} = (N(T-1))^{-1} \sum_{i=1}^N \sum_{t=2}^T \Delta z_{it}$ . En supposant que les éléments dans  $\overline{\overline{\Delta x_{it}}}$  ne sont pas corrélés avec  $\overline{\overline{\Delta w_{it-k}^j}}$  pour chaque valeur de  $k \in \{0, \dots, 5\}$  et chaque valeur de  $j$ , la condition typique afin d'obtenir des estimés  $\hat{\gamma}_k^j$  non biaisés peut être écrite ainsi <sup>2</sup> :

$$\mathbb{E}[\overline{\overline{\Delta \epsilon_{it}}} | \overline{\overline{\Delta w_{it}^j}}, \dots, \overline{\overline{\Delta w_{it-5}^j}}] = 0, \quad (4)$$

pour chaque valeur de  $j$  séparément. Cette dernière condition sera forcément violée si

$$\mathbb{E}[\overline{\overline{\Delta \epsilon_{it}}} \overline{\overline{\Delta w_{it-k}^j}}] \neq 0, \quad (5)$$

pour une quelconque valeur de  $k \in \{0, \dots, 5\}$ . Il est important de noter que si l'équation (5) n'est pas vérifiée et  $\mathbb{E}[\overline{\overline{\Delta \epsilon_{it}}} \overline{\overline{\Delta w_{it-k}^j}}] = 0 \forall k \in \{0, \dots, 5\}$ , cela n'implique pas forcément que (4) soit vraie. Toutefois, les biais générés dans le cas où l'équation (5) n'est pas vérifiée ont toutes les chances d'être relativement faibles. C'est pourquoi nous considérons que les coefficients  $\hat{\gamma}_k^j$  présenteront une quasi-absence de biais si (5) n'est pas vérifiée pour une quelconque valeur de  $k \in \{0, \dots, 5\}$ .

Considérant cela, il est alors possible de réécrire la condition entraînant la quasi-absence de biais dans les coefficients  $\hat{\gamma}_k^j$  de la façon suivante pour une valeur de retard  $k$  donnée :

$$\mathbb{E}[(\Delta \epsilon_{it} - \overline{\overline{\Delta \epsilon_i}} - \overline{\overline{\Delta \epsilon_t}} + \overline{\overline{\Delta \epsilon}})(\Delta w_{it-k}^j - \overline{\overline{\Delta w_{i,k}^j}} - \overline{\overline{\Delta w_{t-k}^j}} + \overline{\overline{\Delta w_k^j}})] = 0,$$

où  $\overline{\overline{\Delta w_{i,k}^j}} = (T-1)^{-1} \sum_{t=2}^T \Delta w_{it-k}^j$ ,  $\overline{\overline{\Delta w_{t-k}^j}} = N^{-1} \sum_{i=1}^N \Delta w_{it-k}^j$ , et où le dernier terme  $\overline{\overline{\Delta w_k^j}} = (N(T-1))^{-1} \sum_{i=1}^N \sum_{t=2}^T \Delta w_{it-k}^j$ . En isolant les termes qui sont multipliés par  $\Delta \epsilon_{it}$  dans l'équation plus haut, il est possible de montrer que

$$\begin{aligned} \mathbb{E}[\overline{\overline{\Delta \epsilon_{it}}} \overline{\overline{\Delta w_{it-k}^j}}] &= \frac{(N-1)(T-2)}{N(T-1)} \mathbb{E}[\Delta w_{it-k}^j \Delta \epsilon_{it}] - \frac{T-2}{N(T-1)} \sum_{m \neq i} \mathbb{E}[\Delta w_{mt-k}^j \Delta \epsilon_{it}] \\ &\quad - \frac{N-1}{N(T-1)} \sum_{l \neq t} \mathbb{E}[\Delta w_{il-k}^j \Delta \epsilon_{it}] + \frac{1}{N(T-1)} \sum_{m \neq i} \sum_{l \neq t} \mathbb{E}[\Delta w_{ml-k}^j \Delta \epsilon_{it}]. \end{aligned} \quad (6)$$

Chacun des quatre termes du côté droit de l'équation (6) va être égal à zéro peu importe la valeur de  $N$  et  $T$  si toutes les espérances contenues dans l'équation (6) sont égales à zéro. Si ces espérances

---

2. Une telle hypothèse peut facilement être relâchée en appliquant le théorème de Frisch-Waugh-Lovell (FWL) sur l'équation (3) afin d'éliminer tous les termes qui sont une fonction de  $\Delta x_{it}$  et de  $\Delta w_{it-l}^j \forall l \neq k$ . De manière générale, tous les résultats présentés dans cette annexe sont maintenus après avoir substitué chaque  $\Delta w_{it-k}^j$  par  $\widehat{\Delta \nu}_{it-k}^j$ , où  $\widehat{\Delta \nu}_{it-k}^j$  est le terme d'erreur obtenu par l'application du théorème FWL sur le vecteur contenant la variable  $\Delta w_{it-k}^j$  pour chaque valeur de  $k \in \{0, \dots, 5\}$ .

ne sont pas toutes égales à zéro, cela entraînera un biais en échantillon fini dans les estimés pour des valeurs de  $N < \infty$  et  $T < \infty$ . L'équation (6) montre aussi qu'un biais asymptotique sera présent si  $\mathbb{E}[\Delta w_{it-k}^j \Delta \epsilon_{it}] \neq 0$ , ce qui est le cas lorsqu'il y a endogénéité contemporaine entre le financement communautaire  $\Delta w_{it-k}^j$  et le terme d'erreur  $\Delta \epsilon_{it}$ .

Si nous nous concentrons seulement sur le biais asymptotique (soit le premier terme dans le côté droit de l'équation (6)), il est possible de substituer  $\Delta \epsilon_{it}$  par  $\overline{\Delta \epsilon_i}$ ,  $\overline{\Delta \epsilon_t}$  ou encore  $\overline{\overline{\Delta \epsilon}}$  afin d'obtenir les trois conditions asymptotiques suivantes (en plus de  $\mathbb{E}[\Delta w_{it-k}^j \Delta \epsilon_{it}] = 0$ ) :

$$\begin{aligned}\mathbb{E}[\Delta w_{it-k}^j \overline{\Delta \epsilon_i}] &= (T-1)^{-1} \sum_{l=2}^T \mathbb{E}[\Delta w_{it-k}^j \Delta \epsilon_{il}] = 0, \\ \mathbb{E}[\Delta w_{it-k}^j \overline{\Delta \epsilon_t}] &= N^{-1} \sum_{m=1}^N \mathbb{E}[\Delta w_{it-k}^j \Delta \epsilon_{mt}] = 0, \\ \mathbb{E}[\Delta w_{it-k}^j \overline{\overline{\Delta \epsilon}}] &= (N(T-1))^{-1} \sum_{m=1}^N \sum_{l=2}^T \mathbb{E}[\Delta w_{it-k}^j \Delta \epsilon_{ml}] = 0.\end{aligned}$$

Ces trois conditions doivent être remplies afin de garantir la quasi-absence de biais asymptotique dans les estimés  $\hat{\gamma}_k^j$ . Autrement dit, ces conditions empêchent la présence d'effets de rétroaction entre chacune des régions (quand  $\Delta \epsilon_{ml}$  est corrélé avec  $\Delta w_{it-k}^j$  pour  $l < t - k$  peu importe la valeur de  $m$ ), tout comme la présence d'anticipations (ou d'effets de débordement) entre chacune des régions (quand  $\Delta \epsilon_{ml}$  est corrélé avec  $\Delta w_{it-k}^j$  pour  $m \neq i$  et  $l > t - k$ ). Notez que les effets de  $\Delta w_{ml-k}^j$  sur  $\Delta \epsilon_{it}$  quand  $m = i$  et  $t - 5 \leq l - k \leq t$  sont pris en compte dans l'équation (2) par les six variables  $\Delta w_{it-k}^j$  avec  $k = \{0, \dots, 5\}$ . Toutefois, la présence potentielle d'anticipations limite la plausibilité d'une absence totale de corrélation entre  $\Delta w_{ml-k}^j$  et  $\Delta \epsilon_{it}$  lorsque  $t - 5 \leq l - k < t$ .

Le but de cet exercice est de montrer que les conditions garantissant la quasi-absence de biais asymptotique dans les estimés sont complexes et risquent de ne pas souvent être satisfaites. Autrement dit, pour obtenir des estimés non biaisés, il est nécessaire que, en plus de l'absence d'endogénéité contemporaine, aucun effet de rétroaction des dépenses de santé vers le financement communautaire, ni d'anticipations, ni d'effets retardés du financement communautaire sur les dépenses de santé (au-delà de ceux inclus dans le modèle) ne doit être présent dans les données, et ce, autant au sein d'une même région qu'entre les régions. Cela est considéré comme très exigeant en pratique.

Cet ensemble de conditions est d'ailleurs souvent négligé dans les analyses de panel à effets fixes doubles (*two-way fixed effects*, voir notamment la section 3.2 de Millimet et Bellemare (2023) pour une discussion sur le sujet). Un tel ensemble de conditions est nécessaire même dans une analyse standard en niveau sans tendance linéaire par région/unité car les effets fixes doubles requièrent une double soustraction de moyenne, soit la soustraction de la moyenne temporelle par région ( $\overline{\Delta w_{i,k}^j}$ ), suivie d'une soustraction de la moyenne régionale par année ( $\overline{\Delta w_{t-k}^j}$ ). En résumé, cet exercice montre qu'il est important d'argumenter l'exogénéité complète des régresseurs dans les deux dimensions (temps et unités) lorsqu'un estimateur classique en niveau est employé pour une analyse de panel à effets fixes doubles.

Si nous appliquons l'opérateur de seconde différence  $\Delta^2$  à l'équation (1), où  $\Delta^2 z_{it} = \Delta z_{it} - \Delta z_{it-1} = z_{it} - 2z_{it-1} + z_{it-2}$  pour n'importe quelle variable  $z_{it}$ , nous obtenons l'expression suivante :

$$\Delta^2 y_{it} = \Delta^2 x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta^2 w_{it-k}^j + \Delta^2 \delta_t + \Delta^2 \epsilon_{it}, \quad (7)$$

où la condition afin d'obtenir des estimés quasiment non biaisés de  $\gamma_k^j$  peut être écrite ainsi

$$\mathbb{E}[(\Delta^2 w_{it-k}^j - \overline{\Delta^2 w_t^j})(\Delta^2 \epsilon_{it} - \overline{\Delta^2 \epsilon_t})] = 0.$$

Il est possible de réécrire cette dernière condition de la manière suivante :

$$\mathbb{E}[(\dot{w}_{it-k}^j - 2\dot{w}_{it-k-1}^j + \dot{w}_{it-k-2}^j)(\dot{\epsilon}_{it} - 2\dot{\epsilon}_{it-1} + \dot{\epsilon}_{it-2})] = 0, \quad (8)$$

où  $\dot{z}_{it} = z_{it} - N^{-1} \sum_{i=1}^N z_{it} \equiv z_{it} - \bar{z}_t$  pour n'importe quelle variable  $z_{it}$  exprimée en niveau. Cette condition est plus facile à satisfaire que la précédente, car elle n'est pas fondée sur l'exogénéité complètement stricte des régresseurs au sein des régions, mais plutôt sur une absence de corrélation entre  $\epsilon_{it}$  et  $w_{ms}^j$  pour  $s = \{t-k-2, \dots, t-k+2\}$  et n'importe quelle paire  $(i, m) = \{1, \dots, N\}^2$ . Effectivement, si nous considérons que la véracité de l'équation (8) entraîne une (quasi-)absence de biais, alors l'estimation non biaisée des coefficients  $\hat{\gamma}_k^j$  exige que les termes d'erreurs et les dépenses communautaires régionales soient indépendantes entre elles à l'intérieur d'une fenêtre de plus ou moins 2 périodes, ce qui peut être considéré comme irréaliste dans certains cas. Cela est toutefois bien moins exigeant que l'exogénéité stricte dans les deux dimensions qui est nécessaire pour l'estimateur en première différence (ou pour l'estimateur en niveau en l'absence de tendance linéaire). Les paragraphes suivants présentent un exemple de la taille relative des différents biais étant générés par un estimateur en seconde différence.

Supposons que le terme d'erreur  $\epsilon_{it}$  suit un processus auto-régressif d'ordre 1, soit que  $\epsilon_{it} = \rho \epsilon_{it-1} + v_{it}$  où  $|\rho| < 1$  et où  $v_{it}$  est le terme d'erreur qui est potentiellement endogène avec moyenne nulle et variance finie<sup>3</sup>. Supposons aussi qu'il existe un effet d'anticipation à un an, une endogénéité contemporaine et une infinité d'effets de rétroaction entre le financement communautaire et les dépenses en santé au Québec lorsque ces deux variables sont exprimées en déviations par rapport à la moyenne des régions. Plus précisément, supposons que  $\mathbb{E}[\dot{w}_{it-l}^j \dot{v}_{it}] \neq 0$  pour  $l = \{-\infty, \dots, 1\}$  et pour n'importe quelle valeur de  $j$ , avec  $\mathbb{E}[\dot{w}_{it-l}^j \dot{v}_{it}] = 0$  autrement. Dans un tel cas, nous pouvons écrire que

$$\begin{aligned} \mathbb{E}[\dot{w}_{it-l}^j \dot{\epsilon}_{it}] &= \mathbb{E}[\dot{w}_{it-l}^j (\rho \dot{\epsilon}_{it-1} + \dot{v}_{it})], \\ &= \rho \mathbb{E}[\dot{w}_{it-l}^j (\rho \dot{\epsilon}_{it-2} + \dot{v}_{it-1})] + \mathbb{E}[\dot{w}_{it-l}^j \dot{v}_{it}], \\ &= \sum_{m=0}^{\infty} \rho^m \mathbb{E}[\dot{w}_{it-l}^j \dot{v}_{it-m}], \end{aligned}$$

où nous avons que

$$\mathbb{E}[\dot{w}_{it-l}^j \dot{v}_{it-m}] = 0 \text{ si } l - m \geq 2,$$

---

3. Appliquer l'endogénéité sur  $\epsilon_{it}$  ou  $v_{it}$  ne change pas les conclusions de l'exercice, pour autant que le terme  $\epsilon_{it}$  soit autocorrélé.

par hypothèse. Par conséquent, nous pouvons écrire que

$$\mathbb{E}[\dot{w}_{it-l}^j \dot{\epsilon}_{it}] = \begin{cases} \sum_{m=0}^{\infty} \rho^{m+l-1} \mathbb{E}[\dot{w}_{it-1}^j \dot{v}_{it-m}] & \text{if } l \geq 2, \\ \sum_{m=0}^{\infty} \rho^m \mathbb{E}[\dot{w}_{it-l}^j \dot{v}_{it-m}] & \text{if } l < 2, \end{cases} \quad (9)$$

pour toute valeur de  $l \in \mathbb{Z}$ , où  $\mathbb{Z}$  est l'ensemble contenant tous les nombres entiers.

En regroupant les termes partageant une même distance temporelle entre eux, il est possible de réécrire l'équation (8) de la manière suivante :

$$\mathbb{E}[\dot{w}_{it-k+2}^j \dot{\epsilon}_{it} - 4\dot{w}_{it-k+1}^j \dot{\epsilon}_{it} + 6\dot{w}_{it-k}^j \dot{\epsilon}_{it} - 4\dot{w}_{it-k-1}^j \dot{\epsilon}_{it} + \dot{w}_{it-k-2}^j \dot{\epsilon}_{it}] = 0.$$

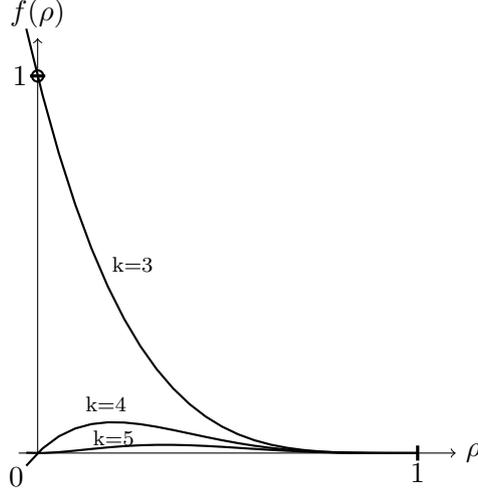
En utilisant l'équation (9) pour substituer les termes correspondants, on peut finalement écrire l'équivalent de l'équation (8) pour chaque valeur de  $k \geq 0$  :

$$\begin{aligned} k = 0 : \quad & \mathbb{E}[\dot{w}_{it+2}^j \dot{\epsilon}_{it}] - 4\mathbb{E}[\dot{w}_{it+1}^j \dot{\epsilon}_{it}] + 6\mathbb{E}[\dot{w}_{it}^j \dot{\epsilon}_{it}] + (\rho - 4) \sum_{m=0}^{\infty} \rho^m \mathbb{E}[\dot{w}_{it-1}^j \dot{v}_{it-m}] = 0, \\ k = 1 : \quad & \mathbb{E}[\dot{w}_{it+1}^j \dot{\epsilon}_{it}] - 4\mathbb{E}[\dot{w}_{it}^j \dot{\epsilon}_{it}] + (\rho^2 - 4\rho + 6) \sum_{m=0}^{\infty} \rho^m \mathbb{E}[\dot{w}_{it-1}^j \dot{v}_{it-m}] = 0, \\ k = 2 : \quad & \mathbb{E}[\dot{w}_{it}^j \dot{\epsilon}_{it}] + (\rho^3 - 4\rho^2 + 6\rho - 4) \sum_{m=0}^{\infty} \rho^m \mathbb{E}[\dot{w}_{it-1}^j \dot{v}_{it-m}] = 0, \\ k \geq 3 : \quad & (\rho^{k+1} - 4\rho^k + 6\rho^{k-1} - 4\rho^{k-2} + \rho^{k-3}) \sum_{m=0}^{\infty} \rho^m \mathbb{E}[\dot{w}_{it-1}^j \dot{v}_{it-m}] = 0. \end{aligned}$$

On remarque tout de suite que, pour les valeurs de  $k < 3$ , l'équation (8) sera forcément violée et les coefficients  $\hat{\gamma}_k^j$  correspondant seront biaisés. Lorsque  $k = 3$ , l'absence d'autocorrélation dans le terme d'erreur crée un biais important en raison de l'effet d'anticipation à un an. Toutefois, plus l'autocorrélation est forte, plus ce biais diminue, ce qui est représenté au graphique A.1 de la page suivante. Lorsque  $k > 3$ , les valeurs de l'expression  $\rho^{k+1} - 4\rho^k + 6\rho^{k-1} - 4\rho^{k-2} + \rho^{k-3}$  sont toujours très près de zéro pour n'importe quelle valeur de  $\rho > 0$ . Le biais devient cependant très élevé si l'autocorrélation du terme d'erreur est fortement négative, ce qui est toutefois peu probable en réalité. Considérant cela, il est fort probable que les résultats pour  $k = \{0, 1, 2, 3\}$  soient biaisés, ce qui n'est toutefois pas le cas pour les résultats lorsque  $k = \{4, 5\}$ . La même logique s'applique aisément lorsque  $\epsilon_{it}$  suit un processus auto-régressif d'ordre  $p$ , mais une faible valeur de biais sera garantie seulement pour les coefficients associés à des valeurs de  $k > p + 2$ . Notons aussi qu'en l'absence d'effet d'anticipation, une faible valeur de biais sera garantie pour les coefficients associés à des valeurs de  $k > p + 1$ , impliquant un faible biais lorsque  $k = \{3, 4, 5\}$  dans l'exemple présenté, et ce peu importe la valeur de  $\rho$ .

## 1.2 ROBUSTESSE FACE AUX BRIS STRUCTURELS

En plus d'être moins exigeant du point de vue de l'exogénéité des régresseurs, l'estimateur en seconde différence est plus robuste aux changements de tendances et aux variations dans les effets



Graphique A.1 : Valeur du terme  $f(\rho) = \rho^{k+1} - 4\rho^k + 6\rho^{k-1} - 4\rho^{k-2} + \rho^{k-3}$  lorsque représenté comme une fonction de  $\rho$  pour différentes valeurs de  $k \geq 3$ .

fixes unitaires/régionaux que les estimateurs en niveau ou en première différence. Pour voir cela, supposons que le vrai modèle est décrit par l'équation suivante :

$$y_{it} = \begin{cases} x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j w_{it-k}^j + \alpha_{i1} \times t + \delta_{i1} + \delta_t + \epsilon_{it} & \text{si } t < T_0, \\ x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j w_{it-k}^j + \alpha_{i2} \times t + \delta_{i2} + \delta_t + \epsilon_{it} & \text{si } t \geq T_0, \end{cases} \quad (10)$$

où  $\alpha_{i1} \neq \alpha_{i2}$  et  $\delta_{i1} \neq \delta_{i2}$ . Ce modèle suppose la présence d'un seul bris structurel qui survient à la même période  $T_0$  pour toutes les régions, mais la taille du bris peut varier entre les régions. La démonstration suivante se généralise relativement bien à la présence de plusieurs bris structurels à des périodes identiques ou différentes pour chaque région. En appliquant à l'expression ci-haut un opérateur de déviations par rapport à la moyenne temporelle, nous obtenons le résultat suivant :

$$\ddot{y}_{it} = \begin{cases} \ddot{x}_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \ddot{w}_{it-k}^j + \alpha_{i1} t - \frac{(\alpha_{i1} - \alpha_{i2})(T_0 - 1)T_0}{2T} - \frac{\alpha_{i2}(T+1)}{2} + \frac{T - (T_0 - 1)}{T}(\delta_{i1} - \delta_{i2}) + \ddot{\delta}_t + \ddot{\epsilon}_{it} & \text{si } t < T_0, \\ \ddot{x}_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \ddot{w}_{it-k}^j + \alpha_{i2} t - \frac{(\alpha_{i1} - \alpha_{i2})(T_0 - 1)T_0}{2T} - \frac{\alpha_{i2}(T+1)}{2} + \frac{T_0 - 1}{T}(\delta_{i2} - \delta_{i1}) + \ddot{\delta}_t + \ddot{\epsilon}_{it} & \text{si } t \geq T_0, \end{cases}$$

où  $\ddot{z}_{it} = z_{it} - T^{-1} \sum_{t=1}^T z_{it} = z_{it} - \bar{z}_i$  pour n'importe quelle variable  $z_{it}$ . On constate que la déviation par rapport à la moyenne n'annule pas les effets fixes régionaux et que cela va biaiser les estimés  $\gamma_k^j$  et  $\beta$  si  $\alpha_{i1}$ ,  $\alpha_{i2}$ ,  $\delta_{i1}$  et  $\delta_{i2}$  ne sont pas indépendants des régresseurs  $x_{it}$  et  $w_{it-k}^j$ . Spécifiquement, la taille du biais pour  $\gamma_j^k$  dépendra de la corrélation entre  $\delta_{i1} - \delta_{i2}$  et  $\ddot{w}_{it-k}^j$ , de la corrélation entre  $\alpha_{iz} t - \frac{(\alpha_{i1} - \alpha_{i2})(T_0 - 1)T_0}{2T} - \frac{\alpha_{i2}(T+1)}{2}$  et  $\ddot{w}_{it-k}^j$ , où  $z = \{1, 2\}$ , et aussi du ratio  $T_0/T$  pour chaque valeur de  $j$  et de  $k$ .

Notez que le biais généré par le bris dans les effets fixes  $\delta_{i1} - \delta_{i2}$  dépend largement du ratio  $T_0/T$ . Si ce ratio est proche de zéro (proche de 1), alors le bris dans les effets fixes affectera surtout les observations où  $t < T_0$  ( $t \geq T_0$ ), avec une incidence totale plus importante si  $T_0/T$  est proche de 1/2.

Il est possible d'appliquer l'opérateur de première différence à l'équation (10), ce qui nous donne l'expression suivante :

$$\Delta y_{it} = \begin{cases} \Delta x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta w_{it-k}^j + \alpha_{i1} + \Delta \delta_t + \Delta \epsilon_{it} & \text{si } t < T_0, \\ \Delta x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta w_{it-k}^j + (\alpha_{i2} - \alpha_{i1})T_0 + \alpha_{i1} + \delta_{i2} - \delta_{i1} + \Delta \delta_t + \Delta \epsilon_{it} & \text{si } t = T_0, \\ \Delta x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta w_{it-k}^j + \alpha_{i2} + \Delta \delta_t + \Delta \epsilon_{it} & \text{si } t > T_0. \end{cases}$$

Comme pour l'estimateur en niveau, il est possible de montrer que l'estimateur à double effets fixes sur ce modèle en première différence entraîne un biais qui ne convergera pas forcément vers zéro lorsque  $T \rightarrow \infty$ . En effet, la moyenne temporelle par région obtenue à partir de la précédente équation est définie ainsi (en prenant soin de ne pas inclure la première période)

$$\overline{\Delta y_i} = \overline{\Delta x_i}^\top \beta + \sum_{k=0}^5 \overline{\gamma_k^j \Delta w_{i,k}^j} + \frac{\alpha_{i2}T - \alpha_{i1} + \delta_{i2} - \delta_{i1}}{T-1} + \overline{\Delta \delta} + \overline{\Delta \epsilon_i},$$

où les barres au-dessus des variables représentent des moyennes et ont les mêmes significations que dans la section précédente. La déviation par rapport à la moyenne du modèle en première différence peut donc s'écrire ainsi

$$\ddot{\Delta y}_{it} = \begin{cases} \ddot{\Delta x}_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \ddot{\Delta w}_{it-k}^j + \frac{T}{T-1}(\alpha_{i1} - \alpha_{i2}) - \frac{1}{T-1}(\delta_{i2} - \delta_{i1}) + \ddot{\Delta \delta}_t + \ddot{\Delta \epsilon}_{it} & \text{si } t < T_0, \\ \ddot{\Delta x}_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \ddot{\Delta w}_{it-k}^j + \left(\frac{T}{T-1} - T_0\right)(\alpha_{i1} - \alpha_{i2}) + \frac{T-2}{T-1}(\delta_{i2} - \delta_{i1}) + \ddot{\Delta \delta}_t + \ddot{\Delta \epsilon}_{it} & \text{si } t = T_0, \\ \ddot{\Delta x}_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \ddot{\Delta w}_{it-k}^j + \frac{1}{T-1}(\alpha_{i1} - \alpha_{i2}) - \frac{1}{T-1}(\delta_{i2} - \delta_{i1}) + \ddot{\Delta \delta}_t + \ddot{\Delta \epsilon}_{it} & \text{si } t > T_0. \end{cases}$$

On remarque ici aussi que la déviation par rapport à la moyenne n'annule pas complètement les effets fixes régionaux. Bien que les termes  $\frac{1}{T-1}(\alpha_{i1} - \alpha_{i2})$  et  $\frac{1}{T-1}(\delta_{i2} - \delta_{i1})$  convergent vers zéro lorsque  $T \rightarrow \infty$ , le terme  $\frac{T}{T-1}(\alpha_{i1} - \alpha_{i2})$  qui est présent dans toutes les périodes pré-bris ne sera pas éliminé asymptotiquement. Par conséquent, un biais asymptotique sera présent dans les coefficients d'intérêt  $\gamma_k^j$  et ce biais dépendra de la corrélation entre  $(\alpha_{i1} - \alpha_{i2})$  et  $\ddot{\Delta w}_{it-k}^j$  et de la période où se trouve le bris,  $T_0$ . Le biais total convergera vers zéro si et seulement si le ratio  $T_0/T \rightarrow 0$  lorsque  $T \rightarrow \infty$  (car le poids des observations pré-bris deviendra marginal).

Si nous nous rapportons au contexte de la note socioéconomique, il est possible de constater qu'un bris structurel est apparu à la fois dans l'évolution des dépenses de santé et dans l'évolution du financement communautaire lors du début de la pandémie de COVID-19, soit à l'année 2019-2020 pour le financement communautaire et à l'année 2020-2021 pour les dépenses de santé (voir les graphiques 3 et 4 de la note). Par conséquent, nous avons que  $\Delta \alpha_i = \alpha_{i1} - \alpha_{i2} < 0$  pour un ratio  $T_0/T$  autour de 0,8. Comme le financement communautaire a aussi subi un tel bris structurel, nous avons que  $\ddot{\Delta w}_{it-k}^j$  est négatif pour la plupart des observations pré-COVID considérant que la moyenne temporelle est rehaussée significativement par les observations après 2019-2020. Nous pouvons donc réalistement affirmer que  $\text{Cov}[\ddot{\Delta w}_{it-k}^j, \Delta \alpha_i] > 0$  pour la plupart des observations pré-COVID avec  $k \in \{0, \dots, 5\}$  et pour n'importe quelle catégorie  $j$ , ce qui explique pourquoi l'estimateur en première différence tend à biaiser positivement les estimés de  $\gamma_k^j$ . Cela entraîne alors une sous-estimation de l'effet du financement communautaire sur les dépenses de santé. Cette intuition est confirmée par des

simulations où des bris structurels corrélés sont introduits dans la variable dépendante et dans une des variables indépendantes (voir la section 4 de l'annexe).

Finalement, en appliquant l'opérateur de seconde différence plutôt que l'opérateur de première différence à l'équation (10), nous obtenons l'expression suivante :

$$\Delta^2 y_{it} = \begin{cases} \Delta^2 x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta^2 w_{it-k}^j + \Delta^2 \delta_t + \Delta^2 \epsilon_{it} & \text{si } t \neq \{T_0, T_0 + 1\}, \\ \Delta^2 x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta^2 w_{it-k}^j + (\alpha_{i2} - \alpha_{i1})T_0 + \delta_{i2} - \delta_{i1} + \Delta^2 \delta_t + \Delta^2 \epsilon_{it} & \text{si } t = T_0, \\ \Delta^2 x_{it}^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta^2 w_{it-k}^j + (\alpha_{i1} - \alpha_{i2})(T_0 - 1) + \delta_{i1} - \delta_{i2} + \Delta^2 \delta_t + \Delta^2 \epsilon_{it} & \text{si } t = T_0 + 1. \end{cases}$$

On remarque alors que le biais pour  $\gamma_k^j$  n'affectera que deux observations seulement par région (soit lorsque  $t = \{T_0, T_0 + 1\}$ ) et que la taille du biais pour ces deux observations dépendra de la corrélation entre  $(\alpha_{i1} - \alpha_{i2})T_0$  et  $\Delta^2 w_{it-k}^j$ , de la corrélation entre  $(\delta_{i2} - \delta_{i1})$  et  $\Delta^2 w_{it-k}^j$  et de l'emplacement du bris  $T_0$ . Contrairement aux autres estimateurs, l'estimateur en seconde différence ne génère aucun biais additionnel en échantillon fini pour les observations qui ne sont pas directement affectées par le bris (soit quand  $t \neq \{T_0, T_0 + 1\}$ ). Il s'agit du principal avantage de l'estimateur en seconde différence. Comme notre échantillon possède seulement 15 périodes, il est fort probable que cela ait un impact majeur sur nos estimés, considérant le fort risque de bris structurels dans les séries. Cette intuition est confirmée par simulations, simulations qui montrent que l'estimateur en seconde différence permet une réduction du biais variant entre 0 et 50% par rapport aux biais obtenus avec les estimateurs en niveau ou en première différence (selon la variance et le niveau de corrélation dans les bris structurels).

### 1.3 ROBUSTESSE FACE AUX RACINES UNITAIRES

Il importe aussi de mentionner que l'estimateur en seconde différence est plus robuste à la présence de racines unitaires d'ordre  $\geq 2$  dans le terme d'erreur  $\epsilon_{it}$  que les estimateurs en niveau ou en première différence. Cela peut être constaté à même le graphique A.1 où le biais est pratiquement nul lorsque le terme d'erreur suit un processus AR(1) avec  $\rho = 1$ . L'absence de biais et la convergence des estimés dans le modèle à effets fixes en niveau sont garanties si et seulement si le terme d'erreur est stationnaire, ce qui peut fréquemment ne pas être le cas en pratique. Si le terme d'erreur n'est pas stationnaire, alors l'estimateur en niveau est fallacieux et biaisé, et l'estimateur en première différence est efficace lorsque le modèle en niveau ne contient pas de bris structurels dans aucun des coefficients. Dans un tel contexte, l'estimateur en seconde différence ne sera pas efficace, sauf si le processus de moyenne mobile qui résulte de la seconde différence est pris en compte à l'aide d'un estimateur des moindres carrés généralisés. Cela n'a toutefois pas été fait dans la note afin de ne pas trop sous-estimer la variance des coefficients (ce qui augmenterait ainsi le nombre de faux positifs, soit l'erreur de type 1).

En pratique, il est toutefois possible que le terme d'erreur suive un processus autorégressif non-stationnaire positif et explosif, soit que

$$\epsilon_{it} = \rho \epsilon_{it-1} + \nu_{it},$$

où  $1 < \rho < 2$  et où  $\nu_{it}$  est une variable aléatoire i.i.d. centrée à zéro avec variance finie. Si tel est le

cas, la première différence du terme d'erreur nous donne

$$\Delta\epsilon_{it} = (\rho - 1)\epsilon_{it-1} + \nu_{it}.$$

Le problème ici est que, même si  $0 < (\rho - 1) < 1$ ,  $\epsilon_{it-1}$  est un processus non stationnaire explosif. La première différence  $\Delta\epsilon_{it}$  sera donc toujours non stationnaire explosive, ce qui risque de mener à des estimés fallacieux. En prenant la seconde différence du terme d'erreur, on obtient plutôt que

$$\begin{aligned}\Delta^2\epsilon_{it} &= (\rho - 1)\Delta\epsilon_{it-1} + \Delta\nu_{it}, \\ &= (\rho - 1)^2\epsilon_{it-1} + (\rho - 1)\nu_{it} + \Delta\nu_{it},\end{aligned}$$

ce qui atténue le problème de non-stationnarité car  $(\rho - 1)^2 < (\rho - 1)$  si  $1 < \rho < 2$ . La troisième différence permettrait encore plus d'atténuer ce problème et ainsi de suite. Notez que si  $0 \leq \rho \leq 1$ , alors les modèles en première et seconde différence vont donner des résultats identiques si le modèle est correctement spécifié. Cette équivalence ne tient pas si  $\rho < 0$ , mais une autocorrélation négative entre les termes d'erreur en niveau semble peu probable dans notre contexte, considérant l'apparente stationnarité des résidus en première différence.

Finalement, il importe aussi de noter que la seconde différence est robuste aux racines unitaires dites « doubles ». Cela est vrai si le terme d'erreur est généré de la façon suivante :

$$\Delta\epsilon_{it} = \rho\Delta\epsilon_{it-1} + \nu_{it},$$

avec  $\rho = 1$  et où  $\nu_{it}$  est i.i.d. Dans un tel cas, l'estimateur en seconde différence est effectivement l'estimateur le plus approprié pour faire en sorte que le terme d'erreur soit stationnaire. Si  $\rho$  n'est pas égal à 1 mais en est proche, l'estimateur en seconde différence permet aussi de fournir des estimés crédibles, ce qui milite davantage en sa faveur comparativement aux estimateurs en première différence et en niveau.

#### 1.4 EFFET DYNAMIQUE ET COINTÉGRATION

Une des craintes partagées par certains auteurs à propos des estimateurs en première ou en seconde différence est le fait que ces estimateurs sont moins efficaces que l'estimateur en niveau s'il existe une relation de cointégration entre deux variables non stationnaires dans le modèle (voir notamment les sections 9.2 et 9.5 de [Verbeek \(2012\)](#)). Une telle situation est très peu probable en pratique, surtout dans un contexte où les bris structurels sont directement visibles dans les séries. De plus, la relation de long terme identifiée par les coefficients de cointégration est équivalente à la relation de long terme qui peut être identifiée grâce à l'introduction d'une variable dépendante retardée dans le modèle. Cependant, il est bien connu que l'introduction d'une variable dépendante retardée dans un modèle à effets fixes en première ou seconde différence introduit un biais important dans le coefficient d'autorégression (le fameux *biais de Nickell*, voir [Nickell \(1981\)](#)). Or, les approches du type Arellano-Bond et « system GMM » souvent utilisées afin de contrer ce biais reposent sur des conditions rarement rencontrées en pratique, notamment sur l'hypothèse que la variable dépendante est assez persistante afin d'être utilisée comme instrument pour elle-même dans le futur ([Arellano et Bond, 1991](#); [Blundell](#)

et Bond, 1998; Roodman, 2009). De plus, un tel instrument ne respecte que rarement les restrictions d'exclusion en raison de la présence d'autocorrélation (d'ordre 1 ou supérieur) dans le terme d'erreur du modèle en niveau.

Pour toutes ces raisons, nous n'utilisons pas une approche de type panel dynamique afin d'estimer le modèle, ni une approche du type « correction d'erreur » (de l'anglais *error-correction model*). Il importe aussi de mentionner que de telles approches sont inutiles si l'effet de (très) long terme n'est pas l'effet recherché, car les modèles à variables *indépendantes retardées* peuvent capter une partie substantielle des effets dynamiques à moyen terme (sur 6 ans dans notre cas). De plus, ces modèles n'imposent aucune forme fonctionnelle particulière aux effets dynamiques, contrairement aux modèles auto-régressifs d'ordre 1 qui imposent que les chocs contemporains doivent se propager à un rythme constant (exponentiellement parlant) dans le futur.

## 1.5 ERREURS DE MESURE

Finalement, il importe de considérer une critique importante qui est souvent faite contre les modèles en première différence, critique qui s'applique aussi aux modèles en seconde différence et plus généralement à tous les modèles à effets fixes. La critique va comme suit : l'utilisation de tels estimateurs accentue le biais d'atténuation généré par la présence d'une erreur de mesure dans les régresseurs d'intérêt contrairement aux modèles à effets aléatoires et autres modèles en coupe transversale. Le biais s'aggrave en fonction du niveau d'autocorrélation des régresseurs d'intérêt, ce qui peut même conduire à l'estimation d'effets nuls peu importe la valeur réelle des coefficients d'intérêt. Un tel phénomène est clairement expliqué et démontré dans la section 11.5 de Wooldridge (2010).

Il importe de rappeler que le biais d'atténuation, s'il existe, est effectivement un biais d'atténuation : la valeur absolue des coefficients d'intérêt sera plus proche de zéro si un tel biais est présent lorsque les estimateurs en première ou seconde différence sont utilisés. Or, nos résultats empiriques et les simulations effectuées à la fin de ce document montrent que l'estimateur en niveau et en première différence est biaisé positivement en présence d'un bris structurel corrélé, biais qui atténue déjà la valeur des coefficients d'intérêt. Par conséquent, si l'estimateur en seconde différence conduit à des résultats qui sont plus loin de zéro comparativement aux autres estimateurs, cela signifie que le biais d'atténuation engendré par de potentielles erreurs de mesure est plus que complètement annulé par l'utilisation de l'estimateur en seconde différence. Nos résultats empiriques confirment une telle interprétation.

## 2. Somme des effets annuels et ratios bénéfiques-coûts

Pour calculer la somme des effets annuels, nous employons une propriété particulière des modèles linéaires. Si nous récrivons l'équation (1) en seconde différence ici, mais en ajoutant et en soustrayant la même quantité du côté droit de l'équation afin de maintenir l'égalité, nous obtenons la formulation

suivante :

$$\begin{aligned}
\Delta^2 \ln(y_{it}) &= \Delta^2 \ln(x_{it})^\top \beta + \sum_{k=0}^5 \gamma_k^j \Delta^2 \ln(w_{it-k}^j) + \Delta^2 \delta_t + \Delta^2 \epsilon_{it} + \Delta^2 \ln(w_{it}^j) \sum_{k=1}^5 \gamma_k^j - \Delta^2 \ln(w_{it}^j) \sum_{k=1}^5 \gamma_k^j, \\
&= \Delta^2 \ln(x_{it})^\top \beta + \Delta^2 \ln(w_{it}^j) \sum_{k=0}^5 \gamma_k^j + \sum_{k=1}^5 \gamma_k^j \Delta^2 \ln(w_{it-k}^j) + \Delta^2 \delta_t + \Delta^2 \epsilon_{it} - \Delta^2 \ln(w_{it}^j) \sum_{k=1}^5 \gamma_k^j, \\
&= \Delta^2 \ln(x_{it})^\top \beta + \Delta^2 \ln(w_{it}^j) \sum_{k=0}^5 \gamma_k^j + \sum_{k=1}^5 \gamma_k^j (\Delta^2 \ln(w_{it-k}^j) - \Delta^2 \ln(w_{it}^j)) + \Delta^2 \delta_t + \Delta^2 \epsilon_{it}.
\end{aligned}$$

On voit donc bien que le coefficient associé à la variable contemporaine  $\Delta^2 \ln(w_{it}^j)$  sera égal à la somme de tous les coefficients  $\gamma_k^j$  pour  $k \in \{0, \dots, 5\}$ . Il suffit donc simplement de retrancher  $\Delta^2 \ln(w_{it}^j)$  à chacune des autres variables retardées pour obtenir la somme directement. L'avantage d'une telle méthode est que l'écart-type calculé sur la somme des coefficients tient compte naturellement de toutes les covariances entre chacun des  $\gamma_k^j$ . Cela évite donc d'avoir à calculer manuellement la somme des éléments de la matrice de variance-covariance de  $\gamma^j$  afin d'estimer les écarts-types pour la somme des six coefficients.

Afin d'obtenir les ratios bénéfices-coûts, nous avons employé la formule suivante :

$$R_{BC} = \frac{-(\sum_{k=0}^5 \hat{\gamma}_k^j) C_{2022-2023}}{F_{2022-2023}^j},$$

où  $C_{2022-2023}$  correspond aux dépenses totales des établissements de santé au Québec pour l'année 2022-2023, et où  $F_{2022-2023}^j$  correspond aux montants versés aux organismes communautaires de la catégorie  $j$  pour la même année. Notez que ces ratios sont invariants au pourcentage d'augmentation du financement des organismes communautaires, considérant que ce pourcentage est inscrit à la fois au numérateur et au dénominateur. L'intervalle de confiance à 95% de ces ratios a été obtenu grâce à la formule suivante :

$$IC_{R_{BC}, 95\%} = \frac{-([\sum_{k=0}^5 \hat{\gamma}_k^j] \pm \hat{\sigma}_{\gamma^j} \times 1.96) C_{2022-2023}}{F_{2022-2023}^j},$$

où l'écart-type associé à la somme  $\sum_{k=0}^5 \gamma_k^j$ ,  $\hat{\sigma}_{\gamma^j}$ , est obtenu grâce à un estimateur de la variance robuste à l'hétéroscélasticité et à l'autocorrélation au sein des régions (plus précisément l'estimateur  $CV_2$ , tel que défini dans [MacKinnon et al. \(2023\)](#)). D'autres estimateurs robustes à l'hétéroscélasticité et l'autocorrélation au sein des régions ont été employés et cela ne change pas fondamentalement les conclusions de la note socioéconomique.

La somme des élasticités, les dépenses totales de santé et les montants versés aux organismes communautaires sont ajustés par groupe de régions dans la section 3.2 de la note afin de calculer des ratios bénéfices-coûts qui représentent correctement chaque groupe de régions. Finalement, prenez note aussi que les ratios bénéfices-coûts sont basés sur les montants les plus récents, ce qui est une décision relativement arbitraire. Toutefois, utiliser une autre année de référence ne change pas substantiellement la largeur des intervalles de confiance. L'idéal aurait été d'utiliser les élasticités et les montants associés à l'année 2022-2023 et aux cinq années suivantes afin d'obtenir des ratios bénéfices-coûts dynamiques plus réalistes, ce qui est toutefois impossible.

Tableau A.1 : Résultats supplémentaires du modèle en seconde différence en utilisant les dépenses de santé de 2010-2011 à 2022-2023 comme variable dépendante

Variables	Coefficients estimés		
	Mission SSS seulement (1)	Mission hors SSS (2)	Total communautaire (3)
Ordonnée à l'origine	-0,039 (0,008)***	-0,029 (0,009)**	-0,033 (0,010)**
% de gens âgés de 65 ans et +	0,918 (0,545)	0,341 (0,608)	0,687 (0,696)
% de gens âgés de 80 ans et +	-0,077 (0,243)	0,036 (0,283)	-0,015 (0,277)
% de femmes	-0,756 (2,622)	0,280 (3,153)	-0,183 (3,154)
Revenu disponible	0,178 (0,121)	0,263 (0,136)	0,200 (0,119)
Revenu disponible retardé d'un an	-0,078 (0,169)	0,071 (0,198)	0,042 (0,182)
Salaire horaire médian	0,030 (0,044)	0,022 (0,042)	0,040 (0,041)
Salaire horaire médian retardé d'un an	0,033 (0,055)	0,016 (0,053)	0,032 (0,052)
Densité de médecins spécialistes	0,077 (0,081)	0,069 (0,069)	0,060 (0,069)
Densité de médecins spécialistes retardé d'un an	-0,096 (0,071)	-0,080 (0,078)	-0,106 (0,076)
Densité de médecins généralistes	0,069 (0,107)	0,103 (0,114)	0,090 (0,111)
Densité de médecins généralistes retardé d'un an	-0,084 (0,071)	-0,090 (0,067)	-0,088 (0,066)
Taux d'emploi	-0,035 (0,050)	-0,049 (0,057)	-0,062 (0,057)
Taux d'emploi retardé d'un an	-0,089 (0,074)	-0,084 (0,071)	-0,105 (0,076)
Taux de mortalité	-0,012 (0,058)	0,023 (0,061)	0,028 (0,056)
Taux de mortalité retardé d'un an	0,016 (0,058)	0,028 (0,060)	0,042 (0,057)
Effets fixes temporels	Oui	Oui	Oui
$R^2$	0,805	0,771	0,779
Nombre d'observations	193		

**NOTES** : Les écarts-types robustes de type CV2 sont présentés entre parenthèses. \* = valeur-p < 0,05 ; \*\* = valeur-p < 0,01 ; \*\*\* = valeur-p < 0,001.

### 3. Résultats supplémentaires

Le tableau A.1 présente les résultats supplémentaires qui sont associés au tableau 1 de la note socioéconomique. Notez qu'aucune des variables incluses dans chacun des modèles n'est significative à un niveau de confiance de 95%, à l'exception de l'ordonnée à l'origine pour chaque modèle. Par souci de concision, les résultats pour les effets fixes temporels ne sont pas présentés dans le tableau A.1. Plusieurs effets fixes temporels sont toutefois fortement significatifs dans chacun des trois modèles,

notamment ceux associés à l'année 2021-2022 où les dépenses en santé ont fortement augmenté dans toutes les régions du Québec en raison de la pandémie de COVID-19 (voir le graphique 1 dans la note).

## 4. Simulations

Le processus générateur de données pour les simulations effectuées dans cette section est défini par l'équation suivante :

$$y_{it} = \begin{cases} x_{it,1}\beta_1 + x_{it,2}\beta_2 + \alpha_{i1} \times t + \delta_{i1} + \delta_t + \epsilon_{it} & \text{si } t < T_0, \\ x_{it,1}\beta_1 + x_{it,2}\beta_2 + \alpha_{i2} \times t + \delta_{i2} + \delta_t + \epsilon_{it} & \text{si } t \geq T_0, \end{cases}$$

où  $(\beta_1; \beta_2) = (-0, 2; 0, 3)$  pour toutes les simulations,  $\epsilon_{it} \sim N(0, 1)$  et  $x_{it,2} \sim N(0, 2)$  pour toutes les valeurs de  $i \in \{1, \dots, N\}$  et de  $t \in \{1, \dots, T\}$ , où  $N = 16$ ,  $T = 15$  et  $T_0 = 13$  afin de simuler les bris structurels dans les dépenses de santé causés par la pandémie dans les trois dernières années étudiées (voir le graphique 1 dans la note). La variable  $x_{it,1}$  est définie de la manière suivante :

$$x_{it,1} = \begin{cases} \tau_{i1} \times t + \lambda_{i1} + \nu_{it} & \text{si } t < T_0, \\ \tau_{i2} \times t + \lambda_{i2} + \nu_{it} & \text{si } t \geq T_0, \end{cases}$$

où  $\nu_{it} \sim N(0, \sigma_X^2)$ . Les tendances linéaires pré-bris  $\alpha_{i1}$  et  $\tau_{i1}$  sont des variables aléatoires qui suivent chacune une loi normale centrée à 0,2 avec variance 0,01 (donc écart-type de 0,1) afin de garantir qu'elles demeurent relativement faibles et majoritairement positives. Les effets fixes pré-bris  $\lambda_{i1}$  sont aussi des variables aléatoires qui suivent la même distribution. Les effets fixes  $\delta_{i1}$  et  $\delta_t$  sont définis comme des fonctions linéaires des covariables  $x_{it,1}$  et  $x_{it,2}$  de la façon suivante :

$$\begin{aligned} \delta_{i1} &= 0,3 \times \bar{x}_{i,1} - 0,2 \times \bar{x}_{i,2}, \\ \delta_t &= 0,2 \times \bar{x}_{t,1} - 0,3 \times \bar{x}_{t,2}, \end{aligned}$$

où  $\bar{x}_{i,k} = \frac{1}{T_0-1} \sum_{t=1}^{T_0-1} x_{it,k}$  et où  $\bar{x}_{t,k} = \frac{1}{N} \sum_{i=1}^N x_{it,k}$  pour  $k \in \{1, 2\}$ .

Les bris structurels dans les tendances linéaires et les effets fixes sont créés de la façon suivante :

$$\begin{aligned} \alpha_{i2} &= \alpha_{i1} \times \Delta a_i, \\ \delta_{i2} &= \delta_{i1} \times \Delta a_i, \\ \tau_{i2} &= \tau_{i1} \times \Delta a_i + \eta_{i,1}, \\ \lambda_{i2} &= \lambda_{i1} \times \Delta a_i + \eta_{i,2}, \end{aligned}$$

où  $\Delta a_i \sim N(\mu_\Delta, \sigma_\Delta^2)$  et où  $\eta_{i,k} \sim N(0, \sigma_{\eta,k}^2)$  pour  $k \in \{1, 2\}$ . Cela permet de définir les bris structurels de manière relative où les différentes tendances linéaires et effets fixes affectés par le bris bougent de façon semblable relativement parlant. Par exemple, si  $\Delta a_i = 1,1$ ,  $\sigma_\Delta^2 = 0$  et  $\sigma_{\eta,k}^2 = 0$  pour  $k = 1$  et  $k = 2$ , cela implique que les tendances linéaires et effets fixes régionaux pour chaque région augmentent de 10% à  $t=13$ . Cette augmentation relative commune s'estompe à mesure que l'on augmente la variance  $\sigma_{\eta,k}^2$  pour chaque valeur de  $k$ , ce qui devrait diminuer le biais lorsque l'on estime  $\beta_1$ , et ce,

Tableau A.3 : Biais du coefficient estimé  $\hat{\beta}_1$  pour chaque estimateur

Scénario	Biais du coefficient $\hat{\beta}_1$		
	En niveau	En première différence	En seconde différence
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,1505	0,0530	0,0383
$\mu_{\Delta} = 1, 5; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,1447	0,0520	0,0377
$\mu_{\Delta} = 0, 7; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,1494	0,0530	0,0385
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 1, 0; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,3746	0,1749	0,1347
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 1, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,5257	0,3034	0,2467
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 1; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,0069	0,0013	0,0007
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 1, 5; \sigma_{\eta,k}^2 = 0$	0,0763	0,0245	0,0174
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 0, 5; \sigma_{\eta,k}^2 = 0$	0,3699	0,1720	0,1318
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 0, 2; \sigma_{\eta,k}^2 = 0$	0,6577	0,4823	0,4211
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0, 2$	0,0955	0,0457	0,0356
$\mu_{\Delta} = 1, 1; \sigma_{\Delta}^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0, 5$	0,0323	0,0233	0,0201

**NOTES** : Des biais positifs élevés témoignent d'une plus grande sur-estimation des coefficients pour la méthode correspondante. Notez que  $\sigma_{\eta,k}^2 = \sigma_{\eta,1}^2 = \sigma_{\eta,2}^2$ .

peu importe la méthode. Cette façon de générer les bris structurels est aussi plus réaliste considérant qu'une hausse de niveau commune entre les séries est pratiquement impossible lorsque la moyenne de chaque série est très différente (comme c'est le cas pour le financement communautaire et les dépenses de santé).

L'estimation des paramètres  $\beta_1$  et  $\beta_2$  est ensuite effectuée à l'aide d'un estimateur en niveau avec effets fixes doubles (régions et années) et tendances linéaires par région, un estimateur en première différence avec effets fixes doubles (sans tendance linéaire) et un estimateur en seconde différence avec effets temporels seulement (comme celui utilisé pour obtenir les résultats présentés dans la note). Le biais sur les paramètres estimés est calculé en faisant la différence entre la moyenne des estimés obtenus (1000 répliquations par scénario) et le vrai paramètre associé. Rappelons que nous nous attendons à obtenir un biais positif important pour  $\hat{\beta}_1$  et dans une moindre mesure pour  $\hat{\beta}_2$ , biais qui devrait diminuer avec l'emploi de l'estimateur en première différence et encore plus avec l'emploi de l'estimateur en seconde différence.

C'est d'ailleurs ce que montre exactement les résultats présentés au tableau A.3. On remarque tout de suite que le biais est toujours plus petit pour l'estimateur en seconde différence que pour les estimateurs en niveau et en première différence. De plus, le biais de l'estimateur en niveau est souvent très élevé, créant une inversion de signe pour le coefficient  $\beta_1$  dont la valeur réelle est  $-0, 2$ . Autrement dit, des bris structurels fortement corrélés dans les tendances linéaires et effets fixes des régresseurs et de la variable dépendante peuvent, selon nos résultats, aisément conduire à des conclusions erronées lorsque l'estimateur en niveau est employé. Ce biais est amoindri par l'emploi d'un estimateur en

Tableau A.4 : Biais du coefficient estimé  $\hat{\beta}_2$  pour chaque estimateur

Scénario	Biais du coefficient $\hat{\beta}_2$		
	En niveau	En première différence	En seconde différence
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	-0,0002	-0,0007	-0,0011
$\mu_\Delta = 1, 5; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	0,0004	-0,0006	-0,0010
$\mu_\Delta = 0, 7; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	-0,0007	-0,0009	-0,0012
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 1, 0; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	-0,0002	-0,0007	-0,0010
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 1, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	-0,0002	-0,0007	-0,0010
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 1; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0$	-0,0004	-0,0009	-0,0012
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 1, 5; \sigma_{\eta,k}^2 = 0$	-0,00004	-0,0007	-0,0011
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 0, 5; \sigma_{\eta,k}^2 = 0$	-0,0004	-0,0008	-0,0011
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 0, 2; \sigma_{\eta,k}^2 = 0$	-0,0004	-0,0008	-0,0010
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0, 2$	-0,0014	-0,0003	0,00004
$\mu_\Delta = 1, 1; \sigma_\Delta^2 = 0, 5; \sigma_X^2 = 1, 0; \sigma_{\eta,k}^2 = 0, 5$	-0,0013	-0,0004	-0,00004

**NOTES** : Des biais positifs élevés témoignent d'une plus grande sur-estimation des coefficients pour la méthode correspondante. Notez que  $\sigma_{\eta,k}^2 = \sigma_{\eta,1}^2 = \sigma_{\eta,2}^2$ .

première différence, mais l'est encore plus par l'emploi d'un estimateur en seconde différence. Il importe aussi de préciser que, lorsque les bris structurels ne sont pas corrélés (soit quand  $\sigma_{\eta,k}^2 \gg 0$  pour  $k \in \{1, 2\}$ ), l'estimateur en seconde différence demeure moins biaisé que les deux autres, mais dans une moindre mesure car le biais disparaît lorsque  $\sigma_{\eta,k}^2$  augmente pour au moins une des deux valeurs de  $k \in \{1, 2\}$ .

Il est intéressant de noter que le biais dans  $\hat{\beta}_1$  augmente de manière importante lorsque  $\sigma_\Delta^2$  augmente ou lorsque  $\sigma_X^2$  diminue. Le premier est causé par le fait qu'une plus grande variabilité dans la taille des bris structurels pour chaque région limite la capacité de l'effet fixe temporel commun à capter ce bris considérant sa forte hétérogénéité. Inversement, une forte variabilité dans le terme d'erreur  $\nu_{it}$  diminue le biais en « masquant » par du bruit la corrélation entre les bris structurels, à l'image d'une augmentation des variances de  $\eta_{i,1}$  et  $\eta_{i,2}$ .

Finalement, le tableau A.4 montre que le biais du coefficient estimé  $\hat{\beta}_2$  est beaucoup plus petit en valeur absolue que pour  $\hat{\beta}_1$ , et ce, peu importe la méthode employée. L'estimateur en seconde différence est rarement l'estimateur à privilégier pour réduire le biais de  $\hat{\beta}_2$ , mais cela importe peu considérant la faible taille du biais en valeur absolue. Nous pouvons donc considérer que les estimés obtenus avec l'estimateur en seconde différence sont crédibles du point de vue de la direction des effets, mais sont probablement conservateurs considérant la potentielle présence d'un biais positif comme cela est montré au tableau A.3.

## Références

- Manuel ARELLANO et Stephen BOND : Some Tests of Specification for Panel Data : Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58 (2):277, avril 1991. ISSN 00346527. URL <https://academic.oup.com/restud/article-lookup/doi/10.2307/2297968>.
- Richard BLUNDELL et Stephen BOND : Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, page 29, 1998.
- Raphaël LANGEVIN : Hausser l'impôt des plus riches : des bénéfiques qui dépassent les inconvénients. *Note socioéconomique*, page 12, avril 2021. URL [https://iris-recherche.qc.ca/wp-content/uploads/2021/04/Fiscalite\\_des\\_plus\\_riches.pdf](https://iris-recherche.qc.ca/wp-content/uploads/2021/04/Fiscalite_des_plus_riches.pdf).
- James G. MACKINNON, Morten Orregaard NIELSEN et Matthew D. WEBB : Cluster-robust inference : A guide to empirical practice. *Journal of Econometrics*, 232(2):272–299, février 2023. ISSN 03044076. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304407622000781>.
- Daniel L MILLIMET et Marc F BELLEMARE : On the (Mis)Use of the Fixed Effects Estimator. *Document de travail*, septembre 2023.
- Stephen NICKELL : Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426, 1981. ISSN 0012-9682. URL <https://www.jstor.org/stable/1911408>.
- David ROODMAN : How to do Xtabond2 : An Introduction to Difference and System GMM in Stata. *The Stata Journal*, 9(1):86–136, 2009.
- Marno VERBEEK : *A Guide to Modern Econometrics*. Wiley, Hoboken, New Jersey, 4ième édition, 2012. ISBN 978-1-119-95167-4.
- Jeffrey M. WOOLDRIDGE : *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts, 2ième édition, 2010. ISBN 978-0-262-23258-6.