Statistique sur les dépenses d'assurance maladie

Méthodologie statistique

Cnamts

DSES

Dernière mise à jour : septembre 2016

SOMMAIRE

Introduction

Des statistiques mensuelles sur les dépenses de santé sont produites chaque mois par la Cnamts et diffusées sur le site de l'Assurance Maladie : www.ameli.fr > Statistiques et publications.

Elles portent sur :

- les dépenses de santé en date de remboursement, remboursées mensuellement par le régime général de l'Assurance Maladie y compris les sections locales mutualistes, en France métropolitaine ; ces dépenses sont en valeur brute ;

- les dépenses de santé en date de soins, remboursées mensuellement par le régime général de l'Assurance Maladie y compris les sections locales mutualistes, en France métropolitaine ; ces dépenses sont d'une part complétées et d'autre part corrigées des jours ouvrables et des variations saisonnières.

Les statistiques en date de remboursement relatives à un mois M donné sont publiées au début du mois M+2.

Les statistiques en date de soins relatives à un mois M donné sont publiées au début du mois M+4.

Ce document explique les méthodologies employées pour corriger les séries statistiques en date de soins. La première partie porte sur la correction des jours ouvrables et des variations saisonnières et la seconde partie présente la méthode de complétude des données en date de soins.

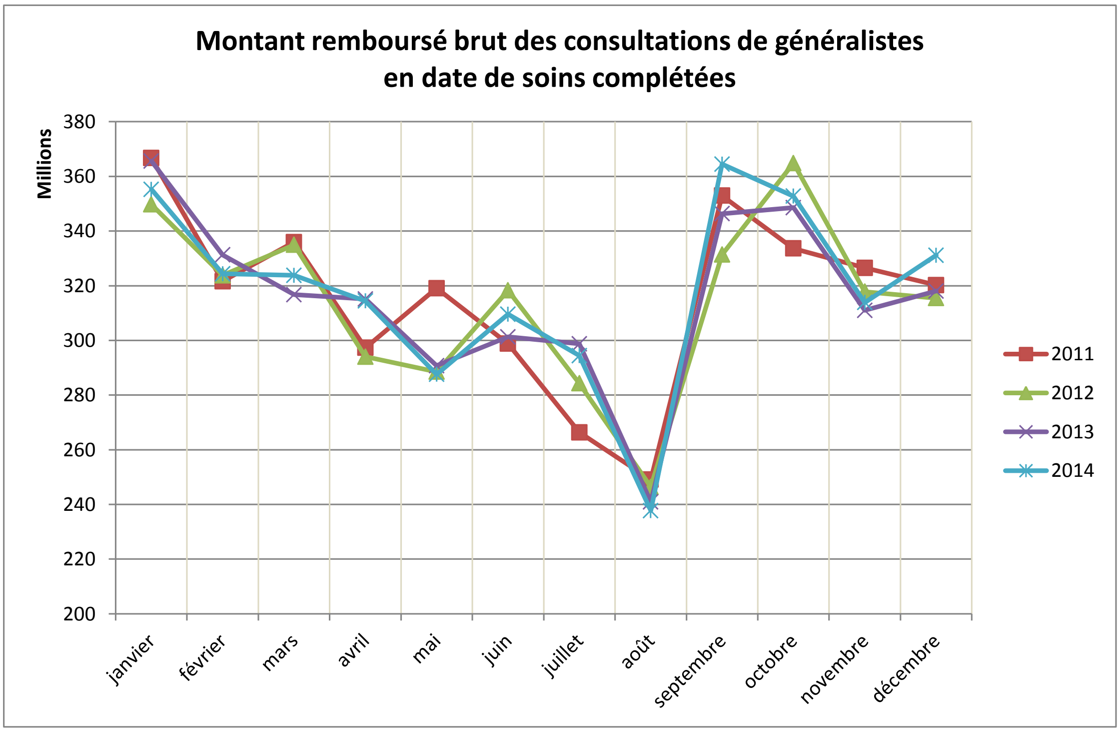
1. Correction des variations saisonnières et des jours ouvrables

1.1 Principe

Les statistiques mensuelles sur les dépenses de santé produites par la Cnamts sont des données agrégées brutes issues du système national d'information interrégimes de l'Assurance Maladie (Sniiram). Ces données reflètent les remboursements de soins pris en charge par l'Assurance Maladie.

Ces données permettent de suivre de manière conjoncturelle les dépenses de l'Assurance Maladie. Cependant, l'analyse mois par mois des séries brutes ne permet pas une analyse pertinente de l'évolution des remboursements de soins (figure ci-dessous). En effet, le « signal » est trop dilué dans des phénomènes statistiques et conjoncturels inhérents à toute série chronologique :

- des phénomènes saisonniers : par exemple, on constate une baisse de la consommation de soins en juillet et août chaque année et une augmentation en septembre. Cet effet est dû aux saisons, à des habitudes sociologiques, etc. ;
- un effet lié au nombre de jours ouvrables un mois donné, qui dépend lui-même de la composante journalière du mois. Par exemple, le mois de février comporte en général moins de jours ouvrables que le mois de mars. Si l'on compare l'évolution du mois de février par rapport à celle du mois de mars nous devons tenir compte de ce phénomène ;
- d'éventuelles mesures de revalorisations ou des changements règlementaires ayant un impact financier sur les remboursements de soins ;
- la situation conjoncturelle : l'emploi, la croissance économique, etc.



Une fois les séries en date de soins complétées, celles-ci sont corrigées des variations saisonnières et des jours ouvrables selon la méthode décrite ci-dessous, afin de permettre la comparaison des mois de soins entre eux.

1.2 Méthodologie

1.2.1 <u>Correction des mesures de revalorisations, changements règlementaires</u>

L'objectif est d'isoler les effets des variations de prix, pour ne projeter théoriquement que les variations de volume. Les séries statistiques sont déflatées d'indices de prix mensuels par branche ou par secteur :

- pour les médicaments comme pour les actes, un indice de Laspeyre[1] chaîné mensuellement, calculé en date de remboursement sur les prix réels est utilisé. Un autre indice permet également de corriger les séries de consommation de médicaments de l'impact de l'évolution du taux de générique ;
- pour les consultations médicales, la Cnamts calcule un indice de prix moyen ;
- pour les indemnités journalières, la Cnamts utilise un indice d'évolution du salaire mensuel de base publié par la Dares (Direction de l'Animation de la Recherche, des Etudes et des Statistiques) qu'elle corrige des mesures règlementaires.

[1] L'Insee utilise également un indice de Laspeyre pour le calcul de l'inflation en France.

1.2.2 Correction des jours ouvrés

La correction de l'effet des jours ouvrés part de l'hypothèse selon laquelle l'évolution d'un indicateur peut être décomposée en deux composantes non corrélées :

- une composante liée uniquement aux effets de jours ouvrés ;
- une composante corrigée des jours ouvrés (CJO), qui contient en particulier la composante saisonnière de la série.

La composante qui correspond à l'effet des jours ouvrés est alors estimée indépendamment de la saisonnalité.

La méthode de base utilisée est la suivante : il s'agit d'une régression de la variable mensuelle sur des variables représentant respectivement le nombre de lundis ouvrés, mardis ouvrés, … samedis ouvrés et de dimanches de chaque mois. (Pour ne pas intégrer d'effet propre à la saisonnalité, ces variables sont désaisonnalisées, en ne conservant que l'écart à la moyenne sur chaque mois de ces nombres de jours). L'indicateur brut en volume (c'est-à-dire corrigé des tarifs) est donc décomposé selon :

$$I_t = \alpha_1 N_{lun}^{ouv} + \alpha_2 N_{mar}^{ouv} + \ldots + \alpha_6 N_{sam}^{ouv} + \alpha_7 N_{dim} + I_t^{cjo} \qquad I_t = \sum_{i=1}^{7} \alpha_i N_t^i + \varepsilon_t$$

Où I_t est la série brute en volume mensuelle

N_{lun}^{ouv} est le nombre de lundis ouvrés désaisonnalisé, N_{mar}^{ouv} est le nombre de mardis ouvrés désaisonnalisé…

I_t^{cjo} est l'indicateur corrigé des jours ouvrés

N_t^i sont les nombres mensuels de jours ouvrés désaisonnalisés (lundis ouvrés,…, samedi ouvrés, dimanches y compris dimanches fériés).

Le modèle est toujours estimé en glissement annuel afin de travailler sur séries stationnaires (1) :

$$\Delta_{12} (I_t) = \sum_{i=1}^{7} \alpha_i \Delta_{12}(N_t^i) + \varepsilon_t \quad (1)$$

L'effet du nombre de jours fériés n'est pas identifiable. En effet, la somme des jours ouvrés, des dimanches (fériés ou non) et des jours fériés (exceptés les dimanches fériés) est égale au nombre de jours du mois. La somme des jours ouvrés et des jours fériés est presque toujours constante (à cause des années bissextiles) d'une année sur l'autre :

$$\sum_{i=1}^{7} \Delta_{12}(N_t^i) + \Delta_{12}(F_t) = 0$$

Où F_t désigne le nombre de jours fériés du mois t (non compris les dimanches fériés qui sont comptabilisés dans N_t^7).

Les coefficients α_i doivent alors être interprétés en référence aux jours fériés. Ainsi, le coefficient des lundis ouvrés ne correspond pas à l'effet brut d'un lundi ouvré, mais à l'effet de ce jour relativement à celui d'un jour férié. Par exemple, si la consommation médicale est inférieure pendant un jour férié, alors le coefficient devant le nombre de jours ouvrés est positif.

Afin de retenir dans l'estimation les effets uniquement dus aux jours ouvrés, les variables de nombres de jours ouvrés sont désaisonnalisées par écart à la moyenne, selon la méthode de Buys-Ballot[2].

Deux évaluations de l'effet des jours ouvrés peuvent être menées : une évaluation additive ou multiplicative. L'évaluation multiplicative consiste à estimer l'effet des jours ouvrés sur le logarithme de l'indicateur.

Modèle additif : $\Delta_{12}(I_t) = \sum_{i=1}^{7} \alpha_i \Delta_{12}(N_t^i) + \varepsilon_t$

Modèle multiplicatif : $(\Delta_{12}(\ln I_t)) = \sum_{i=1}^{7} \alpha_i \Delta_{12}(N_t^i) + \varepsilon_t$

[2] Celle-ci consiste à estimer l'effet saisonnier des jours fériés d'un mois donné comme la moyenne de la variable de nombres de jours ouvrés pour ce mois, sur l'ensemble des années observées, diminuée de la moyenne totale.

$$(\Delta_{12} \ln \boldsymbol{I}_t) = [\sum_{i=1}^{7} \boldsymbol{\alpha}_i (\Delta_{12} \boldsymbol{N}_t^i) + \boldsymbol{e}_t$$

$$\ln \boldsymbol{I}_t - \ln \boldsymbol{I}_{t-12} = [\sum_{i=1}^{7} \boldsymbol{\alpha}_i (\Delta_{12} \boldsymbol{N}_t^i)] + \boldsymbol{e}_t$$

$$\ln(\frac{\boldsymbol{I}_t}{\boldsymbol{I}_{t-12}}) = [\sum_{i=1}^{7} \boldsymbol{\alpha}_i (\Delta_{12} \boldsymbol{N}_t^i)] + \boldsymbol{e}_t$$

$$\frac{\boldsymbol{I}_t}{\boldsymbol{I}_{t-12}} = \exp[\sum_{i=1}^{7} \boldsymbol{\alpha}_i (\Delta_{12} \boldsymbol{N}_t^i)] + \boldsymbol{e}_t]$$

$$\boldsymbol{I}_t = \exp[\sum_{i=1}^{7} \boldsymbol{\alpha}_i (\boldsymbol{N}_t^i)] \frac{\boldsymbol{I}_{t-12}}{\exp[\sum_{i=1}^{7} \boldsymbol{\alpha}_i (\boldsymbol{N}_{t-12}^i)]} \exp(\boldsymbol{e}_t)$$

$$\boldsymbol{I}_t = \boldsymbol{cjo}_t * \boldsymbol{composante\ corrigée\ des\ jours} * \boldsymbol{résidu}_t$$

En pratique, le travail statistique pour estimer l'effet des jours ouvrés s'effectue en suivant différentes étapes :

✦ Première étape : choix de la différenciation

Le modèle est estimé en glissement annuel pour travailler sur des données stationnaires. Malgré tout, le glissement annuel de l'indicateur peut ne pas être stationnaire. **Un test de Dickey-Fuller** est effectué sur le résidu de l'équation (1). Si l'hypothèse de non stationnarité n'est pas rejetée, l'équation (1) est évaluée en différence première.

✦ Deuxième étape : test de l'existence d'effets des jours ouvrés

Il s'agit **d'un test de Fisher** sur les paramètres du modèle (1) (différencié le cas échéant). Si le test conclut à la non significativité jointe des coefficients, aucune correction des jours ouvrés n'est effectuée sur l'indicateur.

✦ Troisième étape : test de différenciation des mois de juillet et août

Les mois de juillet et août peuvent perturber l'estimation de l'effet des jours ouvrés ; ce sont des mois pendant lesquels la consommation médicale est particulièrement faible, ce qui peut conduire à atténuer l'effet des jours fériés. Pour tester l'effet du calendrier sur les mois de juillet et août, les effets sont tout d'abord évalués sur les seuls mois d'été. Un **test de Fisher** permet alors de conclure à la significativité ou non des coefficients.

S’ils ne sont pas significatifs, les coefficients des mois de juillet et août sont contraints à être nuls, c’est-à-dire qu’aucun effet des jours ouvrés n’est retranché à la série brute mensuelle en volume pour les mois d’été ; les coefficients des jours ouvrés sont alors estimés sur tous les autres mois.

Si le test conclut à la significativité des effets de calendrier pour les mois d’été, l’évaluation des coefficients est effectuée sur l’ensemble des mois de l’année, sans aucune différenciation.

✦ Quatrième étape : test des regroupements

Il s’agit de cinq tests pour retenir des regroupements courants. Il est en effet intéressant de retenir des spécifications parcimonieuses, qui aident à la compréhension des coefficients retenus et à la précision des effets estimés. Chacun des tests correspond à un test de Fisher sur des coefficients particuliers. Ces tests sont effectués avec un modèle dans lequel les résidus suivent un modèle autorégressif d’ordre 1. Les regroupements testés sont les suivants :

- le coefficient du dimanche est contraint à zéro

- les coefficients des jours de la semaine sont contraints à être égaux et le paramètre du dimanche est contraint à zéro

- les paramètres du samedi et du dimanche sont contraints à zéro

- les coefficients des jours de la semaine sont contraints à être égaux et les paramètres du samedi et du dimanche sont contraints à zéro

- les coefficients des jours de la semaine sont contraints à être égaux.

✦ Cinquième étape : existence de rattrapages ou non

Une fois le regroupement choisi, il s’agit de tester si les variables explicatives retardées et avancées d'un mois sont significatives ou non. Il est en effet possible que certains médecins rattrapent l’effet d’un jour férié en augmentant ensuite leur activité sur plusieurs semaines, ou à l’inverse anticipent cet effet sur plusieurs semaines précédant le jour férié. Dans ces deux cas, l'équation (1) est estimée en intégrant non seulement les variables de nombre de jours ouvrés du mois en cours mais également du mois précédent ou du mois suivant. Le test est alors un test de Fisher sur les variables avancées, retardées et les deux ensembles.

✦ **Sixième étape : tests de stabilité**

Il est important de tester la stabilité des coefficients estimés. Plusieurs changements de comportement pourraient en effet avoir modifié l'impact des jours fériés sur la consommation de soins. Des tests de Chow sont estimés pour chaque année. Si des ruptures apparaissent, les coefficients de jours ouvrés sont estimés sur des plages glissantes de plusieurs années.

1.2.3 Désaisonalisation à l'aide de la procédure X11

Une fois la série corrigée des effets de calendrier, il reste des effets saisonniers. Pour les corriger nous utilisons la procédure X11 de SAS.

Si nous considérons une série (Xt) telle que Xt = $(TC)_t$[3] + S_t + I_t , avec TC la composante tendance et cycle, l'algorithme de base de X11 se décompose en 8 étapes :

1. **Estimation de la Tendance-Cycle par moyenne mobile 2x12 :**
   $TC_t^{(1)} = M_{2x12}(X_t)$
2. **Estimation de la composante Saisonnier-Irrégulier :**
   $(S_t + I_t)^{(1)} = X_t - TC_t^{(1)}$
3. **Estimation de la composante Saisonnière par moyenne mobile 3x3 sur chaque mois**
   $S_t^{(1)} = M_{3x3}\left[(S_t + I_t)^{(1)}\right]$ et normalisation $Snorm_t^{(1)} = S_t^{(1)} - M_{2x12}(S_t^{(1)})$
4. **Estimation de la série corrigée des variations saisonnières**
   $Xsa_t^{(1)} = (TC_t + I_t)^{(1)} = X_t - Snorm_t^{(1)}$

5. **Estimation de la Tendance-Cycle par une moyenne de Henderson (sur 13 termes)**
   $TC_t^{(2)} = H_{13}(Xsa_t^{(1)})$
6. **Estimation de la composante Saisonnier-Irrégulier**
   $(S_t + I_t)^{(2)} = X_t - TC_t^{(2)}$
7. **Estimation de la composante Saisonnière par moyenne mobile 3x5 sur chaque mois**
   $S_t^{(2)} = M_{3x5}\left[(S_t + I_t)^{(2)}\right]$ et normalisation $Snorm_t^{(2)} = S_t^{(2)} - M_{2x12}(S_t^{(2)})$
8. **Estimation de la série corrigée des variations saisonnières**
   $Xsa_t^{(2)} = (TC_t + I_t)^{(2)} = X_t - Snorm_t^{(2)}$

Source : COMPRENDRE LA METHODE X11, Dominique LADIRAY, Benoît QUENNEVILLE 1999

2. Méthode de complétude des données en date de soins

2.1 Principe

La date de soins correspond à la date de début effective d'exécution des soins.

L'assuré dispose d'un délai de deux ans pour demander à sa caisse d'assurance maladie le remboursement de ses soins médicaux ou le paiement de ses indemnités journalières.

Le dernier mois de soins connu est le mois de soins précédant de 24 mois le dernier mois de remboursement.

Pour analyser la conjoncture récente en date de soins, il convient de prévoir les remboursements à venir pour les mois de soins les plus récents. La méthodologie de cette prévision est appelée « complétude ».

De manière simplifiée, la méthode considère qu'un mois est parfaitement connu lorsque l'on connaît 25 mois de remboursement (avec les données de remboursement de septembre 2015 on considère le mois d'août 2013 comme parfaitement connu), on ne « complète » donc plus ce mois.

Pour le mois dont on ne connaît complétement que 24 mois de remboursement (ici septembre 2013), la méthode consiste à regarder la part des soins remboursés à vingt-quatre mois sur les douze mois précédents (exemple : part des soins d'août 2013 remboursés à fin septembre 2015, la part des soins de juillet 2013 remboursés à fin août 2015, etc.) et de calculer la moyenne de ces parts observées (appelée taux de complétude) pour le mois en cours. Nous obtenons les données en date de soins complétées pour ce mois (ici septembre 2013) en faisant le rapport entre le montant des soins dont les remboursements sont connus (soins de septembre 2013 liquidés à fin septembre 2015) et ce taux de complétude. Nous recalculons ensuite les différents taux de complétude pour ce mois c'est-à-dire pour notre exemple, le taux de complétude de septembre 2013 avec 23 mois de remboursement (somme des remboursements de septembre 2013 à août 2015 /mois complété septembre 2013), puis le taux de complétude avec 22 mois de remboursements, puis 21 mois, etc.

Ensuite, on reproduit la méthode pour le mois suivant incomplet (octobre 2013) en utilisant les données en date de soins complétées du mois de septembre 2013. La méthode se répète ainsi jusqu'au mois de soin M-2 (juillet 2015) pour le mois de remboursement M (septembre 2015).

2.2 Méthodologie

2.2.1 Notations

Les notations utilisées seront les suivantes :

- i : l'indice des mois de soins

- j : l'indice des délais de remboursement (j=0,…,24)

- n : dernier mois de remboursement connu

- Y_{ij} : soins du mois i remboursés en j

		0	1	2	3	**j**	**n**
janv-13	0	10	12	14	16	**18**	20
févr-13	1	11	12,5	14	15,5		
mars-13	2	13	15	17			
avr-13	3	12	14				
…	…						
…	**i**					**Yij**	**Yin**

- C_{ij} : remboursements cumulés des soins du mois i au bout de j mois de remboursement

$$C_{ij}= Y_{i0} + Y_{i1} + Y_{i2} + \ldots +Y_{ij}$$

		0	1	2	3	**j**	**n**
janv-13	0	10	22	36	52	**54**	72
févr-13	1	11	23,5	37,5	53		
mars-13	2	13	28	45			
avr-13	3	12	26				
…	…						
…	**i**					**Cij**	**Cin**

- T_j : proportion de soins remboursés au mois j de développement (taux de complétude)

- So_i : mois de soins i complété

2.2.2 Méthodologie

- Pour les mois de soins i où Y_{i25} est connu : $So_i = C_{i25}$

- Pour le mois de soins i où Y_{ij} connus jusqu'à j et les j sont compris entre 13 et 24 :

- Calcul des T_{ij} pour les mois de soins i connus jusqu'au j[ème] mois de remboursement : Y_{ij}

$$T_{ij} = \frac{Y_{i0} + Y_{i1} + Y_{i2} + \ldots + Y_{ij}}{C_{in}} = \frac{C_{ij}}{C_{in}}$$

- Calcul de la moyenne des T_{ij} sur les douze mois antérieurs à i : $\hat{T}_{ij}$

- On obtient l'estimation du mois de soins i : $So_i = \frac{C_{ij}}{\hat{T}_{ij}}$

  Ce montant est réutilisé dans le calcul du $T_{i-1,j-1}$, et ainsi de suite jusqu'au dernier mois de soins i où Y_{i13} est connu : la <u>méthode est récursive</u>.

- Pour le mois de soins i où Y_{ij} connus jusqu'à j et les j sont compris entre 3 et 12 :

  - Calcul des T_{ij} pour les mois de soins i jusqu'au j[ème] mois de remboursement : Y_{ij}

  - Désaisonnalisation des T_{ij} à l'aide de la méthode Census X11. On note CST_{ij} le coefficient de saisonnalité correspondant, et T_{ij}^* la série des taux de complétude désaisonnalisée : $T_{ij}^* = \frac{T_{ij}}{CST_{ij}}$

  - Calcul de la moyenne des T_{ij}^* des douze mois antérieurs à i : $\hat{T}_{ij}^*$. On repasse ce taux en brut : $\hat{T}_{ij} = \hat{T}_{ij}^* * CST_{ij}$

  - On obtient l'estimation du mois de soins i : $So_i = \frac{C_{ij}}{\hat{T}_{ij}}$

Ce montant est réutilisé dans le calcul du $T_{i-1,j-1}$, et ainsi de suite jusqu'au dernier mois de soins i où Y_{i2} est connu.