POLYNOMICS

Développement de la méthode statistiques d'évaluation d'économicité

Rapport final

Étude menée sur mandat de la FMH, de santésuisse et de curafutura

Développement des méthodes statistiques d'évaluations d'économicité

Étude menée sur mandat de la FMH, de santésuisse et de curafutura

Dr Maria Trottmann, Barbara Fischer, Dr Tobias von Rechenberg, Dr Harry Telser

Septembre 2017

Remerciements
Nous remercions le prof. Dr Stefan Boes (directeur du Centre de la santé, de politique et d'économie auprès de l'Université de Lucerne, directeur du programme SNF Swiss Learning Health Systems) pour sa lecture minutieuse et ses commentaires pertinents sur une version précédente du présent rapport. Nous remercions également le groupe d'accompagnement des donneurs d'ordres, composé de Marc Bill (santésuisse), Mirjam D'Angelo (santésuisse), Dr méd. Andreas Häfeli (FMH), Thomas Kessler (FMH), Dr Philip Moline (NewIndex) et Anke Trittin (curafutura), dont les retours judicieux ont permis de nettement améliorer l'étude.

Table des matières

Liste des abréviations

Adj. R^2	Adjusted R^2; coefficient de détermination corrigé
GAS	Groupe d'âge et de sexe
AIC	Akaike's Information Criterion; critère d'information d'Akaike
ANOVA	Analysis of Variance; analyse de variance
BIC	Bayesian Information Criterion; critère d'information bayésien
COPD	Chronic Obstructive Pulmonary Disease; broncho-pneumopathie chronique obstructive
DDD	Defined Daily Dosis; dose quotidienne définie d'une substance active
GLM	Generalized Linear Model; modèles linéaires généralisés
MAPE	Mean Absolute Prediction Error; erreur de prédiction absolue moyenne
LiMA	Liste des moyens et appareils
N	Nombre d'observations
OLS	Ordinary Least Squares; méthode des moindres carrés
PCG	Pharmaceutical Cost Group; groupes de coûts pharmaceutiques
EC	Effet spécifique au cabinet
PPV	Positive Predictive Value; valeur prédictive positive
RSS	Residual Sum of Squares; somme des carrés résiduelle
SCD	Standardized Cost Difference; rapport de coûts standardisé entre les coûts observés et attendus d'un cabinet médical
Dév. std.	Déviation standard
Tarmed	Système tarifaire de rémunération des prestations médicales ambulatoires en Suisse
RCC	Registre des codes-créanciers de Sasis SA

1 En bref

1.1 Situation de départ

Selon l'art. 56 al. 6 LAMal, les assureurs et les fournisseurs de prestations doivent convenir d'une méthode visant à contrôler le caractère économique des prestations. La première étape de cette méthode identifie les cabinets dont les coûts, étant donnée la patientèle, sont nettement supérieurs aux coûts moyens du groupe de médecins spécialisés concerné à l'aide de méthodes statistiques. Ces cabinets sont ensuite contrôlés individuellement dans la suite. Depuis 2004, on utilise la méthode appelée «ANOVA» pour cette surveillance statistique (Roth und Stahel 2005, Kaiser 2016). Pour le définir approximativement, le premier niveau de cette méthode consiste à corriger les coûts moyens logarithmiques par cabinet de l'effet du groupe d'âge et de sexe (ci-après «GAS») de ses patients. Dans la seconde étape, on effectue une correction par l'influence du groupe de médecins spécialisés et du canton. La présente étude a pour but de proposer un développement de la méthode actuelle, notamment en tenant mieux compte du taux de morbidité de la patientèle. Les principales questions à se poser sont les suivantes:

1. Discussion sur le modèle estimatif sous l'éclairage de la bibliographie internationale: la spécification proposée par Kaiser (2016) est-elle adéquate ou du moins comment pourrait-on faire évoluer cette spécification?
2. Quels facteurs de morbidité supplémentaires pourraient être intégrés dans le modèle? Comment pourrait-on appliquer ces facteurs de morbidité dans la pratique?
3. Quelles caractéristiques complémentaires des cabinets médicaux, et en particulier leur emplacement, ont une influence importante sur les coûts par cabinet? Comment pourrait-on les intégrer dans le modèle?
4. Quelle est la fiabilité (nombre de cabinets faux positifs et faux négatifs) de la surveillance statistique? Quelles extensions du modèle d'estimation pourraient améliorer cette fiabilité?
5. Un calcul avec des données relatives aux patients pourrait-il nettement améliorer la fiabilité des résultats?

1.2 Évaluation et ajouts au modèle estimatif

1.2.1 Le modèle estimatif à deux niveaux

Kaiser (2016) propose d'utiliser un modèle «fixed effects» pour le premier niveau afin de calculer un effet spécifique au cabinet corrigé de l'influence de la patientèle. Intuitivement, l'effet par cabinet indique en quelle mesure un cabinet dévie des coûts moyens attendus pour un cabinet du même groupe de médecins spécialisés et du même collectif de patients. Selon nous, cette spécification est utile, mais lors de l'interprétation de l'effet par cabinet, il faudrait toutefois tenir compte du fait qu'une valeur élevée ne signifie pas obligatoirement que le cabinet manque d'efficience. Le cabinet pourrait en effet aussi présenter des particularités quant aux prestations proposées, qui feraient en sorte que ses coûts soient différents de la moyenne de son groupe de médecins spécialisés. Le modèle statistique n'est pas capable de corriger ce type d'écarts systématiques sauf s'ils se reflètent dans les facteurs de morbidité.

Dans un modèle «fixed effects» traditionnel comme celui proposé au premier niveau, il ne serait pas possible de tenir compte de facteurs qui ne sont constants par cabinet. Kaiser (2016) propose donc de corriger le résultat dans un deuxième niveau en y incluant des facteurs comme par

exemple l'emplacement du cabinet. Nous considérons que cette estimation à deux niveaux constitue une méthode valide et robuste et représente la meilleure solution pour répondre à la question posée. La littérature spécialisée propose d'autres spécifications du modèle «fixed effects». Selon nous, ces spécifications n'ont toutefois pas été suffisamment testées et elles ne se sont pas suffisamment établies pour que nous puissions recommander leur mise en œuvre opérationnelle dans l'évaluation d'économicité. De ces effets de cabinet corrigés, on calcule ensuite un indice qui, pour chaque cabinet, indique le pourcentage duquel le cabinet est supérieur ou inférieur aux coûts attendus.

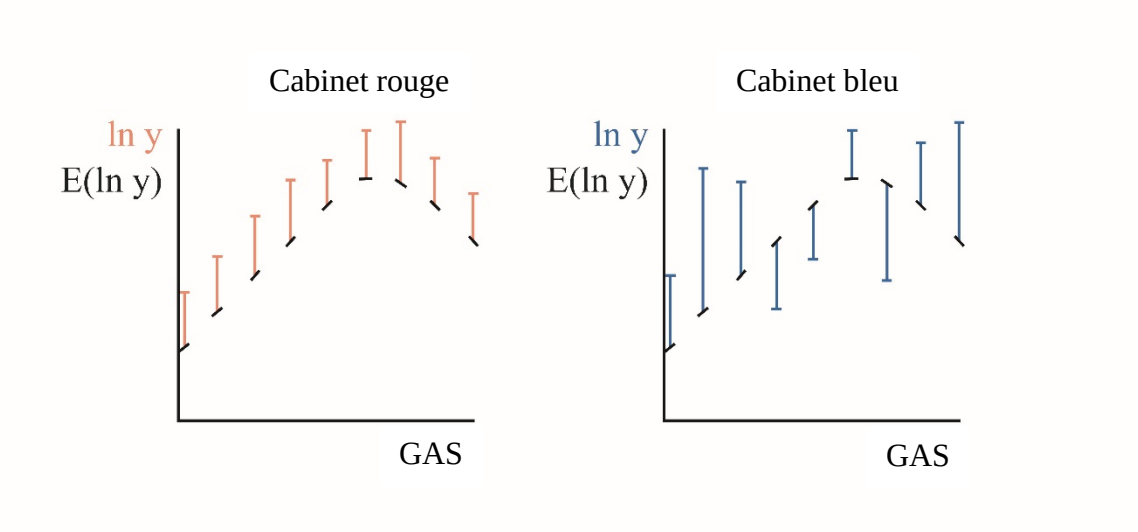
1.2.2 Ajout recommandé d'un facteur d'incertitude

Nous recommandons en complément de tenir compte du fait que l'effet spécifique au cabinet ne peut pas être précis, mais reflètera obligatoirement une certaine incertitude statistique. Pour prendre cette incertitude en compte, nous proposons de calculer un indicateur d'incertitude (qui peut être assimilé à une déviation standard). Intuitivement, l'indicateur d'incertitude peut être interprété comme suit: si un cabinet dévie dans toutes ses observations d'une valeur similaire par rapport aux coûts prévus par le modèle, le facteur d'incertitude sera bas. Mais si certaines observations dévient très fortement dans le sens positif, alors que d'autres ne dévient presque pas ou négativement, l'indicateur d'incertitude sera élevé.

La Figure 1 reprend cette théorie sur la base de deux exemples de cabinets fictifs: les points noirs représentent les résultats de régression du modèle statistique (= valeurs de coûts attendues d'un cabinet pour les variables de morbidité données). Les lignes verticales de couleur sont des déviations des valeurs observées par rapport à la valeur attendue. L'effet spécifique au cabinet correspond à la moyenne des déviations (= longueur moyenne des lignes de couleur). Le cabinet «rouge» dévie de presque la même valeur de la valeur attendue, pour toutes les observations. Dans son cas, la dispersion des valeurs et donc aussi l'indicateur d'incertitude sont faibles. Le cabinet «bleu» quant à lui est très différent: ses déviations varient fortement et n'ont pas toutes le même signe placé devant. Pour le cabinet «bleu», le facteur d'incertitude est nettement supérieur à celui du cabinet «rouge». Mais il est possible que la déviation moyenne et donc aussi l'effet spécifique au cabinet soient les mêmes pour les deux cabinets.

L'indicateur d'incertitude doit permettre de résoudre le problème des valeurs aberrantes qui modifient l'effet spécifique du cabinet même si elles n'ont très probablement pas de rapport avec le caractère économique de ce cabinet. Ces valeurs aberrantes ne concernent que quelques observations du cabinet. Si un cabinet ne présente donc que quelques valeurs supérieures, mais qui le sont nettement par rapport aux coûts attendus, il ne faut pas l'évaluer de la même manière qu'un cabinet pour lequel toutes les observations sont systématiquement supérieures aux coûts attendus. Sur le plan statistique, l'effet spécifique au cabinet peut être identifié avec une plus grande certitude si la variance des déviations par rapport à la valeur attendue (valeurs résiduelles) au sein d'un même cabinet est comparativement faible.

Figure 1 Illustration du facteur d'incertitude



Source: interne, Polynomics.

1.3 Intégration d'autres indicateurs de morbidité

1.3.1 Sélection et application pratique des indicateurs de morbidité

Dans toutes les méthodes de «physician profiling», la littérature spécialisée internationale tient compte de la morbidité de la patientèle. En Suisse aussi, la prise en compte de cet indicateur a plusieurs fois déjà été exigée (Schwenkglenks 2010; Wasem, Lux et Dahl 2010). L'objectif principal du projet consiste donc en une correction améliorée de la morbidité. Dans une analyse préliminaire, nous avons évalué plusieurs indicateurs de morbidité possibles selon les trois critères de «valeur explicative attendue», «caractère exogène» et «disponibilité des données». Puis, nous avons choisi les indicateurs suivants pour le contrôle basé sur des données: franchises à option, hospitalisation pendant l'année précédente et groupes de coûts pharmaceutiques (PCG). D'autres indicateurs liés au diagnostic sont largement répandus dans la littérature spécialisée. En Suisse, ces indicateurs ne satisfont toutefois pas au critère de la «disponibilité des données», de sorte qu'il n'a pas été possible de réaliser de test empirique correspondant.

L'objectif était de préciser les indicateurs de morbidité de manière à ce qu'ils puissent être constitués aussi bien avec les données relatives aux patients qu'avec des données agrégées. Les données agrégées (selon le GAS) proviennent du pool de données et de tarifs de Sasis SA. Il s'agit aujourd'hui de la seule source de données en Suisse qui enregistre les prestations ambulatoires payées par l'assurance-maladie obligatoire à l'échelle du pays. Les données relatives aux patients ont quant à elles été mises à disposition sous forme anonyme par trois assureurs.

- **Application de l'indicateur «franchises à option»**
  En Suisse, les patients ont le choix parmi six niveaux de franchise. Pour l'analyse, nous les avons regroupés en deux groupes: la franchise normale et la première franchise en option (CHF 300 et CHF 500 pour les adultes) sont considérées comme basses et tous les autres niveaux de franchise supérieurs sont considérés comme élevés. Cette classification en «bas» et «élevé» est largement répandue dans la littérature. Une classification plus précise avec d'autres niveaux de franchise n'apporte pratiquement pas d'amélioration au modèle.
  Au niveau du patient, le niveau de franchise représente un indicateur [0/1]. Au niveau du

cabinet, nous utilisons cet indicateur sous forme de part de patients avec un niveau de franchise élevé par GAS et cabinet.

- **Application de l'indicateur «hospitalisation pendant l'année précédente»**
  La variable «Hospitalisation pendant l'année précédente» indique si le patient a séjourné en hôpital pendant l'année précédente. Il s'agit également d'un indicateur [0/1] au niveau du patient. Au niveau du cabinet, nous le spécifions comme la part par GAS et cabinet.

- **Application de l'indicateur «groupes de coûts pharmaceutiques» («PCG»)**
  L'idée que recèlent les groupes de coûts pharmaceutiques consiste à exploiter les décomptes de médicaments pour constituer les indicateurs de morbidité pour le prévisionnel statistique des coûts. Le principal élément pour la formation du PCG réside dans un procédé de classification, qui attribue les substances actives à des domaines d'indication. Nous utilisons dans ce cas un procédé de classification développé sur mandat de l'OFSP pour la compensation des risques entre les assureurs maladie (Trottmann et al. 2015).
  Même si un procédé de classification dédié améliorerait probablement la qualité du modèle, l'option d'utiliser la compensation des risques pour la méthode de contrôle du caractère économique reste une option intéressante pour l'avenir. Ce procédé est toujours mis à jour par l'OFSP pour s'ajuster aux derniers développements dans l'approvisionnement en médicaments.
  La liste de l'OFSP contient actuellement 24 PCG. Mais de notre point de vue, il ne faudrait pas tenir compte de tous les PCG pour tous les groupes de médecins spécialisés. Concrètement, nous préconisons de ne tenir compte d'un PCG que si dans un groupe de médecins spécialisés, plus de 30 cabinets ont prescrit une quantité minimale de médicaments relevant du PCG concerné. Si peu de cabinets ont prescrit des médicaments correspondants, les coefficients PCG estimés peuvent être largement influencés par la dispersion aléatoire.
  Deuxièmement, on peut se demander comment tenir compte de la quantité de substances actives. Dans la compensation des risques, la quantité de substance active est exprimée en doses quotidiennes standardisées (DDD) pour autoriser la comparaison de médicaments différents. Nous considérons qu'il s'agit d'un procédé judicieux. La quantité DDD par observation peut dans ce cas être spécifiée comme variable continue ou il est possible de former des groupes qui dépendent de la distribution empirique des quantités DDD.

1.3.2 Test empirique des indicateurs de morbidité

Les indicateurs de morbidité «franchises en option», «hospitalisation pendant l'année précédente» et «PCG» sont intégrés dans le modèle, au premier niveau (séparément par groupe de médecins spécialisés). Ils améliorent nettement la valeur explicative statistique du modèle global pour la plupart des groupes de médecins spécialisés. Lorsqu'ils n'ont pas d'importance statistique, ils ont au moins le signe attendu (ce qui signifie que les franchises élevées ont une influence négative sur les coûts, les hospitalisations pendant l'année précédente et le PCG ont une influence positive).

Lors du calcul de l'indice, il s'avère également qu'après la correction selon le GAS et les variables de morbidité, les cabinets avec les coûts moyens les plus élevés ne sont plus «automatiquement» considérés comme suspects. Ils ne sont suspects que s'ils présentent des coûts élevés, qui ne s'expliquent pas par les variables explicatives dans le modèle. Sur la base des analyses empiriques et des conclusions de la littérature spécialisée internationale, nous recommandons d'intégrer les indicateurs de morbidité de franchise, d'hospitalisation pendant l'année précédente et du PCG dans le modèle.

1.4 Indicateurs de l'emplacement du cabinet

Dans le cadre du deuxième niveau, nous avons vérifié si en plus du canton, il était possible de tenir compte d'autres caractéristiques relatives à l'emplacement du cabinet. Concrètement, nous avons émis l'hypothèse selon laquelle le taux d'aide sociale, la densité démographique ou le nombre d'étrangers dans la commune du cabinet pouvait influencer l'effet de cabinet calculé. Dans les analyses économétriques, ces facteurs se sont toutefois avérés comme insignifiants statistiquement et d'importance limitée, de sorte qu'il n'est pas obligatoirement nécessaire de les intégrer dans le modèle. Une raison possible à l'influence limitée réside dans le fait que ces indicateurs n'existent qu'au niveau des communes et que cette grandeur est trop imprécise. Les grandes villes ont ainsi par exemple des quartiers avec des taux d'aide sociale très élevés et d'autres quartiers avec des taux faibles. La moyenne pour la commune est dans ce cas probablement peu pertinente.

1.5 Calcul d'indice et qualité du test

À partir de l'effet par cabinet corrigé, on calcule ensuite un indice qui indique le pourcentage avec lequel les coûts d'un cabinet sont supérieurs à la valeur attendue. Les cabinets dont la valeur d'indice est supérieure à 130 (c'est-à-dire 30 % au-dessus de la valeur attendue pour leur groupe de médecins spécialisés) sont considérés comme suspects et sont alors contrôlés. Pour une évaluation définitive de la qualité du test réalisé pour ce procédé, il faudrait connaître les cabinets qui fonctionnent vraiment de manière inefficiente. Cette information n'est toutefois pas disponible et il n'existe pas non plus de procédé qui pourrait améliorer de manière justifiée le caractère économique des cabinets médicaux que la surveillance statistique décrite ici (en langage technique, on dit qu'il n'y a pas de «standard de référence»).

Pour pouvoir malgré tout faire des déclarations sur la fiabilité, nous avons réalisé des simulations. Concrètement, cela signifie que nous avons attribué aléatoirement des effets de cabinet individuels à des cabinets dont près de 10 % ont dépassé la limite des 130 %. Ceci nous permet de savoir quels cabinets simulés sont inefficients et lesquels ne le sont pas. En complément, les cabinets simulés diffèrent dans leur patientèle et dans des déviations de coûts attribués aléatoirement. Nous avons ensuite répété cette attribution plus de 100 fois pour enregistrer la dispersion des valeurs ainsi trouvées.

Cette simulation permet d'analyser la probabilité avec laquelle un cabinet peut être mal évalué à cause de la dispersion aléatoire (p. ex. à cause de valeurs aberrantes élevées dues à des patients gravement malades). Les chiffres clés correspondants sont illustrés dans le Tableau 1. La valeur prédictive positive est particulièrement intéressante pour l'évaluation d'économicité. Elle indique le pourcentage avec lequel les cabinets identifiés comme suspects (valeur de test positive) sont «justement positifs». Dans le calcul d'indice avec l'estimation ponctuelle (sans prise en compte de l'incertitude), il s'agit de plus de trois quarts des cabinets testés positifs. Si on utilise l'indicateur d'incertitude en plus du calcul d'indice, ce pourcentage augmente à plus de 96 %. Ce critère plus strict augmente toutefois aussi la probabilité selon laquelle un cabinet justement positif n'est plus identifié comme tel. Cela se répercute en une sensibilité inférieure: au lieu de près de 90 % avec l'estimation ponctuelle, le total ne s'élève plus qu'à 66 %.

Lors de l'interprétation des chiffres clés, il faut faire attention à ne pas tenir compte des particularités de certains cabinets (par exemple un éventail de prestations particulier) dans la simulation.

Tableau 1 Analyse de la qualité du test d'indice de caractère économique

	Sensibilité	Spécificité	Valeur prédictive positive
Calcul d'indice avec estimation ponctuelle	89,9 %	96,9 %	76,1 %
Calcul d'indice avec limite inférieure de la plage de confiance	66,2 %	99,8 %	96,8 %

Source: interne, Polynomics.

1.6 Calculs à l'aide de données individuelles

1.6.1 Qualité du test à l'aide de données individuelles

Trois assureurs ont mis à disposition des données de décompte se rapportant à des patients individuels (en leur attribuant un pseudonyme). Nous avons analysé ces données une première fois au niveau individuel, puis une seconde fois de manière agrégée (selon la même méthode que celle des blocs de données de Sasis). Selon nos analyses, les données individuelles permettraient d'identifier beaucoup plus de cabinets comme suspects qu'avec les données agrégées. Dans la même idée, la simulation a prouvé que les données individuelles ont fait ressortir moins de valeurs faussement négatives (sensibilité supérieure). On détecte donc une plus grande part de valeurs justement positives. A contrario, nous nous sommes toutefois rendu compte que le nombre de faux positifs était également supérieur (valeur prédictive positive inférieure).

Lors du calcul des données individuelles, l'indicateur d'incertitude était systématiquement inférieur à celui avec les données agrégées. Ceci est dû au fait que les données individuelles par cabinet proposent beaucoup plus de points de données. On peut donc plus facilement séparer l'effet de cabinet systématique de la variation aléatoire.

1.6.2 Estimation

La pré-agrégation entraîne toujours une perte d'informations. Les données agrégées du pool de données ou de tarifs sont aussi beaucoup moins détaillées que les données de décompte relatives à des patients. Concernant la question qui nous intéresse, nos analyses empiriques n'ont toutefois pas prouvé que les données individuelles apportaient une nette amélioration de la surveillance. Le nombre de faux positifs était même inférieur dans nos calculs avec des données agrégées.

Il faudrait faire une nouvelle analyse en ajoutant des informations supplémentaires à la base de données. Les données basées sur le diagnostic pourraient ainsi par exemple autoriser un comparatif des coûts par «phase de traitement», méthode qui s'est imposée dans la littérature spécialisée états-unienne dans le contexte du «physician profiling».

2 Introduction

2.1 Situation de départ

La surveillance statistique constitue la première étape des évaluations d'économicité selon l'art. 56 al. 6 LAMal. L'objectif de la surveillance est d'identifier les fournisseurs de prestations qui présentent un profil de coûts suspect. Ces fournisseurs de prestations sont ensuite analysés dans le cadre d'étapes supplémentaires. Pour que ces étapes supplémentaires soient les plus efficaces

possible et ne génèrent pas d'efforts inutiles de la part des fournisseurs de prestations concernés, il est essentiel d'utiliser une méthode statistique fiable.

Depuis 2004, on utilise la méthode appelée «ANOVA» pour cette surveillance statistique (Roth und Stahel 2005). Pour la définir approximativement, la première étape de cette méthode consiste à corriger les coûts moyens logarithmiques par l'effet du groupe d'âge et de sexe (ci-après GAS). Dans la seconde étape, on effectue ensuite une correction par l'influence du groupe de médecins spécialisés et du canton.

Cette méthode doit être développée, d'une part en y intégrant des indicateurs de morbidité supplémentaires et d'autre part, le cas échéant, aussi en adaptant la méthodologie de calcul. Pour préparer cette démarche, Santésuisse a demandé au printemps 2016 une expertise à la société B,S,S. pour formaliser la méthode préalablement utilisée et mettre en avant les options d'intégration d'autres indicateurs de morbidité. L'expertise comprend notamment aussi des critères permettant de sélectionner les indicateurs de morbidité supplémentaires (Kaiser 2016).

2.2 Objectifs du projet

L'expertise Kaiser (2016) a permis de poser les bases théoriques du développement des méthodes statistiques d'évaluation d'économicité. Dans le présent rapport, ces bases sont appliquées avec les données disponibles et testées dans la pratique. Les principales questions à se poser sont les suivantes:

- Débat sur le modèle d'estimation sous l'éclairage de la bibliographie internationale: la spécification proposée par Kaiser (2016) est-elle adéquate ou du moins comment pourrait-on faire évoluer cette spécification?
- Quels facteurs de morbidité supplémentaires pourraient être intégrés dans le modèle? Comment pourrait-on appliquer ces facteurs de morbidité dans la pratique?
- Quelles caractéristiques complémentaires des cabinets médicaux, et en particulier leur emplacement, ont une influence importante sur les coûts par cabinet? Comment pourrait-on les intégrer dans le modèle?
- Quelle est la fiabilité (nombre de cabinets faux positifs ou faux négatifs) de la surveillance statistique? Quelles extensions au modèle d'estimation pourraient améliorer cette fiabilité?
- Un calcul avec des données relatives aux patients pourrait-il nettement améliorer la fiabilité des résultats?

2.3 Structure du rapport

Globalement, le travail se divise en trois parties: une partie théorie et conceptuelle, une partie empirique avec les données agrégées de Sasis SA et une partie empirique avec les données individuelles. La partie conceptuelle commence dans le chapitre 3 avec une vue d'ensemble des principales conclusions tirées de la littérature spécialisée. Dans le chapitre 4 suit un débat théorique sur la méthode à deux niveaux proposée par Kaiser (2016) et sur les évolutions possibles. Par ailleurs, ce chapitre propose une sélection d'indicateurs de morbidité dont l'intégration est ensuite testée dans la pratique.

Les chapitres 5 à 8 contiennent des analyses effectuées avec les données de Sasis SA. Dans le chapitre 5, nous décrivons comment les données sont préparées, en particulier la définition des

variables cibles et des indicateurs de morbidité. Ces variables sont utilisées dans le chapitre 6 pour calculer le premier niveau du modèle estimatif économétrique. Nous décrivons le deuxième niveau et le calcul d'indice dans le chapitre 7. Dans le chapitre 8, nous faisons une simulation qui sert à évaluer la fiabilité du procédé de surveillance statistique.

Puis dans le chapitre 9 suivent finalement les calculs basés sur les données des assureurs (y compris l'attribution à des patients individuels). Nous testons d'abord différentes variantes de calcul d'indice, puis effectuons une simulation avec pour objectif de comparer la qualité du test des données agrégées à celle du test des données individuelles. Le rapport se termine sur une synthèse dans le chapitre 10.

En annexe (chapitre 11), on trouve des analyses détaillées, des approfondissements d'ordre économétrique et d'autres descriptifs de nature technique. L'annexe s'adresse surtout aux lecteurs intéressés, disposant des connaissances statistiques correspondantes. Son but est d'éliminer les informations inutiles de la partie principale.

En raison de la complexité du sujet, le rapport principal utilise également des termes techniques statistiques et économétriques. Ces termes sont au service de la précision scientifique et de la transparence que nous visons avec le présent rapport. Le but est d'expliquer et de justifier les réflexions et applications statistiques et économétriques de la manière la plus complète possible pour autoriser un contrôle et une compréhension nets. Que ce soit dans le résumé (chapitre 1) ou dans la synthèse (chapitre 10), nous avons essayé de présenter les principales conclusions en éliminant au mieux la terminologie statistique.

3 Bibliographie

L'analyse statistique du caractère économique des médecins est appelée «physician profiling» dans la littérature spécialisée. D'importantes études sur le sujet nous viennent des États-Unis, et pour quelques-unes, aussi des Pays-Bas, c'est-à-dire des pays qui (comme la Suisse) ont un système de santé plutôt compétitif. Dans les pays dans lesquels la santé est gérée par l'État principalement, on privilégie des budgets. Lors du calcul du budget, ce sont souvent les mêmes problèmes méthodiques qui se posent que pour la considération du caractère économique. Ce budget nécessite en particulier l'enregistrement de la morbidité de la patientèle.

Pour l'étude, nous avons effectué une recherche littéraire ciblée dont le but était de répondre aux questions suivantes:

- Quel(le)s indicateurs/méthodes utilise-t-on pour ajuster les risques?
- Quelles méthodes sont utilisées pour la correction de la morbidité et pour le calcul d'indices («physician scores»)?
- Comment évalue-t-on l'adéquation/la fiabilité du procédé?

Nous avons utilisé les mots-clés suivants dans notre recherche littéraire: «physician profiling», «health care cost metrics» «risk adjustment», «capitation». Parmi le grand nombre de résultats, nous avons choisi les études qui contenaient des analyses empiriques avec des données de décompte («real data studies») et basées sur les cabinets de médecins ambulatoires.

Le chapitre est structuré en fonction des thématiques abordées dans les trois principales questions. Aucun travail individuel n'est donc résumé dans son intégralité. Citons quand même les principales sources qui proviennent de trois groupes de chercheurs:

- **Rand Cooperation (Adams 2009; Adams, Mehrotra et McGlynn 2010; Adams et al. 2010)**
  Travail de recherche sur plusieurs années des méthodes courantes aux États-Unis de physician profiling; développement d'un indicateur de fiabilité
- **University of Michigan (Thomas et al., 2004a, 2004b; Thomas und Ward, 2006)**
  Travail de recherche sur plusieurs années de comparaison entre plusieurs méthodes de physician profiling; évaluation de la fiabilité à l'aide de la simulation
- **Université de Rotterdam (Eijkenaar et van Vliet 2014; Eijkenaar et van Vliet 2013)**
  Travail de recherche sur plusieurs années d'évaluation de différentes méthodes statistiques de calcul d'indice

3.1 Correction de la morbidité

3.1.1 Informations de morbidité

Les informations diagnostiques sont de loin les principales informations de morbidité et aussi les plus souvent utilisées dans la littérature spécialisée. Les diagnostics détaillés sont enregistrés selon un système de codification unifié comme ICD-9 ou ICD-10, puis regroupés par un système de classification assisté par ordinateur (groupeur) pour en faire un nombre gérable de groupes d'analyse (Thomas et al., 2004a; Adams et al., 2010a; Wasem et al., 2010; Kristensen et al., 2014).

S'il n'y a pas d'informations diagnostiques, on forme des indicateurs de morbidité à partir des prestations utilisées observées. Les groupes de coûts pharmaceutiques ont particulièrement fait

leurs preuves dans ce cas (PCG) (Eijkenaar und van Vliet, 2013; von Rotz et al., 2008). Le système de classification (groupeur) est développé en attribuant les médicaments en fonction de leurs principales substances actives pour les domaines d'indication. Les patients qui ont acheté des médicaments du domaine d'indication concerné sont classés dans les PCG correspondants.

Un autre indicateur de l'état de santé réside dans les coûts individuels de l'année précédente (von Rotz, Kunze et Beck 2008). Ces coûts autorisent généralement une très bonne prévision des coûts pour l'année suivante. La critique que l'on exprime toutefois à ces coûts, c'est qu'ils ne sont pas indépendants du comportement du médecin (pas une variable exogène, Van de Ven et Ellis 2000).

D'autres indicateurs en corrélation avec les coûts par patient sont les caractéristiques de la région de vie comme l'urbanisation, la part d'étrangers ou des indicateurs socio-économiques (Eijkenaar et van Vliet 2013).

3.1.2 Ajustement au risque: choix du modèle

Episodes of Care

Dans la littérature états-unienne à partir du début des années 2000, on travaille beaucoup selon l'approche des «episodes of care» (Adams, Mehrotra et McGlynn 2010; Thomas, Grazier et Ward 2004a; Thomas, Grazier et Ward 2004b). Des logiciels spécialisés attribuent tous les décomptes (p. ex. prestations médicales, laboratoire, médicaments, etc.) à certains diagnostics et périodes. Dans la mesure du possible, les épisodes sont affectés aux médecins. Cette démarche a lieu selon des règles d'attribution fixes. Ces règles d'attribution sont par exemple:

1. L'épisode compte pour chaque médecin qui a fourni plus de 20 % des prestations médicales de l'épisode.
2. L'épisode compte pour le médecin avec la plus grande part de coûts médicaux, tant que ces coûts sont supérieurs à 30 %.

Dans le premier cas, la règle permet d'attribuer un épisode à plusieurs médecins dans l'évaluation (physician score). Dans les deux cas, il se peut qu'un épisode ne soit pas du tout attribué parce qu'aucun fournisseur de prestations ne peut être identifié comme responsable «décisif».

Dans le cas d'une approche basée sur un épisode, on forme beaucoup de petits groupes, qui s'excluent réciproquement. Les coûts moyens au sein du même épisode sont considérés comme comparables directement de sorte qu'on ne réalise pas de correction de morbidité supplémentaire.

Modèles de régression multidimensionnels

Les modèles de régression multidimensionnels ont pour but de calculer l'influence moyenne de facteurs explicatifs sur une variable cible (p. ex. les coûts). Cette méthode permet d'estimer les répercussions de certaines variables de morbidité (p. ex. âge, sexe, groupes de coûts pharmaceutiques) sur les coûts attendus d'un cabinet médical. L'évaluation du médecin a alors souvent lieu en comparant les coûts attendus d'un médecin avec ses coûts observés. Dans l'évaluation de cabinets médicaux, von Rotz et al. (2008), Eijkenaar et van Vliet (2013) ainsi que de Wasem et al. (2010) ont par exemple utilisé des modèles de régression..

Eijkenaar et van Vliet (2014) abordent les différents aspects du choix du modèle de manière détaillée. Ils comparent un grand nombre de méthodes de régression souvent utilisées dans la littérature sur l'économie de la santé, et notamment les Ordinary Least Squares (OLS), les Genera-

lized Linear Models (GLM), les modèles Two part ou Multilevel («Random Effects»). Ces derniers autorisent de prendre en compte le fait que plusieurs observations sont disponibles pour un médecin. L'examen empirique a permis de se rendre compte que de manière générale, les résultats n'étaient pas très différents. Les auteurs en concluent donc (citation Eijkenaar und van Vliet, 2014)):

«Differences were relatively small and the choice of model may not be as important as other choices such as the set of risk-adjusters, definition of performance index, and method for categorizing provider performance.»

Les auteurs privilégient les modèles basés sur des OLS, parce qu'ils permettent d'utiliser la même spécification pour la modélisation de variables cibles différentes et que la qualité du modèle n'est que légèrement plus mauvaise que pour des modèles plus spécifiques.

3.2 Calcul d'indice

Les résultats de la correction de morbidité doivent ensuite être reportés dans une évaluation de cabinets individuels (calcul d'indice). L'approche de calcul d'indice la plus utilisée dans la littérature est le «predictive ratio» (Thomas, Grazier et Ward 2004a; von Rotz, Kunze et Beck 2008; Wasem, Lux et Dahl 2010). Dans cette approche, les coûts observés par cabinet (i) sont comparés aux coûts attendus. Le «predictive ratio» est supérieur à un lorsque les coûts observés sont supérieurs aux coûts attendus et inférieurs à un lorsqu'ils sont inférieurs.

$$\text{Predictive Ratio}_i = \frac{\text{Valeur moyenne coûts observés}_i}{\text{Valeur moyenne coûts attendus}_i}$$

Cette approche propose plusieurs variantes concernant la manière de calculer la «moyenne des coûts attendus». Dans une approche avec des épisodes de traitement, on calcule d'abord les coûts moyens par épisode, tous médecins confondus. Les moyennes sont ensuite pondérées avec la distribution de fréquence du médecin qui doit être évalué. Dans le cas d'une approche de régression, on peut calculer les coûts attendus par patient directement. On utilise la valeur moyenne pour tous les patients du médecin à évaluer. Cette valeur correspond aux coûts fictifs de ces patients s'ils étaient traités par un médecin moyen.

En alternative au predictive ratio, Thomas et al. (2004b) calculent un indice basé sur la «standardized cost difference» (SCD). Dans ce cas, on utilise la différence absolue entre la valeur moyenne des coûts observés et attendus comme indicateur. La déviation standard des coûts attendus (σ) est alors pondérée sur tous les cabinets, divisée par la racine carrée du nombre de patients (N) du cabinet i.

$$\text{SCD}_i = \frac{\text{Valeur moyenne coûts observés}_i - \text{Valeur moyenne coûts attendus}_i}{\sigma/\sqrt{N_i}}$$

Le grand avantage du SCD réside dans le fait que le nombre de patients est pris en compte directement. Si le nombre de patients est réduit, l'indice ne peut pas être calculé avec la même certitude que pour les grands cabinets. Dans les petits cabinets, il est donc courant que les valeurs d'indice soient très élevées ou très basses (Thomas, Grazier et Ward 2004b). L'inconvénient du SCD réside dans le fait que l'on n'obtient pas de valeur d'indice interprétable comme un chiffre (p. ex. en %). Le SCD permet uniquement de classer les cabinets dans un certain ordre.

3.3 Évaluation du profiling

Par principe, on ne peut bien évaluer la qualité d'un procédé de test que s'il existe un autre procédé de test connu pour sa très grande fiabilité. Dans la littérature spécialisée, on appelle ce test «standard de référence». Pour l'évaluation du caractère économique, il n'existe toutefois pas de standard de référence, car il n'existe pas de procédé qui pourrait constater de manière précise et sans aucun doute les cabinets qui travaillent de manière économique.

Dans les études citées, deux méthodes ont été utilisées pour malgré tout évaluer la qualité du profiling. L'indicateur de la «fiabilité» a été développé par Adams et al. (2010a), puis utilisé par Eijkenaar et van Vliet (2013). L'indicateur s'appuie sur le fait que plusieurs observations sont disponibles pour chaque cabinet. Il compare la variance (σ^2) des observations *au sein* d'un cabinet avec la variance des observations *entre* les cabinets. L'influence d'un cabinet peut être déterminée avec le plus de fiabilité possible si la dispersion entre les cabinets est élevée par rapport à la dispersion au sein du cabinet.

$$\text{Fiabilité} = \frac{\sigma^2_{\text{entre les cabinets}}}{\sigma^2_{\text{entre les cabinets}} + \sigma^2_{\text{au sein d'un cabinet}}}$$

Les auteurs en arrivent donc à la conclusion que comparativement, la variance au sein d'un cabinet est important. On ne peut donc pas négliger le risque de mauvaises classifications.

Thomas et Ward (2006) utilisent une simulation pour évaluer la qualité du profiling. Ils se basent aussi sur le fait que plusieurs observations (épisodes) sont disponibles pour chaque cabinet. Dans une première étape, ils divisent les cabinets médicaux en fonction de leur SCD en trois groupes (efficient, moyen, inefficient). Puis ils forment trois blocs de données pour les différents épisodes. Le bloc de données un contient les épisodes de tous les médecins classés comme «efficients», le bloc de données deux tous les épisodes des «médecins moyens» et le bloc de données trois tous les épisodes des médecins classés comme «inefficients». De plus, ils calculent le type d'épisode survenant à quelle fréquence pour quel groupe de médecins spécialisés.

Cette base permet de simuler des cabinets fictifs. Supposons par exemple qu'il faille simuler un cabinet «efficient» de médecine générale et que le cabinet moyen de médecine générale traite 300 épisodes de «diabète» et 700 épisodes de «sinusite aiguë».[1] On procèderait ensuite à un tirage au sort aléatoire dans le bloc de données des épisodes des médecins efficients (bloc de données un), pour 300 épisodes de diabète et 700 épisodes de sinusite. Ensemble, ces épisodes constitueraient un cabinet fictif.

Puis on réaliserait un «profiling» pour les cabinets simulés. Tous les cabinets fictifs tirés d'épisodes de médecins efficients devraient aussi être évalués comme efficients. Cette démarche permet de calculer la sensibilité, la spécificité et la predictive error du test. Dans le cas des médecins généralistes («family practice»), la predictive error était de près de 20 % pour la découverte de cabinets inefficients. Près d'un cinquième des cabinets évalués comme inefficients ne provenait donc pas de la catégorie correspondante.

[1] La limitation à deux diagnostics n'est donnée qu'à titre indicatif.

4 Théorie

4.1 Modèle estimatif

Selon le mandat du projet, le cadre du modèle présenté dans l'expertise de Kaiser (2016) constitue le point de départ de nos analyses. Dans la suite, nous allons débattre des avantages et des inconvénients du modèle à deux niveaux proposé et aborderons des aspects choisis du procédé de calcul.

4.1.1 Premier niveau: modèle «fixed effects» de calcul des effets spécifiques au cabinet

La pièce maîtresse d'évaluation d'économicité réside dans le modèle «fixed effects» de l'équation (1). Le modèle définit les coûts logarithmisés y_{ij} par cabinet i et groupe d'âge et de sexe (GAS) j en fonction d'un effet par GAS, des variables de morbidité X_{ij} et d'une constante spécifique au cabinet (a_i),

$$\text{In } y_{ij} = \text{GAS}_j\beta_1 + \text{X}_{ij}\ \beta_2 + a_i + \ \varepsilon_{ij} \text{ avec } i \in \{GMS_f\} \quad (1)$$

Le fait que plusieurs observations sont disponibles par cabinet médical autorise l'estimation des constantes spécifiques au cabinet a_i (structure par panels). Pour être utilisée dans la suite, la constante est corrigée de la constante moyenne par groupe de médecins spécialisés. La différence obtenue (effet spécifique au cabinet) indique si les coûts moyens d'un cabinet sont supérieurs ou inférieurs à la moyenne totale avec les variables explicatives.

La variable ε_{ij} définit le terme d'erreur stochastique qui reprend les différences de coûts non systématiques, non explicables avec le modèle. Comme on ne peut pas partir du principe que les variables explicatives comme par exemple la structure d'âge de la patientèle aura la même influence sur les coûts pour tous les groupes de médecins spécialisés, on estime le modèle (1) séparément par groupe de médecins spécialisés (*GMS*). L'estimation est pondérée avec le nombre d'observations disponibles par GAS et cabinet.

Avantages du modèle «fixed effects»

- **Avantages par rapport au comparatif des valeurs moyennes et à la régression linéaire: la variation aléatoire est séparée de l'effet de cabinet inexpliqué**

  Dans cette section, nous abordons les avantages d'un modèle «fixed effects» par rapport à un simple comparatif des valeurs moyennes et à une modélisation des coûts de cabinet avec une régression linéaire normale.
  Dans le cas du comparatif des valeurs moyennes, on ne peut pas tenir compte de facteurs d'influence. En lieu et place, on compare directement les coûts moyens d'un cabinet avec la moyenne totale. Cette démarche part du principe implicite que tous les cabinets ont une patientèle homogène: toute déviation par rapport à la valeur moyenne est considérée comme surefficience ou sous-efficience.
  Dans le cas d'une régression linéaire, on tient compte de différences de coûts explicables pour la détermination des coûts attendus en intégrant des variables correspondantes (comme p. ex. le GAS ou le PCG) dans l'équation estimative (voir p. ex. von Rotz et al., 2008). Cette approche autorise donc l'hétérogénéité entre cabinets. Elle considère toutefois toujours toutes les déviations des coûts attendus ainsi déterminées comme sous-efficience ou surefficience.

Le modèle «fixed effects» est une extension de la régression linéaire normale en divisant les différences non expliquées par le modèle en deux composantes lors de l'estimation: d'une part une composante systématique par cabinet d dans l'équation (1), d'autre part un terme aléatoire (ε_{ij} dans l'équation (1)). Pour évaluer le cabinet, on utilise uniquement la composante systématique (a_i) . Cette composante enregistre alors non seulement tous les facteurs observables, mais aussi tous les facteurs non observables qui pourraient expliquer les différences systématiques dans les coûts spécifiques au cabinet. Pour le dire autrement, il s'agit des déviations par rapport aux coûts attendus, qui sont spécifiques pour un cabinet. Le terme aléatoire peut comprendre des erreurs de mesure ou des déviations aléatoires sur la base du besoin individuel de prestations. Il n'est pas utilisé pour évaluer le cabinet.

- **Avantage par rapport au modèle «random effects»: les variables explicatives ne doivent pas être en corrélation avec l'effet de cabinet**

  S'il y a plusieurs observations par cabinet (modèle par panel), l'effet spécifique au cabinet peut par principe être modélisé de deux manières différentes: avec le modèle «random effects» ou le modèle «fixed effects». Pour obtenir des détails sur ces deux variantes de modèle, reportez-vous par exemple à Wooldridge (2010). La variante «fixed effects» a notamment pour avantage d'autoriser les corrélations entre l'effet spécifique au cabinet a_i et les autres variables explicatives. Il est très probable que dans notre cas, il y ait des caractéristiques de cabinet non observées qui ont une influence sur la patientèle et donc sur les variables explicatives dans le modèle (1). Nous devons donc partir du principe qu'il y a des corrélations entre la constante spécifique au cabinet et les autres variables explicatives. Il convient donc d'utiliser la spécification «fixed effects».

Inconvénients du modèle «fixed effects»

- **Les caractéristiques du cabinet ne peuvent pas être utilisées directement au premier niveau**

  Dans un modèle «fixed effects» défini par l'équation (1), il n'est pas possible de tenir compte des variables qui ne varient pas pour un cabinet. L'influence de ces variables ne pourrait en effet pas être séparée de celle de l'effet spécifique au cabinet. Pour les prendre en compte, la seule manière réside dans la correction au deuxième niveau. Nous parlons de cette correction dans la section 4.1.2.

- **Toutes les particularités non observées du cabinet rentrent dans le «fixed effect»**

  On peut considérer le fait que *toutes* les particularités non observées du cabinet soient intégrées dans l'effet spécifique au cabinet pour sa mesure d'efficience comme un inconvénient de la spécification «fixed effects». Si un médecin dispose par exemple d'un catalogue de prestations particulier, lié à des coûts élevés, il aura un effet de cabinet spécifique élevé. Dans ce cas, cet effet de cabinet n'est toutefois pas dû à son inefficience, mais bel et bien à des particularités non observées de son cabinet. Il faut absolument en tenir compte pour l'interprétation des résultats.
  Dans notre contexte, cet aspect de l'estimation «fixed effects» ne joue pas un rôle important étant donné que la surveillance statistique sert à identifier les cabinets avec des profils de coûts suspects. Ces cabinets sont ensuite contrôlés dans le détail. Les particularités de certains cabinets, comme par exemple un catalogue de prestations spécifiques avec pour conséquence des coûts élevés, seront donc analysées lors du contrôle du cabinet.

Résumé

Pour le problème qui nous intéresse, ce sont les avantages du modèle «fixed effects» qui prédominent. Il se caractérise en particulier par la séparation de l'effet spécifique au cabinet de l'influence des variables explicatives et du terme d'erreur aléatoire.

4.1.2 Deuxième niveau: correction des variables spécifiques au cabinet

La spécification «fixed effects» dans le modèle (1) ne permet pas de tenir compte des variables qui sont constantes au niveau du cabinet. Il faut dans ce cas appliquer une démarche à deux niveaux (Kaiser 2016). Selon la régression «fixed effects» du modèle (1), on réalise une régression linéaire dans la deuxième étape avec l'effet de cabinet spécifique déterminé dans le premier niveau $\hat{a}_i$ comme variable dépendante. Comme variable explicative, on utilise le canton KT_i, le groupe de médecins spécialisés GMS_i et d'autres facteurs Z_i, complétés par une constante d'estimation δ_0 et d'un terme d'erreur u_i.

$$\hat{a}_i = \delta_0 + \delta_1 KT_i + \delta_2 Z_i + \delta_3 GMS_i + u_i \quad (2)$$

Les groupes de médecins spécialisés GMS_i peuvent être interprétés comme «fixed effect» spécifique au groupe de médecins spécialisés et veillent également à ce que la moyenne des valeurs résiduelles par groupe de médecins spécialisés soit égale à 0 ou 1 dans le modèle logarithmisé. L'équation (2) peut alors être modifiée en

$$\hat{u}_i = \hat{a}_i - \hat{\delta}_0 - \hat{\delta}_1 KT_i - \hat{\delta}_2 Z_i - \hat{\delta}_3 GMS_i \quad (3)$$

. Cela montre que la valeur résiduelle estimée $\hat{u}_i$ correspond à l'effet de cabinet spécifique $\hat{a}_i$ du premier niveau, corrigé des valeurs explicatives du deuxième niveau.

Avantage de la démarche à deux niveaux

L'avantage de cette démarche réside dans le fait qu'on peut utiliser un modèle «fixed effects» au premier niveau (avec les avantages cités dans la section précédente), tout en ayant la possibilité d'intégrer des variables qui sont constantes pour un cabinet médical. La littérature spécialisée aborde des approches permettant d'atteindre le même résultats avec une estimation sur un niveau. Aucune norme ne s'est établie pour l'instant (voir p. ex. Plümper et Troeger, 2007; Beck, 2011)).

Inconvénient de la démarche à deux niveaux

Le deuxième niveau peut être considéré comme une aide, étant donné que la méthode «fixed effects» n'est pas capable de tenir compte des variables constantes au niveau du cabinet. Dans la littérature, on utilise souvent ces estimations (voir p. ex. Kristensen et al., 2014). Avec l'approche à deux niveaux, il n'est toutefois pas possible de calculer la covariante des termes d'erreur des deux niveaux. Si on part par exemple du principe que les termes d'erreur de la première estimation varient plus dans certains cantons que dans d'autres, l'estimation des erreurs standards pourrait être distordue au deuxième niveau pour la variable de canton.

Résumé

De notre point de vue, le deuxième niveau décrit par Kaiser (2016) représente une démarche tout à fait valide pour corriger l'effet spécifique au cabinet des caractéristiques du cabinet, comme par exemple son emplacement.

4.2 Calcul d'indice

Dans le modèle proposé par Kaiser (2016), les coûts sont utilisés sous forme logarithmiséé (cf. équation 1). Cette transformation réduit l'asymétrie de la distribution des coûts (variable cible) et des valeurs résiduelles. L'autre avantage de la logarithmisation des coûts réside dans le fait que l'effet de cabinet spécifique ($\hat{a}_i$) ainsi que la valeur résiduelle du deuxième niveau ($\hat{u}_i$) peuvent être interprétés comme grandeurs relatives. L'effet de cabinet au prorata peut être calculé approximativement avec l'indice $100 \cdot \exp(\hat{u}_i)$.

La logarithmisation a toutefois comme inconvénient que la valeur moyenne $exp(\hat{u}_i) \neq 1$ et que la valeur moyenne de l'indice ne s'élève donc pas automatiquement à 100 (voir Kaiser, 2016, p. 8). Comme chaque cabinet doit être évalué en fonction de la valeur moyenne de son groupe de comparaison et que cette valeur moyenne est à comprendre comme coûts attendus, un cabinet avec des coûts de cette valeur moyenne devrait présenter un indice de 100.

Les valeurs d'indice sont donc normalisées avec un facteur d'échelle. La normalisation a lieu par groupe de médecins spécialisés f en multipliant $\exp(\hat{u}_i)$ par $S_f = \left(\frac{1}{\frac{1}{N}\sum_i^N \exp(\hat{u}_i)}\right) * 100$. On obtient ainsi la valeur moyenne de toutes les valeurs d'indice

$$\widehat{U}_i = S_f \times \exp(\hat{u}_i) \text{ pour tous les } i \text{ et } f \in \{1,2 \dots, F\} \quad (4)$$

et une valeur de 100 par groupe de médecins spécialisés.

4.3 Pondération

Dans le procédé proposé par Kaiser (2016), le premier niveau est pondéré par le nombre de patients par médecin et GAS. L'effet souhaité par cette approche est que les observations qui se basent sur un nombre plus grand de patients aient plus d'influence. L'estimation est donc plus précise et efficiente. Si on ne tenait pas compte des variables explicatives, le modèle «fixed effects» estimé pondéré correspondrait à un comparatif de valeurs moyennes des coûts par patient et médecin. Concernant le mode de calcul, il ne s'agit pas d'un mode différent du procédé actuel. La morbidité de la patientèle peut toutefois mieux être prise en compte.

En complément à la pondération proposée par Kaiser (2016) au premier niveau, nous préconisons de pondérer aussi la régression au deuxième niveau et le calcul du facteur d'échelle, cette fois avec le nombre de patients par cabinet. Les motifs sont les mêmes que ceux évoqués dans la section précédente.

4.4 Indicateur d'incertitude et calcul d'une limite inférieure de l'indice

4.4.1 Indicateur d'incertitude pour l'effet spécifique au cabinet

Pour faire une déclaration sur la précision de l'estimation des effets spécifiques au cabinet $\hat{a}_i$ au premier niveau, on calcule un indicateur d'incertitude (à la manière d'une erreur standard). Pour comprendre le calcul, nous vous rappelons d'abord le calcul de l'effet spécifique au cabinet ($\hat{a}_i$)). Comme l'explique Kaiser (2016, p. 6), l'effet spécifique de cabinet peut être calculé comme valeur résiduelle moyenne par cabinet ($\hat{a}_i + \hat{\varepsilon}_{ij} = \hat{\omega}_{ij}$)). L'estimation concernée par cabinet i est décrite dans l'équation (5), J_i désignant le nombre d'observations (nombre GAS) par cabinet, N_{ij} le nombre de patients par cabinet et GAS et N_i le nombre de patients par cabinet.

$$\hat{a}_i = \sum_{j=1}^{J_i} \frac{N_{ij}}{N_i}\left(y_{ij} - \hat{y}_{ij}\right) = \sum_{j=1}^{J_i} \frac{N_{ij}}{N_i}\left(y_{ij} - \left(GAS\hat{\beta}_1 + \mathrm{X}_{ij}\hat{\beta}_2\right)\right)$$
$$= \sum_{j=1}^{J_i} \frac{N_{ij}}{N_i}\ \hat{\omega}_{ij} \quad \text{mit } \hat{\omega}_{ij} = \hat{a}_i + \hat{\varepsilon}_{ij} \text{ et } \sum_{j=1}^{J_i} \frac{N_{ij}}{N_i} = 1 \tag{5}$$

Nous proposons d'appliquer un mode de calcul analogue pour le calcul de l'indice de confiance. Il se base sur les valeurs résiduelles et le nombre d'observations par cabinet. Pour constituer l'indicateur d'incertitude, nous utilisons la formule de calcul de la variance de la valeur moyenne d'une variable aléatoire x avec une distribution au choix de la probabilité et la valeur moyenne réelle μ ainsi que la variance σ_x^2. Dans le cadre d'un échantillonnage avec des observations indépendantes de n, l'estimation de la valeur moyenne $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ correspond à la variance de cette estimation

$$Var(\hat{\mu}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2} n\ \sigma_x^2 = \frac{1}{n}\ \sigma_x^2 \tag{6}$$

Pour la variance inconnue σ_x^2, nous proposons d'utiliser l'estimation ponctuelle $\hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$ (Schira 2009). Dans le modèle (1), l'effet spécifique au cabinet $\hat{a}_i$ correspond à la valeur moyenne des valeurs résiduelles $\hat{\omega}_{ij}$ par médecin, voir l'équation (5). L'estimation de la variance des valeurs résiduelles par cabinet est donc égale à:

$$\hat{\sigma}_{\hat{\omega}_{ij}}^2 = \frac{J_i}{J_i - 1}\sum_{j=1}^{J_i} \frac{N_{ij}}{N_i}\left(\hat{\omega}_{ij} - \hat{a}_i\right)^2. \tag{7}$$

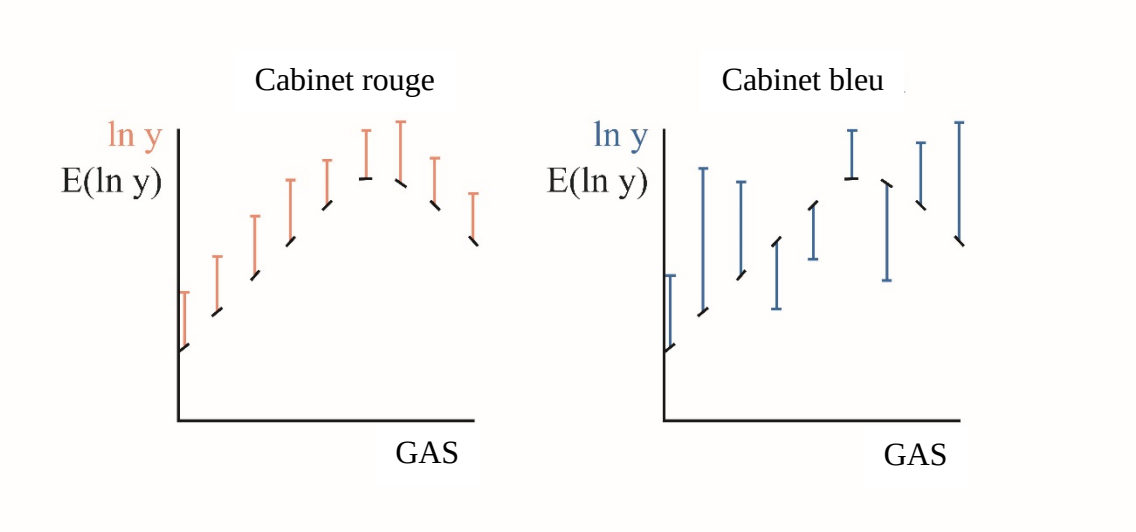
Nous définissons l'indicateur d'incertitude comme suit:

$$\text{Facteur d'incertitude}_{\hat{a}_i} = \sqrt{\frac{1}{J_i}\ \hat{\sigma}_{\hat{\omega}_{ij}}^2} = \sqrt{\frac{1}{J_i}\ \frac{J_i}{J_i - 1}\sum_{j=1}^{J_i} \frac{N_{ij}}{N_i}\left(\hat{\omega}_{ij} - \hat{a}_i\right)^2} \tag{8}$$

L'indicateur d'incertitude a pour but de représenter l'incertitude avec laquelle on peut estimer l'effet spécifique au cabinet. Intuitivement, on interprète l'indicateur comme suit: si un cabinet dévie dans tous ses GAS d'une valeur similaire par rapport aux coûts prévus par le modèle, le facteur d'incertitude sera bas. L'effet spécifique au cabinet ($\hat{a}_i$)) décrit très bien l'écart dans les différents GAS. Si le cabinet dévie toutefois par exemple très fortement dans le sens positif pour un (ou plusieurs) GAS, l'indicateur d'incertitude sera élevé.

La Figure 2 reprend cette théorie sur la base de deux exemples de cabinets fictifs: les points noirs représentent la ligne de régression (= valeurs attendues d'un cabinet pour les variables de morbidité données). Les lignes verticales de couleur sont des déviations des valeurs observées par rapport à la valeur attendue. L'effet spécifique au cabinet correspond à la moyenne des déviations (= longueur moyenne des lignes de couleur). Le cabinet «rouge» dévie d'environ la même valeur de la valeur attendue, pour tous les GAS. Dans son cas, la dispersion des valeurs et donc aussi l'indicateur d'incertitude sont faibles. Le cabinet «bleu» quant à lui est très différent: ses déviations varient fortement et n'ont pas toutes le même signe placé devant. Pour le cabinet «bleu», le facteur d'incertitude est nettement supérieur à celui du cabinet «rouge». Mais il est possible que la déviation moyenne et donc aussi l'effet spécifique au cabinet ($\hat{a}_i$)) soient les mêmes pour les deux cabinets.

Figure 2 Illustration du facteur d'incertitude



Points noirs: valeur attendue par GAS pour les variables de morbidité données (=ligne de régression)

Lignes verticales rouges: déviation du cabinet «rouge» de la valeur attendue

Lignes verticales bleues: déviation du cabinet «bleu» de la valeur attendue

La figure illustre le facteur d'incertitude. Les points noirs représentent les valeurs attendues pour un cabinet avec les variables de morbidité données (ligne de régression), les lignes verticales de couleur sont les déviations des valeurs observées par rapport à la valeur attendue (valeurs résiduelles). L'effet spécifique au cabinet correspond à la déviation moyenne (longueur moyenne des lignes de couleur). Le facteur d'incertitude reflète la dispersion des déviations autour de leur valeur moyenne. Le cabinet «rouge» dévie d'environ la même valeur pour tous les GAS. La dispersion autour de la valeur moyenne est faible. Pour le cabinet «bleu» en revanche, les déviations varient fortement pour les différents GAS. Le facteur d'incertitude est élevé.

Source: interne, Polynomics.

L'indicateur d'incertitude ne tient pas compte de l'incertitude de l'estimation globale des coefficients bêta pour les GAS et les variables de morbidité (dans la Figure 2, il s'agit de la ligne de régression noire). Nous considérons qu'il s'agit d'une démarche valide étant donné que la priorité est donnée à l'incertitude de l'effet *spécifique au cabinet*.[2]

Il faut également préciser que cette démarche est différente du calcul d'erreurs standards robustes des clusters, comme on pourrait par exemple l'utiliser pour évaluer la signification statistique des variables explicatives au 1er niveau du modèle (équation (1)). Les erreurs standards robustes des clusters tiennent compte de l'hétéroscédasticité ainsi que de l'*autocorrélation* au sein d'un cluster. Dans le cas présent, le cluster correspond au cabinet. S'il faut donc calculer l'erreur standard pour l'indicateur «franchise», les données de tous les cabinets médicaux sont intégrées dans ce calcul. On peut donc considérer que la dispersion des termes d'erreur est probablement différente pour tous les cabinets.

[2] La démarche est identique à celle que l'on aurait dans une étude empirique si on considérait uniquement les erreurs standards des coefficients bêta qui doivent être interprétés pour les résultats de ladite étude. Si l'on calculait l'estimation Fixed Effects avec une variable muette (dummy)individuelle par médecin et si l'on calculait des erreurs standards hétéroscédastiques robustes pour les variable muette (dummy), l'on obtiendrait des erreurs standards similaires à l'indicateur d'incertitude proposé ici. La démarche nécessite en revanche beaucoup plus de calculs, en particulier pour les groupes de médecins spécialisés comptant beaucoup de médecins.

Pour calculer l'erreur standard d'un effet spécifique au cabinet $\hat{a}_i$, on utilise toutefois que les valeurs d'*un* cabinet (un autre effet spécifique au cabinet s'applique alors à tous les autres cabinets). Dans le cas présent, il n'existe donc qu'un seul cluster et il n'y a pas de composante dans la dispersion des termes d'erreurs qui pourrait être différente entre clusters. Il n'est donc pas possible de procéder au calcul d'erreurs standards robustes par cluster pertinentes dans notre cas. Si dans cette situation, on utilisait la formule courante pour les erreurs standards robustes par cluster pour l'effet spécifique au cabinet, on obtiendrait des erreurs standards bien trop basses, étant donné que les déviations positives et négatives par rapport à la valeur moyenne s'annuleraient au sein d'un cabinet. La littérature recommande donc aussi de ne commencer à calculer les erreurs standards robustes par cluster qu'à partir d'un nombre minimum de 50 clusters (Cameron et Miller 2015).

4.4.2 Intégration de l'indicateur d'incertitude dans la constitution de l'indice

L'indicateur d'incertitude $\sqrt{\frac{1}{J_i}\,\hat{\sigma}^2_{\hat{\omega}_{ij}}}$ peut servir à calculer une plage de confiance pour l'effet spécifique au cabinet. Dans notre cas, nous sommes particulièrement intéressés par la limite inférieure, parce que la priorité consiste à ne pas identifier trop de cabinets comme suspects par erreur. Nous calculons cette limite inférieure en déduisant 1,96 fois l'indicateur d'incertitude de l'estimation ponctuelle. Pour en arriver là, nous suivons le calcul habituel d'intervalle de confiance de 95 en supposant que les termes d'erreurs sont presque distribués normalement.

$$\hat{a}_i^{low} = \hat{a}_i - 1.96 * \sqrt{\frac{1}{J_i}\,\hat{\sigma}^2_{\hat{\omega}_{ij}}} \tag{9}$$

La valeur limite inférieure de l'effet spécifique au cabinet permet de calculer une valeur limite inférieure pour la valeur d'indice d'un cabinet spécifique. Pour y arriver, on utilise, dans l'équation (3), la limite inférieure de l'effet $\hat{a}_i^{low}$ au lieu de l'estimation ponctuelle $\hat{a}_i$. Pour le reste, le calcul de l'indice ne change pas, même en ce qui concerne le facteur d'échelle. L'objectif de la limite inférieure n'est donc pas de calculer un nouvel indice dont la valeur moyenne serait à nouveau 100. Le but est plutôt de calculer un intervalle de confiance vers le bas pour chaque cabinet, en plus de l'indice existant.

4.5 Transformation des variables cibles

La transformation logarithmique de la variable cible proposée par Kaiser (2016) offre un avantage net pour le calcul des modèles. Si on ne la transformait pas, la variable cible (dépenses de santé) serait fortement asymétrique vers la droite. Même en cas de distribution asymétrique positive, la méthode des moindres carrés (Ordinary Least Squares, OLS) fournit des estimations cohérentes. Plus le nombre d'observations est grand, plus les coefficients estimés s'approchent donc de la réalité. Dans le cas de petits échantillonnages, les coefficients estimés peuvent toutefois manquer de précision. Il est également possible que des coefficients soient «distordus» vers le haut par quelques valeurs très élevées (aberrations) et surestiment donc l'influence réelle (Mihaylova et al. 2011). La transformation logarithmique apporte des valeurs résiduelles presque distribuées normalement (voir évaluations empiriques dans le chapitre 11.3 en annexe). Par ailleurs, elle permet de calculer directement un indice sous forme de grandeur relative.

Mais la transformation logarithmique a aussi ses inconvénients, en particulier lorsque l'objectif consiste à obtenir une prédiction finale en francs absolus. Nous reviendrons sur ces problèmes

dans le chapitre 11.1 en annexe. Dans de nombreuses applications pratiques, les auteurs privilégient l'estimation d'un modèle de coûts non transformé, surtout quand l'objectif de l'analyse est la prédiction de coûts (Buntin et Zaslavsky, 2004; Beck, 2013). De notre point de vue, il n'est pas possible de décider définitivement, en se basant sur des conceptions purement théoriques, si une logarithmisation des variables cibles est indiquée ou non. Dans l'analyse empirique, nous testerons donc deux modèles supplémentaires: un modèle non transformé sans correction des valeurs aberrantes et un modèle avec correction des valeurs aberrantes (winsorisation). Une comparaison empirique est proposée dans la section 11.3 en annexe.

4.6 Intégration d'autres indicateurs de morbidité

L'un des objectifs principaux de cette analyse est de vérifier l'intégration d'autres indicateurs de morbidité dans la surveillance statistique. La littérature spécialisée va d'ailleurs aussi dans ce sens. Dans tous les articles que nous connaissons sur le «physician profiling», on utilise des indicateurs concernant l'état de santé (voir section 3.1). En Suisse aussi, l'intégration d'autres indicateurs de morbidité a déjà été exigée par plusieurs expertises (Schwenkglenks 2010; Wasem, Lux et Dahl 2010).

La première étape consiste à choisir des indicateurs, que nous contrôlerons de manière empirique dans la suite. Les indicateurs doivent répondre aux trois critères suivants (Kaiser 2016, section 3.1):

- **Valeur explicative/pertinence des coûts:** les indicateurs de morbidité doivent avoir une forte influence sur les coûts.
- **Caractère exogène:** les indicateurs de morbidité ne doivent être que peu influençables par le comportement du médecin (pas de fausses incitations ou «moral hazard»).
- **Disponibilité et qualité des données:** les données brutes disponibles doivent permettre d'obtenir des indicateurs de morbidité valides à intégrer dans le modèle.

Le Tableau 2 reprend une liste de variables de morbidité possibles, qui sont évaluées selon les critères de la valeur explicative, du caractère exogène et de la disponibilité des données. Dans la dernière colonne, on voit si l'indicateur est ou non examiné dans les analyses empiriques.

Les informations diagnostiques représentent l'indicateur de morbidité de la patientèle le plus direct et le plus répandu dans la littérature (voir p. ex. Thomas et al., 2004a; ainsi que Adams et al., 2010b)). Dans les données de décompte de l'assurance-maladie suisse, ces données ne sont toutefois pas disponibles. L'état de santé doit donc être estimé en se basant sur des indicateurs de remplacement («proxies»).

Le *niveau de franchise* choisi représente un bon proxy pour l'état de santé d'un patient. Les personnes saines sont les plus à même de choisir un niveau de franchise élevé et plusieurs études ont par ailleurs montré que les coûts moyens de personnes ayant souscrit une franchise élevée étaient nettement inférieurs à ceux de personnes avec une franchise normale (Gardiol, Geoffard et Grandchamp 2005; Schmid und Beck 2015). Par ailleurs, les médecins n'ont que peu d'influence sur le niveau de franchise (variable exogène).

À la manière des niveaux de franchise, les personnes plus saines auront aussi tendance à choisir certains *modèles d'assurance* (p. ex. offres de télémédecine). Mais selon nous, cet indicateur ne doit pas être intégré dans le modèle, et ce pour deux raisons principales: premièrement, le modèle d'assurance n'est pas indépendant du comportement du médecin. Si des cabinets présentent des

coûts inférieurs pour certains modèles d'assurance corrigés du risque, comme l'indique la littérature spécialisée (Kauer 2016), ceci se reflètera aussi dans leurs valeurs d'indice. Deuxièmement, cette variable est difficile à enregistrer dans les données: les modèles sont en effet très différents en fonction de l'assureur, même si leur nom est presque le même.

Tableau 2 Indicateurs de morbidité possibles

	Valeur explicative attendue	Caractère exogène	Disponibilité des données	Test empirique
Franchises	Bonne, justifiée dans les études	Bon, part endogène faible	Bonne dans les données des médecins, très bonne dans les données individuelles	Oui
Modèle d'assurance	Moyenne, dépend du modèle concret	N'est pas indépendant du comportement du médecin	Comparabilité peu claire	Non
Hospitalisations pendant l'année précédente	Peu claire, justifiée dans les études pour les coûts totaux de l'AOS	Moyen, décrit la prise en compte	Très bonne	Oui
Groupes de coûts pharmaceutiques (PCG)	Très bonne, justifiée dans les études	Moyen, décrit l'utilisation, problématique pour les traitements médicamenteux qui sont substituables	Bonne dans les données des médecins, très bonne dans les données individuelles	Oui
Caractéristiques de l'emplacement du cabinet	Peu claire	Bon, part endogène faible	Limitée par la protection des données[1)]	Oui
Canton	Sans correction des effets de prix: claire; avec correction des effets de prix: peu claire	Très bon	Limitée par la protection des données[1)]	Oui
Informations de Tarmed	Élevée pour des groupes de médecins spécifiques, p. ex. pour l'identification de cabinets suspects avec activité chirurgicale)	Dépend de la conception	Bonne, spécification détaillée nécessaire pour l'application	Probablement dans le troisième sous-projet
Groupes de coûts diagnostics	Très bonne, justifiée dans les études	Bon, marge de décision présente	Non disponible	Non

1) Cette limitation s'applique à Polynomics, qui est une agence externe. Elle s'applique toutefois moins aux partenaires tarifaires qui réalisent l'évaluation d'économicité.

Dans cette étape, on choisit les indicateurs de morbidité qui sont ensuite examinés dans l'analyse empirique. Les indicateurs de morbidité sont évalués selon les critères de «valeur explicative attendue», de «caractère exogène» et de «disponibilité des données». Les indicateurs «franchise», «hospitalisation pendant l'année précédente», «groupes de coûts pharmaceutiques» et «variables d'emplacement» sont proposés pour le test empirique de cette étape.

Source: interne.

Les deux indicateurs «*hospitalisation pendant l'année précédente*» et «*groupes de coûts pharmaceutiques*» font déjà l'objet de débats depuis de nombreuses années dans la littérature de compensation des risques (Beck 2013). Depuis 2012, l'hospitalisation pendant l'année précédente est d'ailleurs intégrée dans la compensation des risques. Les groupes de coûts pharmaceutiques le

seront à partir de 2020. Pour les deux indicateurs, il est prouvé qu'ils peuvent contribuer de manière substantielle à la prédiction des coûts, du moins en ce qui concerne les coûts totaux à la charge de l'assurance-maladie obligatoire.

Les deux indicateurs «hospitalisation pendant l'année précédente» et «groupes de coûts pharmaceutiques» sont constitués à partir de la prise en compte préalable des prestations. S'il n'y a pas d'informations diagnostiques, il s'agit de la deuxième meilleure alternative afin de calculer le besoin en prestations de la patientèle. Mais n'oublions pas qu'utiliser des variables de morbidité à partir de la prise en compte d'une prestation n'est pas sans risque. Premièrement, cette prise en compte n'est pas indépendante du comportement du médecin. Ainsi, une prise en compte élevée par le passé peut être due à une patientèle particulièrement malade, mais aussi à d'autres facteurs. Deuxièmement, le fait d'utiliser des traitements pour créer des «facteurs de morbidité» alors qu'ils sont substituables par d'autres traitements peut être à l'origine d'erreurs d'évaluation. Si un suivi médical interne permet ainsi par exemple d'éviter un traitement médicamenteux, la prise en compte des groupes de coûts pharmaceutiques implique le risque que les cabinets qui misent sur des traitements non médicamenteux soient jugés de manière «stricte» parce que leur patientèle sera classée comme plus saine qu'elle ne l'est en réalité. Malheureusement, les outils statistiques ne permettent pas d'évaluer l'importance du risque associé aux groupes de coûts pharmaceutiques. Il faut en tenir compte pour l'interprétation du contenu et l'évaluation des modèles. L'évaluation du contenu pourrait avoir pour conséquence que certains groupes de coûts pharmaceutiques ne seraient pas pris en compte, même si, statistiquement, ils ont une influence significative sur les coûts.

5 Base de données et préparatifs

Dans la section suivante, nous allons décrire comment nous avons choisi les variables cibles, ainsi que la démarche d'opérationnalisation des indicateurs de morbidité choisis avec les pools de tarifs ou de données de la Sasis. Nous nous sommes limités aux principaux aspects de préparation des données pour le modèle. Les informations complémentaires sur l'agrégation et les liens entre les données sont disponibles dans le chapitre 11.2 en annexe.

Les données du pool de tarifs ou de données de la Sasis ne contiennent pas d'identificateur pour des patients individuels.[3] Elles sont pré-agrégées à différents niveaux: les données relatives aux prestations sont agrégées par groupe d'âge, sexe, franchise, modèle d'assurance et type de sinistre;[4] les données des malades sont agrégées par cabinet, groupe d'âge, sexe et hospitalisation pendant l'année précédente et les données de création des groupes de coûts pharmaceutiques sont agrégées par cabinet, groupe d'âge et de sexe.

Pour pouvoir créer un lien entre les blocs de données, nous agrégeons tous les blocs au niveau du cabinet et des groupes d'âge et de sexe (GAS). Les informations relatives aux franchises et aux hospitalisations pendant l'année précédente ne sont pas oubliées, elles servent à créer les variables de morbidité (voir section 5.2). La variable du modèle d'assurance ne sera pas pris en compte dans le modèle (voir section 4.6). Nous estimons que la valeur informelle des variables Type de sinistre (différence entre maladie, accident et maternité) est faible, de sorte qu'il n'est pas nécessaire d'en tenir compte.

5.1 Constitution des variables cibles

5.1.1 Affectation des prestations aux RCC anonymes

Pour constituer les variables cibles, il faut affecter les coûts des cabinets médicaux, identifiés par des numéros RCC anonymes. Les deux colonnes de données «Responsable» ou «Créancier» servent à cela. Les coûts des types de prestations qui sont fournis directement dans le cabinet sont affectés au créancier. On les appelle aussi coûts *directs*. Ils sont énumérés dans le Tableau 3.

Les traitements médicaux décomptés selon Tarmed représentent de loin la plus grande part des types de prestations directes. Il faut faire attention parce qu'ils sont mesurés en *points tarifaires* et pas en francs. Ainsi, la quantité de prestations fournies peut être comparée à l'échelle des cantons. La présente étude ne se demandera pas si les différentes valeurs des points tarifaires sont justifiées.[5]

Il faut également tenir compte du fait que pour les coûts directs, seuls les coûts facturés par le cabinet en lui-même sont affectés à chaque cabinet. Si un médecin envoie un patient à un deuxième médecin, seules les prestations que le premier médecin a fournies sont comptabilisées pour celui-ci. Les prestations du second médecin sont quant à elles imputées au second médecin. Cette affectation repose d'une part sur des raisons liées à la technique des données: en Suisse, les patients n'ont pas besoin de présenter de demande de transfert s'ils souhaitent consulter plusieurs

[3] Pour les évaluations sur la base de données individuelles, voir le chapitre 9.

[4] Techniquement, le «type de prestation» existe comme niveau de regroupement supplémentaire. Mais pour nous, il ne vaut pas comme un regroupement, mais nous l'utilisons pour créer la variable cible (voir section 5.1.1).

[5] Dans le cas d'une réalisation opérationnelle du modèle, il faudrait toutefois se demander si une pondération avec une valeur de point tarifaire unifiée (p. ex. 0,88 comme moyenne pondérée de toutes les valeurs de point tarifaire utilisées en Suisse) ne serait pas plus adaptée qu'une pondération avec la valeur un. La pondération avec une valeur de un signifie que les prestations médicales sont plus valorisées que d'autres prestations (p. ex. médicaments).

médecins. Les transferts demandés entre médecins ne figurent donc pas tous dans les données de facturation. D'autre part, on ne peut pas non plus clairement définir sur le plan du contenu en quelle mesure un médecin qui envoie un patient ailleurs peut contrôler les coûts des autres médecins.

Tableau 3 Les coûts directs sont affectés au créancier anonyme

	Nombre d'observations	Nombre de créanciers	Valeur totale (CHF)[a)]
Traitement médical Tarmed	12 486 072	21 912	7 233 188 488
Médecin ambulatoire/médicaments	4 878 863	16 486	1 821 441 960
Médecin ambulatoire/analyses	4 170 960	13 057	503 105 202
Médecin ambulatoire/autres tarifs	1 457 463	14 684	133 859 594
Médecin ambulatoire/LiMA	796 714	18 921	173 098 286
Médecin ambulatoire/physiothérapie	6199	1571	2 055 681
Total	**23 796 271**		**9 866 749 211**

Les chiffres ont été recueillis en 2015. Seuls les créanciers avec au moins un patient dans le registre des malades sont représentés.

[a)] Il s'agit de points tarifaires pour les prestations Tarmed.

Le tableau reprend les types de prestation pour lesquels les coûts sont imputés au numéro RCC anonyme du créancier («coûts directs»). Toutes les prestations Tarmed sont concernées. Dans le cas d'un transfert d'un médecin vers un autre, la quantité de prestations fournie par chacun est affectée au médecin concerné.

Source: données du pool de données Sasis, registre des prestations 2015, calculs internes.

Le Tableau 4 reprend les types de prestations dont les coûts sont imputés au médecin qui réalise la prestation. Il s'agit des types de prestations que les patients suisses ne peuvent pas facturer à l'assurance-maladie obligatoire sans prescription médicale. La plupart de ces prestations sont des médicaments, suivies par les examens en laboratoire et les prestations de physiothérapie.

Tableau 4 Les coûts indirects sont affectés au responsable anonyme

	Nombre d'observations	Nombre de responsables	Valeur totale (CHF)
Pharmacies/médicaments A&B, autres taxes méd. obligatoires, analyses	6 320 434	21 230	2 398 443 794
Laboratoires/analyses	3 216 904	16 795	693 527 971
Physiothérapeutes/physiothérapie	1 023 183	15 434	650 295 932
Point de fourniture/LiMA	336 314	13 640	228 261 054
Pharmacies/LiMA	628 644	16 409	121 615 503
Total	**11 525 479**		**4 092 144 254**

Les chiffres ont été recueillis en 2015. Seuls les responsables avec au moins un patient dans le registre des malades sont représentés.

Les types de prestations illustrés ici («coûts indirects») sont affectés au numéro RCC anonyme du prestataire responsable. Les coûts des médicaments représentent de loin la plus grande part.

Source: données du pool de données Sasis, registre des prestations 2015, calculs internes.

5.1.2 Agrégation, logarithmisation et winsorisation

Par principe, la variable cible de l'estimation est la moyenne des coûts par cabinet et GAS. Nous disposons en moyenne de 35 observations par cabinet, mais moins pour les groupes de médecins spécialisés en gynécologie, ainsi qu'en médecine pour enfants et adolescents (19 observations en moyenne).

Comme expliqué dans la section 4.5, nous testons trois spécifications différentes de ces variables cibles: la transformation logarithmique (Kaiser 2016), les coûts non transformés et la «winsorisation». Dans le cas de la logarithmisation, il n'est pas possible de tenir compte des coûts qui sont nuls ou négatifs. Ces valeurs sont en effet dues à des annulations, raison pour laquelle on les exclut de toutes les estimations.

La winsorisation est réalisée par groupe de médecins spécialisés, sur le 95^{e} centile. Le 95^{e} centile correspond à la valeur pour laquelle 95 % des coûts observés sont inférieurs et 5 % supérieurs. Les valeurs observées supérieures au 95^{e} centile sont «tronquées» à la valeur du 95^{e} centile. Elles ne sont donc pas exclues et ont toujours une valeur (nettement supérieure au reste de la distribution). Mais elles ne sont pas prises en compte entièrement dans l'estimation.

5.2 Constitution des indicateurs de morbidité

5.2.1 Constitution des indicateurs «franchise» et «hospitalisation pendant l'année précédente»

Les données de la Sasis ne permettent pas de voir spontanément les franchises qu'ont souscrites les différents malades. Dans le registre des prestations, on peut toutefois voir le nombre de consultations par cabinet, GAS et niveau de franchise. Nous avons donc calculé la part de consultations ayant eu lieu pour des assurés avec des franchises élevées par cabinet et GAS. Sont considérés comme franchises élevées tous les niveaux de franchise d'au moins CHF 1000 pour les adultes et CHF 200 pour les enfants. La variable a le format d'une variable continue, d'une valeur comprise entre zéro et un.

Nous spécifions l'indicateur «hospitalisation pendant l'année précédente, qui est utilisé depuis 2012 dans la compensation des risques, à la manière des franchises. Nous calculons pour chaque cabinet et GAS la part de patients ayant été hospitalisée l'année précédente. Ainsi, nous obtenons également une variable continue avec des valeurs comprises entre zéro et un.

5.2.2 Constitution des groupes de coûts pharmaceutiques

Pour constituer les groupes de coûts pharmaceutiques, nous avons utilisé une classification PCG développée pour la compensation des risques entre assureurs maladie. Elle représente une adaptation des groupes de coûts pharmaceutiques, qui a été utilisée en 2014 pour la compensation des risques aux Pays-Bas (Trottmann et al. 2015). Elle contient 24 groupes de coûts pharmaceutiques.

La classification n'a pas été développée exprès pour les évaluations d'économicité, et ce pour des raisons liées aux ressources principalement: développer notre propre concept n'aurait pas été réalisable dans le cadre financier et dans les délais de ce projet. On peut toutefois s'imaginer qu'en définissant des PCG spécifiques par groupe de médecins spécialisés, on pourrait nettement améliorer la valeur explicative des modèles.

Pour l'application opérationnelle de la classification des PCG avec les données existantes, nous avons procédé comme suit.

Calcul de la quantité de substances actives dans les doses quotidiennes standardisées (quantité DDD)

Pour le classement des PCG, il faut rendre les quantités de substances actives de différents médicaments comparables entre elles. Dans la classification PCG de la compensation des risques, on utilise des «doses quotidiennes standardisées» (Defined Daily Dosis, DDD). Cette valeur de mesure[6] créée par l'OMS correspond à une «dose moyenne par jour pour l'indication principale pour adultes» par substance active et forme pharmaceutique. Il ne s'agit aucunement d'une recommandation de traitement et le dosage pour les patients peut être nettement différent. Mais cette valeur de mesure s'est établie dans les calculs statistiques.

La liste PCG contient un groupe de substances actives (code ATC) par PCG, qui est attribué au PCG concerné et pour chacune des substances actives, les formes pharmaceutiques et le DDD par forme pharmaceutique. La figure 3 montre une représentation schématisée de l'affectation d'un code pharmaceutique à un PCG et le calcul de la quantité DDD. Le code pharmaceutique contient 4,5 mg de la substance active Indacatérol (code ATC: R03AC18) en poudre à inhaler (Inhal.powder). Le code ATC et la forme pharmaceutique permettent de relier le code pharmaceutique à la liste PCG. Le code ATC correspond à «COPD/asthme sévère» dans le PCG (PCG 2) et la DDD pour la forme pharmaceutique «poudre à inhaler» s'élève à 0,15 mg. Un emballage de 4,5 mg contient donc 30 DDD.

Figure 3 Affectation des PCG et calcul de la quantité DDD

(1) Code pharmaceutique ou GTIN

Code pharmaceutique	Nom du produit	AtcCode	AdmCode	Quantité de substance active par emballage
4603089	Onbrez Breezhaler, Inh Kaps 0.150 mg, 30 Stk	R03AC18	Inhal.powder	4.5 mg

Combinaison

(2) Affectation PCG

Désignation PCG	N° PCG	AtcCode	AdmCode	DDD
COPD/asthme sévère	2	R03AC18	Inhal.powder	0.15 mg

(3) Classification du code pharmaceutique ou GTIN

Pharmacode	N° PCG	Quantité DDD par emballage
4603089	2	30

La figure montre l'exemple d'une classification d'un emballage de médicament (code pharmaceutique) dans un PCG. Pour la classification, on associe le code pharmaceutique à la liste PCG. Les variables pour cette association sont le code ATC et la forme pharmaceutique. La liste PCG quant à elle attribue le code ATC à une domaine d'indication (PCG) et détermine la quantité de substance active applicable à ce code ATC et à cette forme pharmaceutique comme dose quotidienne standardisée (DDD). On peut ainsi calculer la quantité DDD par code pharmaceutique.

Source: interne, Polynomics.

[6] Voir: https://www.whocc.no/ddd/definition_and_general_considera/, consulté le 25.9.2017.

Les données du pool de tarifs de la Sasis contiennent les différents codes pharmaceutiques facturés par cabinet et GAS. Nous les avons associés à la liste PCG pour calculer la quantité DDD totale facturé par PCG, cabinet et GAS.

Exclusion de PCG spécifiques par groupe de médecins spécialisés

Un PCG spécifique n'est pris en compte pour un groupe de médecins spécialisés que si plus de 30 médecins de ce groupe ont prescrit une quantité minimale ou plus de médicaments du PCG concerné. La quantité minimale pour qu'un cabinet soit comptabilisé a été fixée à une DDD faible de 1,8. Ceci correspond à une dose semestrielle (180 DDD définies) pour 100 patients.

La raison à cette restriction à 30 médecins par groupe de médecins spécialisés repose sur la stabilité des coefficients estimés. En effet, si un groupe de médecins spécialisés ne présentait que peu d'observations avec des médicaments du PCG concerné, il serait problématique d'en tenir compte. Les coefficients estimés pourraient dépendre trop largement d'observations individuelles et ne pas décrire de lien général.

Hiérarchisation

Les PCG utilisés dans la compensation des risques aux Pays-Bas impliquent une «hiérarchisation» de ces PCG. Les domaines d'indication similaires sont regroupés en une hiérarchie (exemple: asthme et COPD). Si un patient satisfait aux critères d'admission pour plusieurs PCG de la même hiérarchie, il n'est attribué qu'au PCG le plus sévère (exemple: un patient qui a reçu des médicaments aussi bien pour l'asthme que pour le COPD ne sera affecté qu'au groupe COPD). Les Pays-Bas ont choisi de procéder à cette hiérarchisation pour limiter les possibilités de manipulation étant donné que la compensation des risques distribue beaucoup d'argent. Mais sur le plan statistique, cette hiérarchisation a plutôt tendance à rendre les résultats plus mauvais: les modèles de prédiction sans cette restriction présentent une meilleure valeur explicative (Trottmann et al. 2015).

La hiérarchisation ne peut par ailleurs pas être appliquée sur des données agrégées. Chaque PCG a donc été pris en compte séparément. Une observation peut ainsi par exemple avoir une valeur positive aussi bien pour le PCG «asthme» que pour le PCG «CODP». On obtient un changement du contenu en supprimant la hiérarchisation pour les PCG 23 (diabète type II avec hypertension) et PCG 8 (diabète type II sans hypertension). Dans la compensation des risques, un patient ne peut être classé dans le PCG 23 que s'il satisfait simultanément aux critères du PCG 8. Mais comme il n'est pas possible d'évaluer quels patients ont reçu des médicaments des deux domaines d'indication, les deux PCG 8 (diabète type II) et 23 (hypertension) sont considérés séparément. Cela conduit à une nette augmentation du nombre de patients dans le PCG 23.

Prise en compte de la quantité de substances actives

Comme nous l'avons évoqué plus haut, nous avons calculé la quantité de substance active (quantité DDD) par PCG, cabinet et GAS dans les données agrégées. Cette quantité DDD doit être utilisée dans la régression comme variable explicative. Il existe plusieurs options pour l'appliquer concrètement. Premièrement, on peut intégrer la quantité DDD directement comme variable continue dans le modèle. Ceci suppose implicitement un lien linéaire entre la quantité DDD prescrite et les coûts par GAS et cabinet. En alternative, il est aussi possible de créer des groupes axés sur la distribution empirique des quantités de substance active (pour le dire plus communément, on forme des groupes pour «quantité faible de substance active», «quantité moyenne de substance active», etc.). Cette deuxième alternative a pour avantage que l'on peut renoncer à l'hypothèse

d'une relation linéaire. Mais l'inconvénient réside dans l'apparition possible d'effets de seuil: si une valeur est proche de la limite de groupe, une petite modification de la quantité de substance active modifie son classement et entraîne donc une prévision de coûts différente.

Dans l'application empirique, nous avons choisi la deuxième option. Pour constituer les groupes, nous avons considéré qu'il était important que chaque groupe dispose d'un nombre suffisant d'observations. Le nombre de groupes a donc été adapté au nombre d'observations disponibles par PCG et groupe de médecins spécialisés. Les seuils pour la répartition dans chaque groupe ont été choisis comme suit:

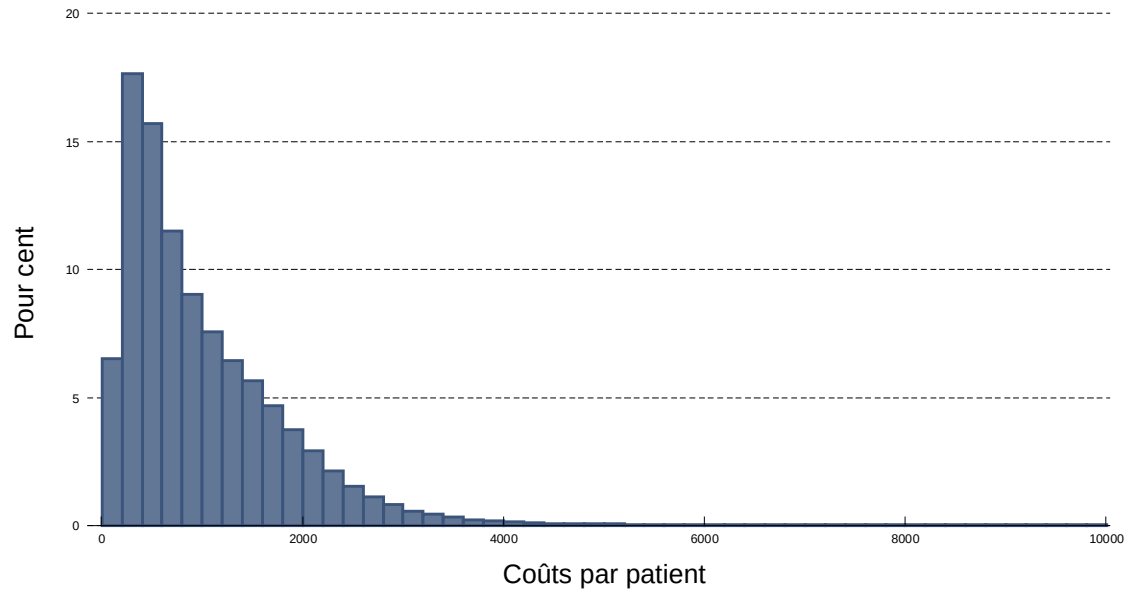
- Catégorie dédiée pour la quantité DDD = 0
- S'il y avait plus de 10 000 observations d'une valeur supérieure à zéro, nous avons créé dix catégories correspondant aux déciles de la répartition.
- S'il y avait entre 1000 et 10 000 observations d'une valeur supérieure à zéro, nous avons créé quatre catégories par quartiles.
- S'il y avait entre 0 et 1000 observations d'une valeur supérieure à zéro, nous avons créé deux catégories. La séparation suit la ligne médiane.

6 Analyses empiriques, premier niveau

6.1 Distribution des variables cibles avant et après la transformation

Les dépenses de santé sont typiquement distribuées avec une forte asymétrie positive. Cela signifie qu'il existe beaucoup d'observations avec des coûts très bas et un petit groupe d'observations avec des coûts très élevés. La figure 4 illustre cet état de fait en montrant la répartition des coûts par patient pour le groupe de médecins spécialisés «Médecine interne générale». La première barre qui contient les observations d'un coût jusqu'à 200 francs comprend près de 6 % des observations. La ligne médiane qui sépare les observations avec les 50 % de coûts les plus élevés des 50 % de coûts les plus bas est de 867 francs, et donc largement inférieure à la valeur moyenne de CHF 1026. L'asymétrie est égale à 6,45, ce qui correspond à une asymétrie nette du côté positif.

Figure 4 Coûts non transformés par patient, médecine interne générale



N = 205 430, 258 observations avec des coûts > CHF 10 000 non représentées.

Quand elles ne sont pas transformées, les dépenses de santé présentent une large asymétrie positive, ce qui signifie que la plupart des observations ont des coûts faibles (50 des observations ont des coûts jusqu'à 867 francs par patient) alors qu'une petite minorité présente des dépenses nettement supérieures.

Source: données de Sasis SA, année 2015, calculs internes.

Le modèle «fixed effects» décrit dans la section 4.1 est estimé avec la méthode de régression habituelle des moindres carrés (OLS). Indépendamment de la distribution des variables cibles, cette méthode fournit des estimations cohérentes. Plus le nombre d'observations augmente, plus les coefficients estimés se rapprochent des valeurs réelles. Dans le cas de petits échantillonnages, les coefficients estimés peuvent toutefois manquer de précision. Dans le cas d'une distribution asymétrique positive, il est alors possible que des coefficients soient «distordus» vers le haut par quelques valeurs très élevées (aberrations) et surestiment donc l'influence réelle .

Une première réduction de l'asymétrie positive est atteinte en «tronquant» les coûts très élevés (winsorisation). Concrètement, cela signifie que les valeurs observées supérieures au 95^{e} centile

sont «tronquées» à la valeur du 95e centile. Elles ne sont donc pas exclues et restent largement supérieures à la plus grande partie de la distribution, sans être aussi extrêmes qu'avant la winsorisation. Comme le montre le Tableau 5, la valeur moyenne baisse de CHF 20, soit près de 2 % suite à la winsorisation chez les médecins généralistes. L'abaissement des valeurs extrêmes réduit donc la somme totale des coûts pris en compte de près de 2 %. L'asymétrie de la distribution diminue nettement et la déviation standard baisse également.[7]

Tableau 5 Distribution des variables cibles, médecine interne générale

	N	Valeur moyenne	Dév. std.	p25	p50	p75	p95	Asymétrie	mas.
Non transformé	205 425	1026	738	503	867	1392	2261	99 783	6,45
Winsorisé	205 425	1004	617	503	867	1392	2261	2591	0,80
Logarithmisé	205 425	6,71	0,70	6,22	6,76	7,24	7,72	11,51	-0,41

Observations pondérées par le nombre de malades.

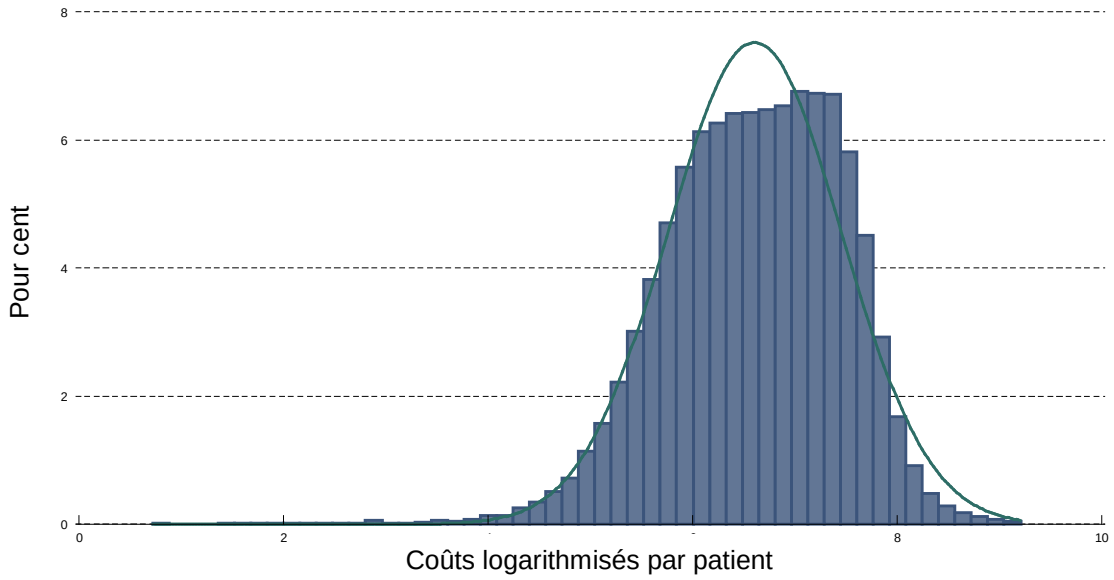
Lorsqu'ils ne sont pas transformés, les coûts présentent une forte asymétrie positive: la ligne médiane se trouve nettement au-dessus de la valeur, l'asymétrie a une valeur élevée de plus de six et les 5 % les plus chers (valeurs comprises entre le 95e centile et le maximum) sont répartis sur une fourchette beaucoup plus que grande que toutes les autres valeurs. Après la transformation logarithmique, la distribution est nettement plus symétrique. L'asymétrie devient même négative, la distribution présente une légère asymétrie négative.

Source: données de Sasis SA, année 2015, calculs internes.

L'effet de la transformation logarithmique est illustré dans la figure 5, la ligne verte représente la fonction de densité de la distribution normale. On voit clairement que la distribution est désormais presque symétrique. Les valeurs plutôt faibles sont écartées par la logarithmisation, les valeurs élevées sont tronquées. Les chiffres clés de la distribution aussi indiquent une variable à distribution presque normale. La ligne médiane se trouve près de la valeur moyenne et l'asymétrie est même légèrement négative (distribution à asymétrie légèrement négative).

[7] On s'aperçoit que le maximum dans le modèle winsorisé ne correspond pas à la valeur du 95e centile de la distribution non transformée. La raison réside dans le fait que les valeurs sont pondérées par la somme de personnes malades (comme c'est le cas pour la régression) dans la figure. Mais pour la winsorisation, le 95e centile a été calculé sans pondération. Etant donné que les valeurs aberrantes concernent souvent les valeurs avec peu de malades, le 95e centile calculé non pondéré est supérieur à la valeur pondérée.

Figure 5 Médecine interne générale: coûts logarithmisés par patient



N = 205 172, 258 observations avec des coûts > CHF 10 000 non représentées.
Après la transformation logarithmique des données de départ, la distribution des coûts est presque symétrique. La ligne tracée présente une distribution comparativement normale.

Source: données de Sasis SA, année 2015, calculs internes.

Le Tableau 6 représente la distribution pour des groupes de médecins spécialisés choisis. Dans le groupe des médecins spécialisés en médecine pour enfants et adolescents, l'asymétrie est extrêmement élevée. Cette valeur extrême est due à quelques valeurs aberrantes très élevées. Il s'agit d'observations avec des coûts de médicaments très élevés qui, certes, sont extrêmes, mais peuvent arriver (pas obligatoirement une erreur de données). Mais elles ont un nombre de patients réduits, ce qui fait qu'elles sont donc peu pondérées dans la régression.

Concernant les médecins en ophtalmologie, les coûts totaux sont réduits le plus par la winsorisation. Dans le modèle winsorisé, la valeur moyenne est 16 % inférieure à la valeur moyenne dans les données originales. Dans ce groupe de médecins spécialisés, les 5 % supérieurs de la distribution comprennent donc non seulement de «vraies aberrations» (avec normalement peu de malades), mais aussi une grande partie des coûts. Probablement, une particularité dans les prestations fournies par une partie de ce groupe de médecins spécialisés est à l'origine des coûts nettement différents.

Tableau 6 Distribution des variables cibles pour des groupes de médecins spécialisés choisis

	N	Valeur moyenne	Dév. std.	p25	p50	P75	p95	Asymétrie	max.
Chirurgie									
Non transformé	15 543	588	424	323	481	724	1303	31 674	5,07
Winsorisé	15 543	561	314	323	481	724	1303	1348	0,97
Logarithmisé	15 543	6,19	0,60	5,78	6,18	6,58	7,17	10,36	0,13
Gynécologie									
Non transformé	23 401	531	284	353	466	671	1035	33 678	5,22
Winsorisé	23 401	520	240	353	466	671	1035	1044	0,52
Logarithmisé	23 401	6,13	0,59	5,87	6,14	6,51	6,94	10,42	-1,28
Cardiologie									
Non transformé	14 295	870	368	671	825	992	1431	23 727	6,59
Winsorisé	14 295	852	274	671	825	992	1431	1593	0,65
Logarithmisé	14 295	6,70	0,36	6,51	6,71	6,90	7,27	10,07	-0,30
Médecine pour enfants et adolescents									
Non transformé	17 115	445	742	304	407	547	751	248 449	226,12
Winsorisé	17 115	437	179	304	407	547	751	1042	0,78
Logarithmisé	17 115	6,00	0,45	5,72	6,01	6,30	6,62	12,42	-0,47
Ophtalmologie									
Non transformé	34 238	534	641	265	366	521	1515	25 118	4,93
Winsorisé	34 238	449	277	265	366	521	1202	1202	1,53
Logarithmisé	34 238	5,99	0,65	5,58	5,90	6,26	7,32	10,13	1,23
Psychiatrie et psychothérapie									
Non transformé	59 562	2368	1322	1535	2145	2898	4645	30 885	2,40
Winsorisé	59 562	2328	1145	1535	2145	2898	4645	5727	0,91
Logarithmisé	59 562	7,63	0,57	7,34	7,67	7,97	8,44	10,34	-0,89

Observations pondérées par le nombre de malades.

L'asymétrie de la distribution est la plus élevée pour le groupe des médecins spécialisés en médecine pour enfants et adolescents. Pour l'ophtalmologie, la winsorisation joue un rôle important dans les coûts totaux pris en compte, la valeur moyenne baisse de 15 %. Une grande partie des coûts se trouve en «queue de distribution».

Source: données de Sasis SA, année 2015, calculs internes.

6.2 Indicateurs de morbidité

6.2.1 Groupes d'âge et de sexe (GAS)

Dans le procédé ANOVA utilisé aujourd'hui, on exploite déjà les groupes d'âge et de sexe pour calculer l'indice. Même s'il ne s'agit pas d'indicateurs directs de l'état de santé, ils restent en corrélation avec les dépenses de santé. Pour constituer l'indicateur, on regroupe cinq années. À partir de 96 ans, il ne reste plus qu'un seul groupe.

Le Tableau 7 indique les chiffres clés d'un modèle de régression qui contient uniquement le GAS et un effet spécifique au cabinet comme variables explicatives. Avec un coefficient de détermination corrigé (Adjusted R^2) de près de 70 à 90 %, on peut affirmer que la valeur explicative est bonne. La valeur explicative du GAS est la plus faible pour les coûts des groupes de médecins spécialisés en médecine pour enfants et adolescents et pour les psychiatres et psychothérapeutes. Dans le cas de la médecine pour enfants et adolescentes, le coefficient R^2 est comparable à d'autres modèles, mais l'erreur de prédiction absolue moyenne (Mean Absolute Prediction Error, MAPE) est nettement supérieure. Cela peut indiquer qu'il y a de fortes aberrations, mais qu'elles s'expliquent facilement par les variables du modèle. Le coefficient R^2 dépend largement de l'explication que l'on peut apporter aux valeurs aberrantes. Il est donc relativement élevé. Le MAPE, qui dépend beaucoup moins des valeurs aberrantes, présente donc une valeur comparativement moins bonne.

Tableau 7 Valeur explicative d'un modèle avec seulement le GAS selon les groupes de médecins spécialisés

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants/ adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	205 430	15 543	23 401	14 295	17 115	34 238	59 562
N médecins	5455	481	1221	425	1020	871	2487
Adj. R^2	0,817	0,729	0,892	0,700	0,745	0,862	0,427
MAPE	0,330	0,347	0,292	0,246	0,508	0,259	0,478

Variable cible: transformée par logarithmisation, pondération de la régression avec le nombre de malades par GAS

N: taille de l'échantillon; adj. R^2: coefficient de détermination corrigé; MAPE: erreur de prévision absolue moyenne.

Le rapport entre les groupes d'âge et de sexe et les dépenses de santé est clairement établi. Un modèle qui ne comprend que le GAS et un effet de cabinet spécifique atteint des valeurs R2 de près de 70 à 90 % (avec les données agrégées). La plus petite valeur explicative est atteinte pour le groupe des médecins spécialisés en psychiatrie et psychothérapie, ainsi que pour la médecine pour enfants et adolescents.

Source: données de Sasis SA, année 2015, calculs internes.

6.2.2 Indicateur «franchise»

Une franchise élevée vaut surtout le coup pour les personnes en bonne santé et n'ayant pas besoin de beaucoup de prestations. Plusieurs études ont montré que les coûts moyens de personnes ayant souscrit une franchise élevée étaient nettement inférieurs à ceux de personnes avec une franchise normale (Schmid et Beck, 2015; Gardiol et al., 2005).

Le Tableau 8 représente les coefficients de cet indicateur pour des groupes de médecins spécialisés choisis. Le modèle d'estimation économétrique correspond au modèle complet avec GAS, franchise, hospitalisation pendant l'année précédente et PCG des variables explicatives. La variable cible est logarithmisée.

Pour la plupart des groupes de médecins spécialisés, on voit le rapport négatif attendu entre la part de franchises élevées et les coûts. Mais dans le groupe des médecins spécialisés en cardiologie, il n'a pas d'importance statistique. Un rapport positif, contraire aux hypothèses, s'observe dans le groupe des médecins spécialisés en gynécologie. La gynécologie représente un cas à part dans la mesure où les prestations en lien avec la maternité sont exclues de la franchise. Pour les

patientes qui consomment beaucoup de ces prestations, une franchise élevée peut toujours être intéressante.

Tableau 8 Influence des franchises, modèle global par groupes de médecins spécialisés

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants/ adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	205 430	15 543	23 401	14 295	17 115	34 238	59 562
N médecins	5455	481	1221	425	1020	871	2487
Coefficient modèle log.	-0,298** (0,00)	-0,069** (-0,00)	0,090** (0,01)	-0,018 (0,31)	-0,445*** (0,00)	-0,128*** (0,00)	-0,068*** (0,00)

Coefficients du modèle global avec GAS, franchises, hospitalisation pendant l'année précédente et PCG, variable cible logarithmisée. Erreur standard entre parenthèses. Niveau de signification: *** p<0.01, ** p<0.05, * p<0.1.

Le tableau représente les coefficients de l'indicateur «franchise» pour des groupes de médecins spécialisés choisis. L'indicateur est significativement négatif pour la plupart des groupes de médecins spécialisés. Dans le groupe des médecins spécialisés en gynécologie, on observe toutefois un effet positif important.

Source: données de Sasis SA, année 2015, calculs internes.

Pour des raisons de place, nous n'avons pas représenté les coefficients des autres groupes de médecins spécialisés, ils sont disponibles dans la feuille Excel en annexe «Polynomics_Wirtschaftlichkeitspruefungen_Regressionskoeffizienten_Stufe_1». De manière générale, nous avons trouvé des coefficients négatifs significatifs dans 20 groupes de médecins spécialisés. Dans 10 groupes de médecins spécialisés, les effets n'étaient statistiquement pas significatifs, et deux parmi ces dix étaient insignificativement positifs. Le groupe de médecins spécialisés en gynécologie évoqué plus haut est le seul pour lequel un rapport positif significatif a été observé.

6.2.3 Indicateur «hospitalisation pendant l'année précédente»

Le Tableau 9 représente les coefficients de l'indicateur des hospitalisations pendant l'année précédente pour des groupes de médecins spécialisés choisis. L'indicateur est positivement significatif pour les groupes de médecins spécialisés en médecine interne générale, chirurgie, médecine pour enfants et adolescents et psychiatrie et psychothérapie. Les plus grands effets, et donc les plus pertinents sur le plan économique, sont ceux de la médecine interne générale, de la médecine pour enfants et adolescents et de la psychiatrie et psychothérapie. Concernant les autres groupes de médecins spécialisés, l'influence n'est pas significative, mais positiv, avec l'exception de l'ophtalmologie.

Tableau 9 Influence des variables hospitalisation pendant l'année précédente, modèle global par groupes de médecins spécialisés

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants/ adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	205 430	15 543	23 401	14 295	17 115	34 238	59 562
N médecins	5455	481	1221	425	020	871	2487
Coefficient modèle log/	0,111*** (0,00)	0,059*** (0,00)	0,024 (0,27)	0,003 (0,80)	0,099*** (0,00)	-0,032 (0,21)	0,093*** (0,00)

Coefficients du modèle global avec GAS, franchise, hospitalisation pendant l'année précédente et PCG, variable cible logarithmisée. Erreur standard entre parenthèses, niveau de signification: *** p<0.01, ** p<0.05, * p<0.1.

Le tableau représente les coefficients de l'indicateur «hospitalisation pendant l'année précédente» pour des groupes de médecins spécialisés choisis. Dans quatre groupes de médecins spécialisés, l'indicateur est significativement positif, dans les autres groupes, il n'a pas d'influence significative.

Source: données de Sasis SA, année 2015, calculs internes.

Tous groupes de médecins spécialisés confondus, l'indicateur «hospitalisation pendant l'année précédente» est significativement positif pour 13 groupes de médecins spécialisés. Dans les autres groupes, il est insignifiant. Un effet négatif significatif, qui irait à l'encontre des intuitions, n'a été observé dans aucun groupe de médecins spécialisés.

6.2.4 Indicateur PCG

Nous avons déjà expliqué comment étaient constitués les groupes de coûts pharmaceutiques dans la section 5.2.2. Nous avons utilisé une liste PCG qui a été développée pour la compensation des risques dans l'assurance-maladie PCG. Les PCG possibles sont répertoriés dans le Tableau 10.

Tableau 10 PCG, dont l'intégration a été testée

PCG	Label	PCG	Label
1	Asthme	13	Tumeurs hormono-dépendantes
2	COPD/asthme sévère	14	Cancer
3	Fibrose cystique/enzymes pancréatiques	15	Maladies rénales
4	Cholestérol élevé	16	Maladies du cerveau/moelle osseuse
5	Maladie de Crohn et colite ulcéreuse	17	Douleurs neuropathiques
6	Dépression	18	Parkinson
7	Diabète type I	19	Psychose, Alzheimer et addictions
8	Diabète type II	20	Rhumatismes
9	Épilepsie	21	Maladies de la thyroïde
10	Glaucome	22	Transplantations
11	Maladies cardiaques	23	Hypertension
12	HIV/SIDA	24	ADHS

Pour l'opérationnalisation des PCG, nous avons testé une classification qui a été développée avec la compensation des risques entre les assureurs maladie. Elle contient 24 groupes de coûts pharmaceutiques.

Source: Trottmann et al. (2015), Tableau 4; modification pour PCG 8 et 23 (voir section 5.2.2).

Comme décrit dans la section 5.2.2, un PCG n'est pris en compte que si plus de 30 médecins du groupe de médecins spécialisés ont prescrit une quantité minimale de médicaments du PCG concerné. Cette quantité minimale correspond à 1,8 DDD par cabinet. Il s'agit d'une grandeur purement statistique, nous avons renoncé à une évaluation des PCG qui doivent être intégrés dans le modèle pour chaque groupe de médecins spécialisés.

Le Tableau 11 montre les PCG qui ont été intégrés dans le modèle pour les groupes de médecins spécialisés concernés. Pour les médecins en médecine interne générale, il s'agit des 24 PCG. Étant donné qu'il s'agit d'un grand groupe de médecins spécialisés, le seuil des 30 médecins prescripteurs est vite atteint. La même règle s'applique aux médecins en psychiatrie et psychothérapie, pour lesquels 13 des 24 PCG ont été intégrés. Les groupes de médecins plus spécialisés comme par exemple en ophtalmologie ne prescrivent en revanche que des médicaments de quelques PCG (par exemple PCG 10 Glaucome). Le PCG 23 (hypertension) qui est très prévalent auprès de la population constitue le PCG le plus souvent intégré dans le modèle.

Tableau 11 Influence des PCG, modèle global

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants / adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	205 430	15 543	23 401	14 295	17 115	34 238	59 562
N médecins	5455	481	1221	425	020	871	2487
PCG pris en compte	Tous 24	4, 6, 23	6, 13, 14, 21, 23	1, 4, 6, 8, 11, 21, 23	1, 24	10	1, 4, 6, 8, 9, 16, 17, 18, 19, 20, 21, 23, 24

Les PCG ne sont intégrés dans le modèle que si dans un groupe de médecins spécialisés, au moins 30 médecins ont prescrit un minimum de médicaments du domaine d'indication (critère purement statistique, pas d'évaluation du contenu). Pour le groupe des médecins spécialisés en médecine interne générale, cela s'applique à presque tous les PCG alors que pour les autres groupes, comme l'ophtalmologie, seuls quelques PCG sont concernés (PCG 10, glaucome).

Source: données de Sasis SA, année 2015, calculs internes.

Les coefficients des PCG sont illustrés dans la feuille Excel en annexe «Polynomics_Wirtschaftlichkeitspruefungen_Regressionskoeffizienten_Stufe_1». Dans aucun des 30 groupes de médecins spécialisés, il n'y a eu de PCG qui ont obtenu des coefficients significativement négatifs dans les statistiques. Certains coefficients étaient en revanche insignificativement négatifs, en particulier dans les groupes avec des quantités faibles de substances actives. Il serait possible de les regrouper avec le groupe «pas de prescription». L'indicateur PCG ne serait alors pris en compte qu'à partir d'une certaine quantité de substance active. Cette démarche correspond par exemple aussi au modèle HCC, qui est utilisé dans le programme Medicare états-unien pour l'ajustement des risques (Pope et al. 2000). En alternative, il serait possible de spécifier la quantité de substance active comme variable continue (cf. section (Pope et al. 2000)) pour entièrement éviter les coefficients négatifs.

6.3 Aperçu du premier niveau

Le Tableau 12 illustre les chiffres clés de la valeur explicative dans le modèle avec les franchises élevées, l'hospitalisation pendant l'année précédente et les PCG. À titre de comparaison, le tableau contient aussi les chiffres clés pour une régression uniquement avec l'effet spécifique au

cabinet (EC) et avec un modèle composé de l'EC et du GAS. Le modèle avec uniquement l'effet spécifique au cabinet correspond dans ce cas au comparatif des valeurs moyennes pondérées, comme on le pratique aussi dans l'indice RSS.

En plus du coefficient R^2 et de l'erreur de prédiction absolue moyenne (MAPE), on y voit aussi les deux critères d'information AIC (Akaike Information Criterion) et BIC (Bayesian Information Criterion). Pour ces deux critères d'information, l'idéal est d'avoir une valeur basse (voire même négative). Contrairement au coefficient R^2 par exemple, ces indicateurs évaluent la valeur explicative supplémentaire de facteurs d'influence. Ils se dégradent lorsque des facteurs supplémentaires qui contribuent peu à la valeur explicative sont intégrés.

Concernant tous les indicateurs, le modèle avec les informations de morbidité s'en sort mieux que les modèles avec seulement un effet de cabinet ou avec le GAS et un effet de cabinet. Chez les médecins dont la patientèle diffère de la moyenne pour ce qui est des variables explicatives, la prise en compte des variables de morbidité conduit à une amélioration nette de l'évaluation.

Les différences entre groupes de médecins spécialisés sont importantes. Ainsi, l'amélioration du coefficient R^2 est nettement meilleure dans les groupes de médecins spécialisés en médecine interne générale, en médecine pour enfants et adolescents et en psychiatrie et psychothérapie que dans les autres groupes. Le MAPE, qui ne réagit pas aussi fortement aux aberrations que le coefficient R^2, s'améliore nettement dans les groupes de médecins spécialisés en médecine interne générale, en gynécologie, en médecine pour enfants et adolescents et en ophtalmologie. Vous trouverez d'autres débats sur le diagnostic de régression dans la section 11.3 en annexe.

Tableau 12 Valeur explicative du modèle avec franchises élevées, hospitalisation pendant l'année précédente et PCG

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants / adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	205 430	15 543	23 401	14 295	17 115	34 238	59 562
N médecins	5455	481	1221	425	020	871	2487
Adj. R^2							
Modèle complet	0,877	0,742	0,899	0,766	0,768	0,865	0,526
EC + GAS	0,817	0,729	0,892	0,700	0,745	0,862	0,427
EC	0,284	0,704	0,633	0,576	0,339	0,580	0,333
MAPE							
Modèle complet	0,279	0,341	0,286	0,223	0,490	0,252	0,438
EC + GAS	0,330	0,347	0,292	0,246	0,508	0,259	0,478
EC	0,627	0,360	0,444	0,285	0,710	0,417	0,515
AIC							
Modèle complet	158,8	6771,4	-12 990,0	-9494,2	-5098,7	-1386,4	54 506,4
EC + GAS	81 834,9	7527,6	-11 438,7	-5934,3	-3508,4	-644,7	65 744,6
EC	362 275,1	8877,1	17 216,2	-1040,0	12 778,3	37 378,5	74 740,8
BIC							
Modèle complet	3747,7	7146,3	-12 482,2	-8956,9	-4703,6	-939,0	55 468,9
EC + GAS	82 254,5	7841,3	-11 108,3	-5624,0	-3190,8	-298,7	66 113,4
EC	362 275,1	8877,1	17 216,2	-1040,0	12 778,3	37 378,5	74 740,8

Variable cible logarithmisée. N: taille de l'échantillon; adj. R^2: coefficient de détermination corrigé; MAPE: erreur de prédiction absolue moyenne; AIC: critère d'information d'Akaike; BIC: critère d'information bayésien; EC: effet de cabinet; GAS: groupes d'âge et de sexe.

Les chiffres en noir indiquent la valeur explicative du modèle avec les niveaux de franchise, l'hospitalisation pendant l'année précédente et les PCG, les chiffres en gris celle des modèles avec seulement l'effet de cabinet (EC) et/ou l'EC et le GAS. Pour tous les indicateurs, le modèle avec les informations de morbidité est celui qui s'en sort le mieux. Chez les médecins dont la patientèle diffère de la moyenne pour ce qui est des variables explicatives, la prise en compte des variables de morbidité devrait conduire à une amélioration nette de l'évaluation.

Source: données de Sasis SA, année 2015, calculs internes.

7 Effet spécifique au cabinet et calcul d'indice

7.1 Distribution de l'estimation ponctuelle pour l'effet spécifique au cabinet

Sur le principe, l'effet spécifique au cabinet est calculé comme pour les autres coefficients. Mais l'effet n'est pas considéré par rapport à une catégorie de référence, mais par rapport à la moyenne du groupe de médecins spécialisés. Ceci permet de s'assurer que l'effet de cabinet moyen soit égal à zéro pour tous les groupes de médecins spécialisés (voir colonne «Valeur moyenne» dans le Tableau 13).

La distribution des effets de cabinet est représentée dans le Tableau 13.[8] Dans le modèle logarithmique, les valeurs multipliées par 100 correspondent à peu près à l'effet de cabinet en %. Un effet de cabinet de 0,16 indique donc que le cabinet a des coûts 16 % supérieurs à ceux attendus en fonction de la structure de sa patientèle. Les chiffres dans le modèle non transformé correspondent aux déviations par rapport à la moyenne pour chaque cabinet en francs. Cette moyenne n'est pas la même pour tous les groupes de médecins spécialisés, les chiffres ne peuvent donc pas être comparés directement. Cette comparaison est toutefois disponible dans les résultats du calcul d'indice dans le chapitre 7.3.

On note surtout la forte augmentation entre le 95^{e} centile, le 99^{e} centile et le maximum. Il existe un petit groupe de cabinets dont les coûts inexpliqués par patient sont largement supérieurs à ceux des autres cabinets. Dans le modèle non transformé, cette différence est particulièrement manifeste, ce qui doit très certainement être considéré comme l'inconvénient de ce modèle. Dans la plupart de ces cas, il faut s'attendre à ce que ces valeurs très élevées soient explicables par un éventail de prestations spécial ou par une autre particularité du cabinet. Dans un sous-projet dédié, l'objectif sera de vérifier, dans la continuité de cette étude, de quelles particularités il s'agit et si elles peuvent être intégrées dans la surveillance statistique.

Tableau 13 Distribution de l'effet spécifique au cabinet

	N	Valeur moyenne	Dév. std.	p25	p50	p75	p95	P99	Asymétrie	max.
Logarithmisé	17 464	0,0	0,36	-0,16	0,00	0,17	0,49	0,96	5,83	-0,53
Non transformé	17 464	0,0	476	-152	-38	83	432	1290	162 239	64
Winsorisé	17 464	0,0	294	-125	-22	95	387	830	23 018	10

Pondéré par nombre de patients par cabinet. Les cabinets avec des groupes de médecins spécialisés de moins de 30 cabinets et les cabinets de groupe ne sont pas représentés.

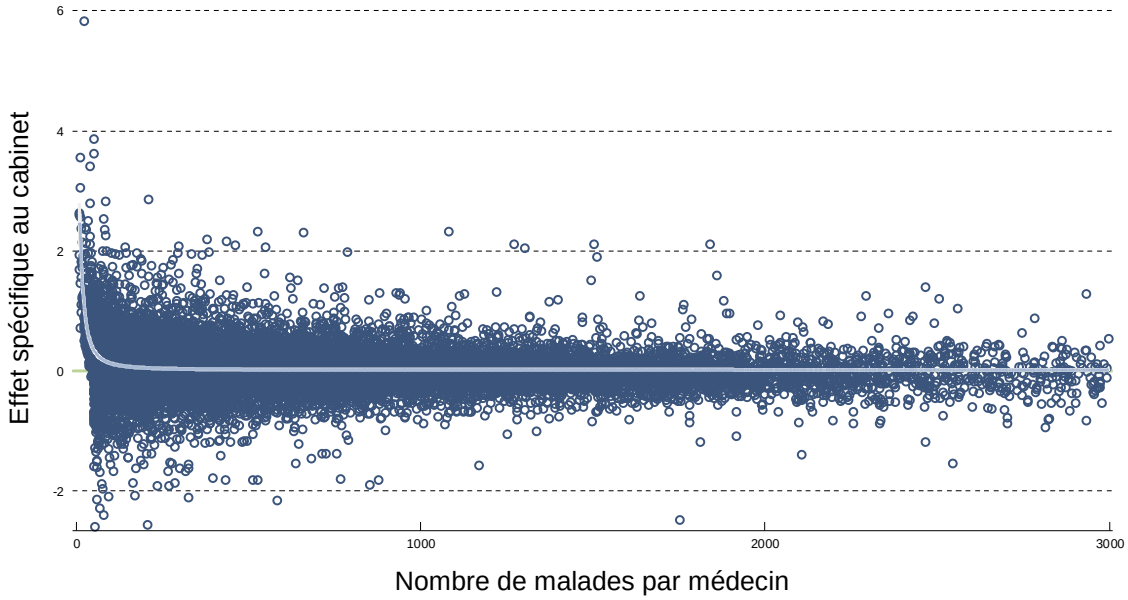
Les effets de cabinet indiquent la différence moyenne d'un cabinet médical par rapport à la moyenne totale du groupe de médecins spécialisés concerné. La valeur moyenne est donc toujours égale à zéro. Les valeurs dans le modèle logarithmisé montrent (multipliées par 100) les déviations en %, alors que les deux autres modèles les indiquent en francs.

Source: données de Sasis SA, année 2015, calculs internes.

[8] Le tableau ne comprend pas les résultats des cabinets de groupe et des groupes de médecins spécialisés de moins de 30 cabinets. Il s'agit en effet de groupes très hétérogènes qui nécessitent beaucoup de prudence pour l'interprétation des effets de cabinet.

La figure 6 représente le rapport entre le nombre de malades par médecin et l'effet spécifique au cabinet calculé. La «forme en entonnoir» du nuage de points indique que dans les très petits cabinets (jusqu'à près de 200 malades), la dispersion de l'effet spécifique au cabinet est supérieure à celle des grands cabinets. La ligne de tendance bleu clair passe à plat sur de grandes parties de la distribution. Il n'y a donc pratiquement pas de rapport systématique entre la taille du cabine et l'effet spécifique au cabinet. Dans les plus petits cabinets, la ligne de tendance tend subitement vers le haut. Dans les cabinets jusqu'à 50 malades environ, l'effet de cabinet calculé est donc aussi systématiquement supérieur. Ceci est principalement dû à un effet de sélection: les très petits cabinets ne sont en effet intégrés dans la surveillance statistique que s'ils ont causé plus de 100 000 francs de coûts (voir section 11.2 en annexe).

Figure 6 Effet spécifique au cabinet et nombre de malades



Chaque point correspond à un cabinet médical, N = 17 078 cabinets médicaux avec plus de 3000 patients (N = 379) non représentés.

La ligne de tendance bleu clair est très plate. Sur de larges parties de la distribution, il n'y a donc pas de rapport entre la taille du cabinet et l'effet de cabinet calculé. Pour les petits cabinets (de moins de 200 malades), les points sont un peu plus dispersés. On remarque que pour les plus petits cabinets (moins de 50 malades environ), la ligne de tendance pointe vers le haut. Ceci est dû à la sélection aléatoire, car les très petits cabinets ne sont inclus que s'ils ont causé plus de CHF 100 000 de coûts bruts.

Source: données de Sasis SA, année 2015, calculs internes.

7.2 Correction et caractéristiques de l'emplacement du cabinet

Bien que les données aient déjà été corrigées des valeurs de point tarifaire différentes, on peut s'imaginer que le traitement de collectifs de patients similaires engendre des coûts différents en fonction de la région. Ceci pourrait p. ex. être dû à des influences culturelles ou à l'influence des principaux centres de formation médicale régionaux. D'autres facteurs comme le degré d'urbanisation ou le taux d'aide sociale pourraient aussi jouer un rôle dans les coûts par patient.

Comme décrit dans la section 4.1.2, nous corrigeons les effets spécifiques au cabinet dans une deuxième régression de l'influence de l'emplacement du cabinet (voir aussi Kaiser, 2016). Dans

cette deuxième régression, la variable cible est l'effet spécifique au cabinet (une valeur par cabinet). Les coefficients de régression sont estimés comme étant la moyenne de tous les groupes de médecins spécialisés, car beaucoup d'entre eux ont trop peu de cabinets pour qu'un calcul par groupe de médecins spécialisés soit judicieux. À ce niveau aussi, nous réalisons une pondération par le nombre de malades par médecin.

7.2.1 Canton du cabinet

L'influence du canton dans lequel se trouve le cabinet est représentée dans la partie gauche du Tableau 14.[9] Sur les 25 cantons, 17 ne sont pas significativement différents de la catégorie de référence (Argovie). Les coefficients significatifs sont indiqués en jaune foncé (coefficient positif) ou en bleu (coefficient négatif). La valeur fortement négative du canton de l'Uri est marquante. La valeur extrême provient très probablement aussi du petit nombre de cabinets qui s'y sont installés. Pour que le calcul des coefficients soit stable, nous considérons qu'un groupe doit contenir au moins 30 cabinets. Les cantons avec peu de cabinets pourraient être regroupés avec des cantons voisins structurellement comparables dans l'analyse.

Comme beaucoup de coefficients de canton sont insignifiants, on peut se demander si le modèle ne pourrait pas être évalué avec une entité géographique plus grande. Le côté droit du Tableau 14 illustre les coefficients des sept grandes régions de l'Office fédérale de la statistique. Dans ce tableau, la région du lac de Genève et celle de Zurich sont significativement positives, ce qui les démarque de la catégorie de référence (Espace Plateau). Le montant des coefficients ne peut pas être comparé directement à celui des coefficients pour les cantons parce que le groupe de référence n'est pas le même.

Dans les deux modèles, le coefficient R^2 est très petit, les variables géographiques n'expliquent donc pas vraiment la variance des effets spécifiques au cabinet. La comparaison des critères d'information AIC et BIC est quant à elle intéressante. Les deux critères d'information «punissent» l'intégration de variables supplémentaires avec peu de valeur explicative dans le modèle. L'AIC est légèrement meilleur (parce qu'inférieur) dans le modèle avec les grandes régions et le BIC dans celui avec les cantons. Mais pour les deux indicateurs, les deux modèles sont très proches. On ne peut donc pas en conclure qu'un modèle soit meilleur que l'autre.

Comme les indicateurs de canton n'apportent pas grand chose à la valeur explicative du modèle, il serait tout à fait justifié, sur le plan statistique, de remplacer les cantons dans le modèle par les grandes régions. L'avantage réside dans le fait que l'on estime alors des coefficients plus stables et que les aberrations du canton d'Uri seraient éliminées. Comme les petits cantons peuvent produire des constellations spéciales, il est probablement plus juste de comparer les cabinets au niveau des grandes régions. Mais cela signifierait aussi tourner le dos à la comparaison actuelle au sein du canton. La question doit faire l'objet d'autres débats sur le fond.

[9] Par rapport aux évaluations du premier niveau, la base de données est quelque peu limitée parce que nous ne disposions pas des informations géographiques de tous les médecins. Concernant les évaluations par canton réalisées en interne chez Sasis pour des raisons liées à la protection des données, nous avons également fait le choix de n'intégrer que les cabinets de plus de 50 malades et avec plus de CHF 100 000 de prestations. Pour garantir la comparabilité, toutes les évaluations sur les variables géographiques ont été réalisées sur cette base de données limitée.

Tableau 14 Coefficients du canton ou de la grande région au deuxième niveau

Canton	Coefficient	Grande région	Coefficient
AG	Référence	Espace Plateau	Référence
AI	-0,027	Région du lac de Genève	0,051***
AR	-0,002	Suisse du Nord-ouest	-0,010
BE	-0,005	Suisse orientale	-0,018
BL	-0,056**	Suisse centrale	-0,014
BS	-0,017	Zurich	0,019*
FR	-0,039*		
GE	0,098***		
GL	-0,037		
GR	-0,024		
JU	-0,087*		
LU	-0,051**		
NE	-0,018		
NW	-0,083		
OW	-0,042		
SG	-0,057***		
SH	0,026		
SO	0,016		
SZ	0,024		
TG	0,001		
TI	0,023		
UR	-0,163**		
VD	0,018		
VS	0,003		
ZG	0,030		
ZH	0,010		
N	15 680		15 680
R^2	0,013		0,008
AIC	12 243,3		12 183,8
BIC	12 542,1		12 635,8

Variable cible logarithmisée, autres variables explicatives: groupe de médecins spécialisés, densité démographique de la commune; niveau de signification: *** p<0.01, ** p<0.05, * p<0.1.

17 des 25 indicateurs de canton ne sont pas significativement différents de zéro. Pour le canton de l'Uri, on obtient un coefficient fortement négatif qui est probablement dû au petit nombre de cabinets. Les critères d'information indiquent que le modèle avec seulement les grandes régions n'est pas statistiquement plus mauvais que le modèle avec les cantons.

Source: données de Sasis SA, année 2015, calculs internes.

7.2.2 Autres caractéristiques de l'emplacement du cabinet

On peut penser que d'autres caractéristiques de l'emplacement du cabinet influencent l'effet de cabinet. Nous avions notamment émis l'hypothèse que la densité démographique, le taux d'aide sociale ou encore la part d'étrangers pourraient jouer un rôle. L'Office fédéral de la statistique[10] met ces données à dispositions pour chaque commune. Nous avons donc exploité ces données et formé 5 à 6 groupes en fonction de la distribution empirique.

Comme illustré dans le Tableau 15, ces indicateurs ont été insignifiants dans toutes les estimations. Les coefficients des groupes les plus élevés (plus forte densité démographique, plus fort taux d'aide sociale, plus grande part d'étrangers) présentent les signes positifs attendus, mais en termes de contenu, leur taille est limitée (les détails sont visibles dans la section 11.4 en annexe). C'est le taux d'aide sociale qui a la plus grande influence: Les cabinets situés dans une zone avec un fort taux de bénéficiaires de l'aide sociale ont en moyenne un effet spécifique au cabinet 4 % supérieur à celui des autres cabinets. Comme les coefficients sont toutefois insignifiants et faibles, on peut les ignorer dans le modèle pour des raisons statistiques. Nous présentons les conséquences d'une renonciation au calcul de l'indice dans la section 7.4.

[10] Portraits régionaux 2015: Communes – chiffres clés; consulté sur https://www.bfs.admin.ch/bfs/de/home/statistiken/regionalstatistik/regionale-portraets-kennzahlen/gemeinden.gnpdetail.2016-0166.html.

Tableau 15 Vue d'ensemble de l'intégration des caractéristiques régionales au deuxième niveau

	Spéc. 1 Canton	Spéc. 1 Grande région	Spéc. 2 Canton	Spéc. 2 Grande région	Spéc. 3 Canton	Spéc. 3 Grande région
Région	Réf.: AG ▪ Pos. sign. GE ▪ Neg. sign. BL, BS, JU, LU, NE, SG, UR	Réf.: espace Plateau ▪ Pos. sign: Lac de Genève, Zurich	Réf.: AG ▪ Pos. sign. GE ▪ Neg. sign. BL, FR, JU, LU, SG, UR,	Réf.: espace Plateau ▪ Pos. sign: Lac de Genève, Zurich	Réf.: AG ▪ Pos. sign. GE ▪ Neg. sign. BL, FR, JU, LU, SG, UR	Réf.: espace Plateau ▪ Pos. sign. Lac de Genève, Zurich ▪ Neg. sign. Suisse orientale
Densité démographique	5 groupes tous insign.	5 groupes tous insign.	5 groupes tous insign.	5 groupes tous insign.		
Taux d'aide sociale	6 groupes tous insign.	6 groupes tous insign.				
Part d'étrangers					5 groupes tous insign.	5 groupes tous insign.
N	15 680	15 680	15 680	15 680	15 678	15 678
Adj. R^2	0,014	0,009	0,013	0,008	0,011	0,005
AIC	12 167,3	12 217,3	12 183,8	12 243,3	12 201,9	12 282,5
BIC	12 657,5	12 554,3	12 635,8	12 542,1	12 653,9	12 581,2

Variable cible logarithmisée, autres variables explicatives: groupe de médecins spécialisés.
Les variables calculées densité démographique, taux d'aide sociale et part d'étrangers n'ont pas eu d'influence significative sur l'effet de cabinet. La teneur explicative aussi n'a que peu varié.

Source: données de Sasis SA, année 2015, calculs internes.

Une possible raison pour le fait que les indicateurs de commune ne sont pas significatifs statistiquement est qu'il ne sont pas assez précis. La densité démographique peut ainsi par exemple aussi être élevée dans des communes rurales, dont la superficie n'est pas grande. Le taux d'aide sociale et la part d'étrangers en revanche sont souvent très différents dans les communes, et en particulier dans les métropoles. La moyenne par commune ne donne donc pas d'informations précises sur la patientèle du cabinet.

La classification à neuf niveaux de l'Office fédéral de la statistique représente une autre répartition des communes suisses. Pour classer les communes, elle tient compte du caractère central d'une commune et l'activité principale des habitants. Les coefficients des neufs types de commune sont représentés dans le Tableau 16. Les communes suburbaines et rurales ont des coûts inférieurs à ceux des communes centrales, de 5 à 8 %. Les communes avec les plus forts revenus présentent les coûts les plus élevés par rapport aux communes centrales.

Tableau 16 Intégration des types de communes de l'OFS au deuxième niveau (modèle log.)

	Spécification 4	Spécification 5
Spécification	6 grandes régions; types de commune OFS;	26 cantons; types de commune OFS;
Région	Groupe de comparaison: espace Plateau ▪ Pos. sign.: Lac de Genève, Zurich ▪ Neg. sign. Suisse orientale	Groupe de comparaison: AG ▪ Pos. sign. GE ▪ Neg. sign. BL, FR, JU, LU, NW, SG, UR
Centres	Groupe de référence	Groupe de référence
Communes suburbaines	-0,05*** (-6,59)	-0,05*** (-6,91)
Communes à fort revenu	0,09*** (5,80)	0,09*** (5,54)
Communes périurbaines	-0,03 (-1,58)	-0,02 (-1,54)
Communes touristiques	0,01 (0,37)	0,02 (1,02)
Communes industrielles et tertiaires	-0,06*** (-4,80)	-0,05*** (-4,17)
Communes dortoirs urbaines	-0,08** (-3,20)	-0,08** (-3,14)
Communes agricoles mixtes	-0,07**(-3,19)	-0,06** (-2,74)
Communes agricoles	-0,06 (-0,68)	-0,052 (-0,60)
N	15 678	15 678
Adj. R^2	0,011	0,018
AIC	12 184,3	12 098,2
BIC	12 513,7	12 580,8

Erreur standard entre parenthèses, niveau de signification: *** p<0.01, ** p<0.05, * p<0.1.

Le type de commune selon la classification à neuf niveaux de l'Office fédéral de la statistique a une influence sur les effets de cabinet calculés. Ainsi, les effets de cabinet sont en moyenne inférieurs dans les régions rurales et en moyenne supérieurs à ceux des communes centrales dans les communes à fort revenu. On peut toutefois se demander, sur le fond, si ces différences sont justifiées ou si les communes à fort revenu présentent p. ex. un plus grand problème de «moral hazard».

Source: données de Sasis SA, année 2015, calculs internes.

Même si les effets sont significatifs sur le plan statistique, on peut se demander sur le fond si le fait d'en tenir compte apporterait une évaluation plus juste des cabinets médicaux. On peut par exemple tout à fait s'imaginer que dans les communes à fort revenu, où l'offre est importante, le problème de «moral hazard» est plus étendu que dans les régions rurales. Dans ce cas, il serait tout à fait normal que les valeurs d'indice soient en moyenne inférieures dans les régions rurales. Le modèle de calcul ne devrait alors pas modifier ces indices.

7.3 Résultats du calcul d'indice

7.3.1 Calcul d'indice avec l'estimation ponctuelle

Les résultats du calcul d'indice proposé par Kaiser (2016) sont visibles dans le Tableau 17. Par définition, la moyenne pondérée (pondération par nombre de patients) des valeurs d'indice est égale à 100 pour tous les groupes de médecins spécialisés. Comme pour les effets spécifiques au

cabinet, la distribution des valeurs d'indice présente une asymétrie positive, ce qui signifie que la fourchette des valeurs supérieures à 100 est largement supérieure à celle des valeurs inférieures à 100.

On note surtout la forte différence entre le 95e centile, le 99e centile et le maximum dans le modèle logarithmisé et le modèle non transformé. Dans ces deux modèles, le maximum prend des valeurs très élevées, ce qui est très probablement dû à des particularités de cabinet, qui ne peuvent pas être détectées dans la surveillance statistique. Il n'est pas plausible que le style de cabinet seul puisse avoir un effet aussi important. L'effet de la winsorisation se manifeste pour le maximum: comme les aberrations ont été «tronquées» dans le calcul de l'effet spécifique au cabinet, la valeur d'indice maximale calculée est aussi nettement inférieure.

Tableau 17 Résultats du calcul d'indice

	N	Valeur moyenne	Dév. std.	p50	p75	p95	p99	Asymétrie	max.
Logarithmisé	17 464	100	60	95	112	152	237	32 569	243,97
Non transformé	17 464	100	44	95	112	153	246	4790	7,45
Winsorisé	17 464	100	28	98	114	148	183	613	0,65

Pondéré par nombre de patients par médecin.

Par définition, les valeurs d'indice sont de 100 en moyenne pour chaque groupe de médecins spécialisés. Environ 5 % des cabinets ont une valeur d'indice supérieure à 150. On note surtout la grande différence entre le 99e centile et le maximum. Dans le modèle logarithmisé et dans le modèle non transformé, il y a des aberrations avec des valeurs d'indice très élevées. Dans le modèle winsorisé, ce problème est beaucoup moins visible du fait du troncage des valeurs les plus élevées.

Source: données de Sasis SA, année 2015, calculs internes.

Avec le seuil utilisé actuellement pour une valeur d'indice de 130, en moyenne 18 % des cabinets sont identifiés comme ayant un profil de coûts suspect (voir Tableau 18). Les recoupements entre les trois variantes de calcul sont élevés, 15 % des cabinets sont identifiés comme suspects dans les trois variantes de calcul.

Dans certains groupes de médecins spécialisés (p. ex. chirurgie ou ophtalmologie), le modèle winsorisé met plus de cabinets suspects en évidence. Cela peut indiquer que par le troncage des aberrations pour le calcul du modèle, des cabinets sont identifiés parce qu'ils présentent des coûts élevés, mais pas «élevés au point d'être des aberrations» par patient. Sans ce troncage, ces cabinets seraient «cachés» par leurs aberrations.

Dans le groupe de médecins spécialisés en psychiatrie et psychothérapie, il y a plus de cabinets suspects que dans les autres groupes de médecins spécialisés. Comme la valeur moyenne de l'indice est égale à 100 pour tous les groupes de médecins spécialisés, cela est probablement dû à la dispersion. Si la dispersion est surtout présente *entre les cabinets*, il est tout à fait légitime que l'indice le montre. Dans une deuxième étape, il faut clarifier s'il existe des différences légitimes (particularités d'un cabinet) qui expliqueraient cette dispersion. Si la dispersion est en revanche élevée *au sein d'un cabinet*, cela indique plutôt que les patients ont des besoins très hétérogènes.[11] Nous reviendrons sur ce point dans la section 7.3.2.

[11] La littérature aussi aborde le point de la différence de dispersion *entre* les cabinets et de dispersion *au sein d'*un

Tableau 18 Part de cabinets suspects avec une valeur d'indice de 130

	Modèle logarithmisé	Modèle non transformé	Modèle winsorisé	Dans tous les modèles
Tous les groupes de médecins spécialisés	18 %	18 %	17 %	15 %
Médecine interne générale	15 %	14 %	12 %	11 %
Chirurgie	17 %	18 %	20 %	16 %
Gynécologie	17 %	17 %	14 %	12 %
Cardiologie	12 %	13 %	9 %	8 %
Médecine pour enfants et adolescents	14 %	13 %	13 %	11 %
Ophtalmologie	13 %	14 %	18 %	11 %
Psychiatrie et psychothérapie	26 %	27 %	27 %	24 %

La part de cabinets avec des profils de coûts suspects est similaire dans les trois modèles. Le modèle qui identifie le plus grand nombre de cabinets est différent par groupe de médecins spécialisés. 15 % des cabinets sont identifiés comme suspects dans les trois modèles.

Source: données de Sasis SA, année 2015, calculs internes.

Le Tableau 19 classe les observations selon le percentile de la distribution des coûts par groupe de médecins spécialisés. Dans le groupe des 90 à 100 %, on trouve donc tous les cabinets dont les coûts font partie des 10 % supérieurs de leur groupe de médecins spécialisés. Dans un modèle sans correction de la morbidité (p. ex. avec une simple comparaison des moyennes), ces cabinets sont presque tous identifiés comme suspects. Après la correction de la morbidité à l'aide du modèle logarithmisé, le nombre de cabinets les plus chers n'est plus de 83 %, alors que le nombre de cabinets avec des coûts moyens identifiés comme suspects est plus élevé. Dans le modèle winsorisé, cet effet est encore plus marqué. Parmi les cabinets avec les coûts moyens les plus élevés, seuls deux tiers sont encore identifiés comme suspects.

cabinet. On considère alors une évaluation des cabinets médicaux comme particulièrement fiable lorsque la dispersion entre les cabinets est élevée comparée à celle au sein des cabinets (voir Adams et al., 2010b).

Tableau 19 Part de cabinets suspects selon la distribution des coûts

Percentile de distribution des coûts [1)]	Sans correction de morbidité	Modèle logarithmisé	Modèle non transformé	Modèle winsorisé
0 – 10 %	0 %	0 %	0 %	0 %
10 – 25 %	0 %	2 %	2 %	2 %
25 – 50 %	0 %	4 %	3 %	3 %
50 – 75 %	4 %	11 %	11 %	12 %
75 – 90 %	55 %	38 %	39 %	41 %
90 – 100 %	98 %	83 %	83 %	69 %

[1)] La classification se base sur la distribution des coûts par groupe de médecins spécialisés.

Sans la correction de la morbidité (p. ex. par une simple comparaison de la moyenne), presque tous les cabinets avec les coûts les plus élevés sont identifiés comme suspects. Après la correction de la morbidité, il y a plus de cabinets dans les 75 % inférieurs de la distribution des coûts qui sont identifiés parce que leur profil de coûts est suspect par rapport aux indicateurs de morbidité mesurés.

Source: données de Sasis SA, année 2015, calculs internes.

Le Tableau 20 donne une vue d'ensemble des différences entre le modèle winsorisé et le modèle logarithmisé. 95 % de tous les cabinets ont une évaluation identique dans les deux modèles. Avec 434 cabinets, le groupe qui ne ressort comme suspect que dans le modèle logarithmisé est légèrement plus grand que celui qui n'est considéré comme suspect que dans le modèle winsorisé.

Tableau 20 Comparatif des suspicions dans le modèle logarithmisé et le modèle winsorisé

		Suspect dans le modèle winsorisé		Total
		Non	Oui	
Suspect dans le modèle logarithmisé	Non	14 006 (80 %)	329 (2 %)	14 335
	Oui	434 (3 %)	2695 (15 %)	3129
	Total	14 440	3024	

95 % des cabinets ont la même évaluation dans les modèles logarithmisé et winsorisé. Les cabinets qui ne ressortent comme suspects que dans le modèle logarithmisé sont légèrement plus grands que ceux qui ne sont considérés comme suspects que dans le modèle winsorisé.

Source: données de Sasis SA, année 2015, calculs internes.

Le bloc de données contient beaucoup de cabinets avec un petit nombre de malades (p. ex. jusqu'à 80 malades). On peut s'attendre à ce que ces cabinets présentent une dispersion supérieure des valeurs d'indice à celle des grands cabinets. Le Tableau 21 représente les valeurs d'indice regroupées par taille de cabinet. Pour les regrouper, nous utilisons les quartiles de la distribution de taille par groupe de médecins spécialisés. Les 25 % de cabinets les plus petits par groupe de médecins spécialisés se trouvent donc dans le premier quartile.

On remarque que la valeur d'indice moyenne des petits cabinets est supérieure à celle des grands cabinets. À partir du 2^{e} quartile de la distribution de taille, les valeurs d'indice sont proches. Il existe plusieurs explications possibles aux valeurs d'indice supérieures pour les petits cabinets.

Premièrement, il y a un effet de sélection dans les petits cabinets: les cabinets de moins de 50 patients ne sont en effet intégrés dans la surveillance que s'ils ont causé plus de CHF 100 000 de prestations brutes (c.-à-d. par patient plus de CHF 2000 par an). Dans la plupart des groupes de médecins spécialisés, cette valeur est nettement supérieure au 95^{e} centile de la distribution (voir Tableau 6). Deuxièmement, les cabinets avec peu de patients pourraient être tentés de fournir plus de prestations à certains patients alors que les cabinets bien fréquentés ne le font pas. Dans ce cas, cette valeur d'indice supérieure serait justifiée. Troisièmement, les aberrations élevées pourraient avoir une influence nettement supérieure sur les coûts totaux pour des *patients individuels* dans les petits cabinets. La valeur d'indice ne doit pas illustrer cet état de fait, car il ne s'agit pas d'un effet de cabinet. Nous reviendrons sur ce point dans la section suivante et lors des simulations à la fin de la section 8.3.

Tableau 21 Valeurs d'indice par taille de cabinet

Taille du cabinet	N	Valeur moyenne	Dév. std.	p50	p75	p95	p99
Modèle logarithmisé							
1er quartile par groupe de médecins spécialisés	4337	121	172	107	137	219	417
2^{e} quartile par groupe de médecins spécialisés	4377	105	47	99	118	168	251
3^{e} quartile par groupe de médecins spécialisés	4371	99	34	96	110	143	217
4^{e} quartile par groupe de médecins spécialisés	4379	96	40	92	108	143	210
Total	**17 464**	**100**	**60**	**95**	**112**	**152**	**237**
Modèle winsorisé							
1er quartile par groupe de médecins spécialisés	4337	111	37	109	131	177	220
2^{e} quartile par groupe de médecins spécialisés	4377	104	28	103	119	153	185
3^{e} quartile par groupe de médecins spécialisés	4371	100	25	99	113	141	179
4^{e} quartile par groupe de médecins spécialisés	4379	97	28	96	111	143	176
Total	**17 464**	**100**	**28**	**98**	**114**	**148**	**183**

Pondéré par nombre de patients par médecin.

Les cabinets avec peu de malades ont en moyenne une valeur d'indice supérieure à ceux avec beaucoup de malades. Dans le modèle winsorisé, cet effet n'est pas aussi marquant.

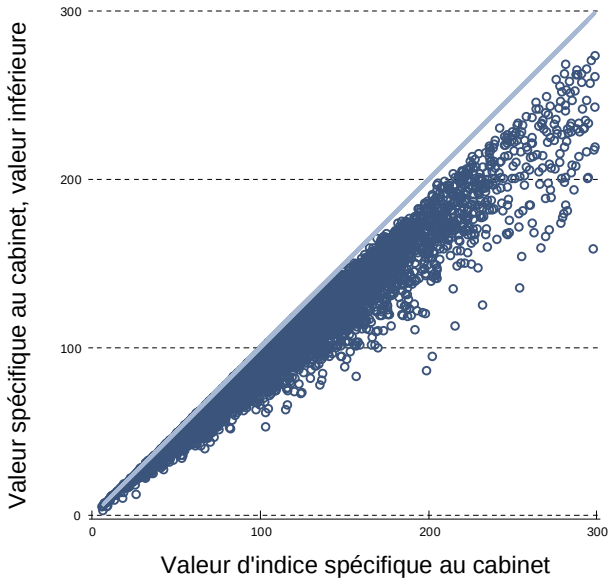
Source: données de Sasis SA, année 2015, calculs internes.

7.3.2 Calcul d'indice avec prise en compte de l'indicateur de confiance

La figure 7 représente la valeur d'indice spécifique au cabinet et la limite inférieure (selon le calcul dans la section 4.4). Les cabinets proches de la ligne du milieu ont une dispersion faible. Si leurs coûts sont différents de la valeur moyenne du groupe de médecins spécialisés (ce qui s'exprime par une valeur d'indice différente de 100), ils le font environ à la même hauteur pour tous les GAS. Les cabinets avec une forte dispersion sont sous la ligne des 45 degrés dans la figure 7.

L’incertitude vient du fait qu'il y a de fortes différences entre les observations (GAS différents). On peut voir que les cabinets avec une valeur d'indice élevée ont tendance à aussi avoir des indicateurs d’incertitude supérieurs. Ils subissent donc une forte variation lorsque l’indice est calculé avec la limite inférieure.

Figure 7 Limite inférieure pour la valeur d’indice



Chaque point correspond à un cabinet médical, N = 17 078 cabinets médicaux avec plus de 3000 patients (N = 386) non représentés.

Cette figure montre le rapport entre la valeur d’indice calculée avec la limite inférieure (axe vertical) et celle calculée avec l’estimation ponctuelle (axe horizontal). En moyenne, les cabinets avec une valeur d'indice élevée ont aussi un indicateur d’incertitude plus élevé que les cabinets avec une valeur faible.

Source: données de Sasis SA, année 2015, calculs internes.

La dernière colonne du Tableau 22 indique la part de cabinets pour laquelle la limite inférieure de la valeur d'indice est supérieure à 130. Par rapport à l’estimation ponctuelle, la part de médecins suspects est nettement réduite. Tous cabinets confondus, le nombre passe ainsi de 18 à 11 %. Le plus marquant est la baisse pour le groupe de médecins spécialisés en psychiatrie et psychothérapie, qui avait d’ailleurs une part importante de cabinets suspects dans l’estimation ponctuelle. Cela peut indiquer qu’une partie de la valeur élevée est due au fait que la grande dispersion *au sein* des cabinets n’a pas permis d’estimer clairement la valeur spécifique au cabinet. Lors du calcul de l’indice avec la limite inférieure, ce groupe de médecins spécialisés ne présente donc plus systématiquement plus de cabinets suspects que les autres groupes.

Tableau 22 Part de cabinets suspects en cas d'utilisation de la limite inférieure pour l'évaluation

	Modèle logarithmisé	Modèle logarithmisé, calcul d'indice avec limite inférieure
Tous les groupes de médecins spécialisés	18 %	11 %
Médecine interne générale	15 %	9 %
Chirurgie	17 %	13 %
Gynécologie	17 %	11 %
Cardiologie	12 %	8 %
Médecine pour enfants et adolescents	14 %	8 %
Ophtalmologie	13 %	11 %
Psychiatrie et psychothérapie	26 %	12 %

La dernière colonne indique pour quelle part de cabinets la limite inférieure de la valeur d'indice est supérieure à 130. La part de cabinets suspects est nettement réduite. Tous cabinets confondus, le chiffre de cabinets suspects passe de 18 à 11 %. La baisse pour le groupe de médecins spécialisés en psychiatrie et psychothérapie est la plus marquante. Il s'agissait du groupe avec la valeur de départ la plus élevée.

Source: données de Sasis SA, année 2015, calculs internes.

Comme nous l'avons montré dans la section précédente, la valeur d'indice calculée pour les cabinets avec un plus petit nombre de malades est systématiquement plus élevée que pour les grands cabinets. Ceci se répercute aussi sur la part de cabinets classés comme suspects. Le Tableau 23 répartit les cabinets en groupes selon le nombre de malades. On peut constater que dans le groupe avec le moins de malades, le nombre de cabinets suspects est nettement supérieur. Si on calcule l'indice avec l'estimation ponctuelle, on obtient ainsi par exemple 35 % des plus petits cabinets. Le nombre supérieur de suspects persiste pour le calcul avec la limite inférieure. Cela indique que le fait que les petits cabinets soient souvent suspects n'est pas dû à la grande dispersion à cause de la variation aléatoire, mais à des valeurs systématiquement supérieures (par sélection aléatoire ou «moral hazard» supérieur). Lors de la simulation dans la section 8.3, nous reviendrons sur le point de la variation aléatoire (Tableau 29).

Tableau 23 Cabinets identifiés avec l'estimation ponctuelle ou la limite inférieure

Nombre de malades par cabinet	N	Estimation ponctuelle logarithmisée	Limite inférieure logarithmisée
1er quartile par groupe de médecins spécialisés	4337	35 %	22 %
2e quartile par groupe de médecins spécialisés	4377	18 %	11 %
3e quartile par groupe de médecins spécialisés	4371	11 %	6 %
4e quartile par groupe de médecins spécialisés	4379	8 %	5 %

Le nombre de cabinets suspects avec un petit nombre de malades est largement supérieur à celui des grands cabinets. Et cela ne change pas quand on utilise la limite inférieure. Nous reviendrons sur ce point dans la section Simulation.

Source: données de Sasis SA, année 2015, calculs internes.

7.4 Tests de spécification du deuxième niveau

Comme évoqué dans le chapitre 7.2, les caractéristiques de l'emplacement du cabinet n'ont eu que peu d'influence statistiquement significative au deuxième niveau. Dans cette section, nous allons vérifier leur influence sur le calcul d'indice. À cette fin, le Tableau 24 représente la manière dont la part des cabinets classés comme suspects varie en fonction de la spécification du deuxième niveau.[12]

Comme grandeur de référence, nous avons calculé l'indice uniquement après le premier niveau (tous les aspects du calcul d'indice restent inchangés). Ainsi 17,6 % des cabinets seraient identifiés comme suspects. Si l'on tient compte de la grande région, la part baisse à 17,3 %, soit 56 cabinets qui ne sont plus considérés comme suspects. En ajoutant le taux d'aide sociale et le groupe de médecins spécialisés[13], le nombre de cabinets baisse à nouveau de 37, avec un taux de 17,0 % de cabinets suspects. Si on utilise la densité démographique à la place du taux d'aide sociale, la part baisse encore plus pour atteindre 16,9 %. Si on utilise les deux variables de taux d'aide sociale et de densité démographique dans le modèle, le taux reste à 16,9 % de cabinets suspects.

Tableau 24 Part de cabinets suspects avec spécification différente du 2e niveau

	N	Nombre de cabinets suspects	Part de cabinets suspects
2e niveau avec groupe de médecins spécialisés, taux d'aide sociale, densité démographique et grande région	15 766	2661	16,9 %
2e niveau avec groupe de médecins spécialisés, densité démographique et grande région	15 766	2660	16,9 %
2e niveau avec groupe de médecins spécialisés, taux d'aide sociale et grande région	15 766	2682	17,0 %
2e niveau avec groupe de médecins spécialisés, grande région	15 766	2719	17,3 %
1er niveau	15 766	2775	17,6 %

Par la correction supplémentaire au deuxième niveau, le nombre de cabinets identifiés comme suspects baisse de 114 (modèle global avec toutes les variables). Si on n'utilisait uniquement les grandes régions, ce chiffre serait de 56 cabinets.

Source: données de Sasis SA, année 2015, calculs internes.

[12] La base de données reste limitée comme décrit dans le chapitre 7.2.

[13] La valeur moyenne de chaque groupe de médecins spécialisés est égale à zéro (voir section 7.1). Le groupe de médecins spécialisés n'est intégré dans le modèle au deuxième niveau que si d'autres variables explicatives sont également prises en compte.

8 Simulation pour la détermination des faux positifs et faux négatifs

8.1 Erreurs de 1er et 2e type

La fiabilité d'un modèle statistique se mesure généralement aux erreurs de prédiction. On différencie en général les erreurs de 1er type des erreurs de 2e type. Une erreur de 1er type (aussi appelée erreur α) est présente lorsque l'hypothèse nulle sur laquelle se base le test d'hypothèse est rejetée bien qu'elle soit vraie dans la réalité. L'autre mauvaise décision possible est l'erreur de 2e type (aussi appelée erreur β). On parle de cette erreur lorsqu'on rejette par erreur l'hypothèse alternative.

Dans le cas d'évaluation d'économicité, l'hypothèse nulle est donnée par le fait qu'un cabinet médical travaille de manière efficiente et ne présente pas de coûts excessifs. L'hypothèse alternative est que le cabinet ne travaille pas de manière efficiente et présente donc des coûts excessifs. On obtient donc quatre options pour le test des hypothèses, qui sont illustrées dans le Tableau 25.

Tableau 25 Tableau de décision pour les tests d'hypothèse statistiques

	Faits réels: hypothèse nulle (le cabinet est efficient)	Faits réels: hypothèse alternative (le cabinet est inefficient)
Résultat statistique au test négatif: non-rejet de l'hypothèse nulle	Décision correcte justement négatif	Erreur 2e type faux négatif
Résultat statistique au test positif: rejet de l'hypothèse nulle	Erreur 1e type faux positif	Décision correcte justement positif

Un test d'hypothèse statistique ne peut avoir que quatre résultats. Soit l'hypothèse nulle n'est pas rejetée, et ce à juste titre, soit elle est rejetée à juste titre en faveur de l'hypothèse alternative. Dans ce cas, on parle de résultats justement négatifs et justement positifs. L'hypothèse nulle peut toutefois aussi être rejetée par erreur ou ne pas être rejetée par erreur. Dans ce cas, on parle d'erreurs de 1er et 2e type ou de résultats faux positifs et faux négatifs.

Source: interne, Polynomics.

Un résultat statistique peut être positif ou négatif. Dans le premier cas, on ne rejette généralement pas l'hypothèse nulle et dans le second cas, on la rejette en faveur de l'hypothèse alternative. Selon que l'hypothèse nulle ou alternative était vraie, on obtient des bonnes décisions ou des résultats faux positifs (erreur de 1er type). Cela signifie donc que des cabinets sont estimés comme inefficients par erreur, alors qu'ils sont efficients. D'un autre côté, on obtient aussi des résultats faux négatifs (erreur de 2e type) en quel cas des cabinets inefficients sont détectés comme efficients.

L'objectif est d'avoir la meilleure fiabilité possible avec une méthode statistique et donc de réduire au minimum les erreurs de 1er et 2e type. En même temps, la probabilité d'obtenir des résultats corrects augmente. Il y a toutefois aussi des limites à la fiabilité parce que le fait de réduire une erreur de 1er type augmente automatiquement la probabilité d'erreurs de 2e type. Les deux types d'erreur ne peuvent donc pas être réduits en même temps. En réalité, il y a un conflit et il faut réfléchir si l'on préfère éviter les résultats faux positifs ou plutôt les résultats faux négatifs.

Dans notre cas, la plus grande difficulté réside dans le fait que nous ne sommes même pas en mesure de déterminer les erreurs α et β parce que nous ne connaissons pas la vérité. Nous n'avons

en effet aucune information sur le fait que les cabinets contenus dans le bloc de données sont réellement efficients ou non. Il n'existe pas non plus de test qui nous permettrait de constater de manière univoque les cabinets efficients ou inefficients (pas de standard de référence). Les modèles statistiques avec les données réelles ne peuvent donc pas être vérifiés quant à la présence de plus ou moins de résultats faux positifs ou faux négatifs. La seule possibilité consiste à utiliser des simulations. Pour y arriver, nous simulons toutes les données du modèle et générons des cabinets qui sont inefficients. Cela permet de comparer la «réalité» avec les résultats statistiques et d'évaluer la fiabilité des modèles.

8.2 Mode opératoire de la simulation

8.2.1 Calcul des coûts attendus

Pour obtenir des coûts (efficients) attendus pour tous les cabinets médicaux, nous estimons le modèle du premier niveau (1) avec les coûts absolus, c'est-à-dire non logarithmisés y_{ij}. Puis, nous calculons pour chaque cabinet les coûts attendus $\hat{y}_{ij}$ ainsi que les valeurs résiduelles $\hat{\varepsilon}_{ij}$ sans l'effet spécifique au cabinet $\hat{a}_i$ des paramètres de modèle obtenus. Les formules sont visibles dans les équations (10) et (11). L'estimation économétrique ainsi que le calcul consécutif des coûts attendus et valeurs résiduelles sont séparés par groupe de médecins spécialisés.

$$\hat{y}_{ij} = GAS\hat{\beta}_1 + X_{ij}\hat{\beta}_2 \tag{10}$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \hat{y}_{ij} - \hat{a}_i \tag{11}$$

pour tous les $i \in GMS_f$ et $f = \{1,2,\dots,F\}$

8.2.2 Simulation inefficience et terme d'erreur

Les coûts attendus $\hat{y}_{ij}$ représentent les coûts sans inefficience et variation aléatoire. Dans la simulation, on génère des valeurs d'inefficience par cabinet ainsi que des valeurs résiduelles par observation en plusieurs opérations. La constitution des coûts simulés $\dot{y}_{ij}$ est représentée dans l'équation (12). Les valeurs d'inefficience I_i proviennent d'une distribution exponentielle, qui est ensuite multipliée pour obtenir les coûts attendus $\hat{y}_{ij}$.

$$\dot{y}_{ij} = \hat{y}_{ij} \times (1 + I_i) + \tilde{\varepsilon}_{ij} \text{ et } I_i \sim Exp(\lambda) \tag{12}$$

Un médecin est considéré comme inefficient si ses coûts sont 30 % supérieurs à la moyenne. Les paramètres sont alors choisis de manière à générer environ 10 % de cabinets inefficients par opération.[14]. Pour générer les valeurs résiduelles $\tilde{\varepsilon}_{ij}$, nous utilisons les véritables valeurs résiduelles $\hat{\varepsilon}_{ij}$ de l'équation (3). Les valeurs résiduelles sont formées par groupe de médecins spécialisés comme les coûts attendus associés. L'objectif est d'obtenir une dispersion la plus réaliste possible

[14] Mathématiquement, nous avons cherché une distribution exponentielle pour laquelle 10 % des valeurs sont supérieures à au moins 30 % au-dessus de la moyenne. Ceci s'applique à la fonction exponentielle avec $\mu = 0.3$. μ correspond alors à l'efficience moyenne par groupe de médecins spécialisés. Dans la simulation, les médecins qui sont 30 % au-dessus de l'efficience moyenne sont considérés comme inefficients. L'équation suivante doit donc s'applique à la valeur limite: $\bar{\dot{y}}_i > 1.3 \times \bar{\hat{y}} \rightarrow \bar{\dot{y}}_i > 1.3 \times 1.3 \times \bar{y} = 1.69 \times \bar{y}$, = $\bar{\hat{y}}$ représentant les coûts moyens avec inefficience et $\bar{y}$ les coûts moyens sans inefficience.

des données. Dans chaque opération de simulation, nous extrayons une valeur résiduelle par observation $\tilde{\varepsilon}_{ij}$ de la distribution de $\hat{\varepsilon}_{ij}$ et l'ajoutons aux coûts attendus $\hat{y}_{ij}$.[15]

8.2.3 Contrôle de la fiabilité

Après la génération des coûts simulés $\dot{y}_{ij}$, nous effectuons l'estimation économétrique du modèle à deux niveaux et formons les valeurs d'indice standardisées pour identifier les cabinets inefficients. Dans chaque étape, le modèle à deux niveaux est estimé trois fois. La structure du modèle est identique, ce qui n'est pas le cas de la spécification des variables cibles. Les coûts simulés $\dot{y}_{ij}$ sont utilisés une fois en valeurs absolues, une fois avec winsorisation et une fois avec logarithmisation. La fiabilité par opération est évaluée pour chaque modèle. Pour calculer la fiabilité, nous déterminons le nombre de cabinets *justement positifs*, *faux positifs*, *faux négatifs* et *justement négatifs* (voir Tableau 26). En complément, nous calculons les chiffres clés de *sensibilité*, *spécificité* et *PPV* (valeur prédictive positive) pour chaque opération.[16]

Tableau 26 Catégories de contrôle de la fiabilité

	Effet de cabinet simulé ≤ 30 %	Effet de cabinet simulé > 30 %
Valeur d'indice calculée ≤ 130	Justement négatif	Faussement négatif
Valeur d'indice calculée > 130	Faussement positif	Justement positif

Les résultats sont divisés en quatre catégories pour examiner la fiabilité. Négatif signifie que l'effet de cabinet a été identifié comme non suspect par la surveillance statistique et positif qu'il a été identifié comme suspect. L'objectif est d'obtenir un maximum de cabinets bien identifiés (justement positifs et négatifs) et de réduire au minimum les cabinets mal identifiés (faux positifs et faux négatifs).

Source: interne.

Les formules de calcul des chiffres clés sont expliquées dans les équations (13) à (15). Le chiffre clé *Sensibilité* indique la part de cabinets médicaux inefficients qui ont été bien identifiés.

$$\text{Sensibilité} = 100 * \frac{\text{Nombre justement positifs}}{\text{Nombre justement positifs + faux négatifspositiv}} \quad (13)$$

Le chiffre clé *Spécificité* indique la part de cabinets médicaux efficients qui ont été bien identifiés.

$$\text{Spécificité} = 100 * \frac{\text{Nombre justement négatifs}}{\text{Nombre justement négatifs + faux positifs}} \quad (14)$$

Le chiffre clé *PPV* (Positive Predictive Value) indique la part de cabinets médicaux réellement inefficients parmi les cabinets classés comme tels.

$$\text{PPV} = 100 * \frac{\text{Nombre justement positifs}}{\text{Nombre justement positifs+faux positifs}} \quad (15)$$

[15] Pour évaluer la solidité de ce mode opératoire, nous avons testé une spécification alternative dans laquelle l'inefficience était ajoutée par addition. Les résultats étaient similaires.

[16] Ces chiffres clés sont définis pour évaluer la qualité du test, p. ex. pour les procédés de diagnostic médical.

8.3 Résultats de la simulation

Nous avons effectué 110 opérations de simulation au total, avec à chaque fois un calcul de tous les chiffres clés. Les résultats sont synthétisés dans le Tableau 27. Le tableau indique respectivement la valeur moyenne et, entre parenthèses, la déviation standard des chiffres clés *Sensibilité*, *Spécificité* et *PPV*. Pour l'interprétation des résultats, il faut tenir compte du fait que nous n'avons pas intégré de *particularités de cabinet* au moment de la génération des données. Les cabinets médicaux générés (N = 17 464) se différencient uniquement par la structure de leur patientèle (composante systématique) et par la variation aléatoire. Les cabinets identifiés comme faux positifs ne sont donc positifs parce qu'au tirage au sort aléatoire, des valeurs résiduelles élevées leur ont été attribuées.

Dans la partie gauche (bleu foncé) du tableau, vous trouvez les résultats du calcul d'indice avec l'estimation ponctuelle (cf. section 7.3.1). Le modèle avec la variable cible logarithmisée (modèle log.) atteint les meilleures valeurs pour les trois chiffres clés et semble donc le plus adapté à la surveillance statistique. Les déviations standards des chiffres clés sont faibles et les résultats des différentes opérations de simulation ne varient pas trop.

Dans le modèle logarithmisé, la sensibilité s'élève à 90 %, un dixième des cabinets réellement positifs ne sont donc pas identifiés dans la surveillance. Parmi les justement négatifs qui représentent la majorité, 97 % sont bien identifiés. Le PPV est particulièrement intéressant pour la surveillance statistique d'évaluation d'économicité. Il s'agit de la part de cabinets classés positifs et qui le sont réellement. Dans le modèle logarithmisé, la part identifiée est de 76 %. Cela signifie donc que pour près d'un quart des cabinets classés suspects, cette classification n'est pas correcte. Le modèle avec la variable cible absolue propose un PPV légèrement inférieur (72 %). Le modèle winsorisé quant à lui fait nettement moins bien: seulement 60 % des cabinets identifiés comme positifs le sont réellement.

Dans la partie droite (vert-jaune) du Tableau 27, on peut voir les chiffres clés pour un calcul d'indice avec la limite inférieure. Par rapport à l'estimation ponctuelle, on obtient un PPV largement supérieur, de l'ordre de 97 %. La probabilité pour qu'un cabinet réellement négatif ne soit classé comme positif que par la variation aléatoire baisse donc à 3 %. Pour la sensibilité, tout le contraire se produit: la probabilité pour qu'un cabinet réellement positif soit identifié dans la surveillance baisse à moins de 70 %.

Le comparatif des trois variantes de calcul «modèle logarithmisé», «winsorisé» et «non transformé» montre donc que dans le modèle non transformé, le PPV est même légèrement meilleur que dans le modèle logarithmisé. Ceci est probablement dû au fait que les effets de cabinet sont fortement dispersés dans le modèle transformé. Dans le modèle non transformé, le calcul d'indice avec la limite inférieure est donc particulièrement radical. La sensibilité baisse alors à 40 %, une valeur très faible dans le modèle non transformé. De manière générale, le modèle logarithmisé propose une meilleure fiabilité, et ce également dans le calcul avec la limite inférieure.

Tableau 27 Chiffres clés par modèles

	Calcul d'indice avec estimation ponctuelle			Calcul d'indice avec limite inférieure		
	Sensibilité	Spécificité	PPV	Sensibilité	Spécificité	PPV
Modèle log. M1	89,9 % (0,9)	96,9 % (0,2)	76,1 % (1,2)	66,2 % (1,5)	99,8 % (0,05)	96,8 % (0,5)
M2 winsorisé 95e centile	73,9 % (3,1)	94,5 % (1,0)	60,0 % (3,2)	35,9 % (3,3)	98,9 % (0,5)	79,5 % (4,9)
M3 absolu	80,1 % (1,5)	96,6 % (0,2)	72,4 % (1,1)	40,0 % (1,4)	99,9 % (0,01)	99,7 % (0,2)

La valeur moyenne et la déviation standard sont indiquées entre parenthèses pour chaque chiffre clé. Les chiffres clés permettent de mesurer la capacité des modèles à bien classer les cabinets médicaux. Pour chacun des chiffres clés, le modèle logarithmisé s'en sort aussi bien voire mieux que les deux autres modèles.

Source: données simulées sur la base des données de Sasis SA, calculs internes.

Le Tableau 28 reprend les chiffres clés «sensibilité», «spécificité» et «PPV» pour les groupes de médecins spécialisés avec le plus de médecins pour le modèle logarithmisé. Les chiffres clés *Sensibilité* et *Spécificité* sont proches. Les *cardiologues* ont les chiffres les plus élevés pour la sensibilité (93 %) et la spécificité (98 %). Le groupe de médecins spécialisés en *médecine pour enfants et adolescents* a la valeur la plus basse pour la sensibilité (84 %) et la spécificité (92 %). Pour le chiffre clé *PPV*, les valeurs des groupes de médecins spécialisés varient un peu plus. Le groupe de médecins spécialisés *Médecine pour enfants et adolescents* présente la plus faible valeur, avec 54 %.

Tableau 28 Chiffres clés par groupe de médecins spécialisés

	Calcul d'indice avec estimation des points			Calcul d'indice avec limite inférieure		
	Sensibilité	Spécificité	PPV	Sensibilité	Spécificité	PPV
Tous les groupes de médecins spécialisés	89,9 % (0,9)	96,9 % (0,2)	76,1 % (1,2)	66,2 % (1,5)	99,8 % (0,05)	96,8 % (0,5)
Médecine interne générale	89,4 % (1,5)	97,7 % (0,3)	81,1 % (2,0)	72,1 % (20,1)	99,9 (0,1)	96,1 (9,3)
Chirurgie	92,9 % (4,7)	97,4 % (0,9)	80,1 % (6,5)	72,2 % (7,6)	99,8 (0,2)	97,2 (2,9)
Gynécologie	91,5 % (3,2)	96,5 % (0,8)	74,4 % (5,2)	68,5 % (5,3)	99,7 (0,2)	96,7 (2,4)
Cardiologie	92,4 % (4,3)	98,3 % (0,7)	85,6 % (6,4)	74,9 % (6,8)	99,9 (0,2)	98,3 (2,6)
Médecine pour enfants et adolescents	83,9 % (4,1)	92,1 % (1,0)	54,1 % (5,0)	56,6 % (5,1)	99,2 (0,3)	88,7 (4)
Ophtalmologie	93,2 % (3,2)	97,9 % (0,6)	83,1 % (4,8)	71,9 % (5,7)	99,8 (0,2)	98 (1,8)
Psychiatrie et psychothérapie	91,0 % (2,4)	96,3 % (0,5)	73,4 % (3,3)	63,2 % (3,8)	99,8 (0,1)	97 (1,6)

La valeur moyenne et la déviation standard sont indiquées entre parenthèses pour chaque chiffre clé. Les chiffres clés sont préparés selon les groupes de médecins spécialisés choisis. Les chiffres Sensibilité et Spécificité sont relativement proches. Pour le chiffre clé PPV, les valeurs des groupes de médecins spécialisés varient un peu plus. Cette variation se réduit toutefois quand on utilise la limite inférieure.

Source: données simulées sur la base des données de Sasis SA, calculs internes.

La faible qualité du test pour le groupe de médecins spécialisés en *médecine pour enfants et adolescents* est en partie due à la conception de la simulation: les cabinets de ce groupe de médecins spécialisés sont fortement concentrés en groupes d'âge et de sexe. Comme le montre l'équation (12), les valeurs résiduelles réelles sont extraites sans pondération par le nombre de malades au moment de la génération des données. Mais les coûts moyens par patient de ce groupe de médecins spécialisés présentent une forte asymétrique positive (voir Tableau 6) et les cas dans lesquels les coûts sont les plus élevés sont particulièrement causés par des patients masculins de plus de 20 ans. Comme ce groupe de patients ne comporte qu'un faible nombre de malades par rapport aux autres groupes d'âge et de sexe, il n'est pas réaliste que les valeurs résiduelles soient extraites avec la même probabilité que pour tous les autres groupes. On obtient donc beaucoup de cabinets avec des valeurs résiduelles simulées en moyenne supérieures à zéro, ce qui rend l'identification de médecins réellement inefficients difficile.

Les résultats de la simulation permettent aussi d'analyser en quelle mesure la grande part de valeurs positives est due à la variation aléatoire pour les très petits cabinets. Cela pourrait par exemple être le cas si des aberrations élevées distordaient la valeur spécifique au cabinet de manière non justifiée pour les petits cabinets. Le Tableau 29 illustre la qualité du test par taille de cabinet (calcul d'indice avec l'estimation ponctuelle). On peut y voir que les petits cabinets n'ont pas systématiquement plus de faux positifs que les autres cabinets. On ne peut donc pas en conclure que le nombre supérieur de valeurs positives soit dû à des aberrations plus élevées. Les motifs sont donc plutôt à chercher dans les effets décrits de la sélection ou dans les incitations.

Tableau 29 Qualité du test par taille de cabinet

Nombre de malades par cabinet	Sensibilité	Spécificité	PPV
1er quartile par groupe de médecins spécialisés	84,9 % (1,9)	96,9 % (0,3)	75,6 % (2,1)
2e quartile par groupe de médecins spécialisés	90,3 % (1,5)	96,9 % (0,3)	76,4 % (1,9)
3e quartile par groupe de médecins spécialisés	91,7 % (1,4)	96,8 % (0,3)	76,4 % (2,2)
4e quartile par groupe de médecins spécialisés	92,7 % (1,3)	96,7 % (0,3)	75,9 % (2,1)

Pour les petits cabinets, les tests sont presque d'aussi bonne qualité que pour les grands. On ne peut donc pas en conclure que le nombre supérieur de valeurs positives soit dû à des aberrations plus élevées.

Source: données simulées sur la base des données de Sasis SA, calculs internes.

9 Calcul d'indice à l'aide de données individuelles

En complément aux données agrégées, nous avons pu disposer d'un bloc de données avec les données individuelles de patients (données individuelles). L'objectif est d'examiner, à l'aide de ces données, si la surveillance basée sur des données individuelles ou corrigée de certains coûts de patient élevés avant l'agrégation améliore la fiabilité. Nous avons une nouvelle fois fait appel à la simulation pour cette tâche.

9.1 Base de données

Les données individuelles contiennent des données des cantons de Berne et d'Argovie de trois assureurs (part de marché totale env. 30 %). Elles comprennent les coûts directs, les coûts prescrits et le registre des patients (une description détaillée de la préparation des données est disponible en annexe). Le bloc de données contient les sept groupes de médecins spécialisés suivants (voir Tableau 30):

- Médecine interne générale
- Chirurgie
- Gynécologie
- Cardiologie
- Médecine pour enfants et adolescents
- Ophtalmologie
- Psychiatrie et psychothérapie

Les indicateurs de morbidité disponibles sont, comme pour le bloc de données agrégées, les variables «franchise élevée», «hospitalisation pendant l'année précédente» et «PCG». Pour le PCG, c'est la quantité DDD prescrite par le médecin à certains patients qui est enregistrée (par médecin et patient). La quantité DDD a été additionnée par patient. Ceci sert aussi de facteur de morbidité, parce qu'un patient avec un PCG donné peut aussi présenter des données supérieures chez d'autres médecins et pas seulement celui qui lui a prescrit les médicaments. D'autre part, la quantité DDD a aussi été additionnée au niveau du médecin sur les GAS, puis divisée par le nombre de patients. On obtient ainsi une quantité DDD totale moyenne par patient et GAS pour chaque médecin.

Tableau 30 Vue d'ensemble bloc de données individuelles

Groupe de médecins spécialisés	Observations	Nombre de patients	Nombre de cabinets
Médecine interne générale	266 818	240 470	701
Chirurgie	11 677	11 491	57
Gynécologie	70 204	67 876	148
Cardiologie	17 478	17 033	54
Médecine pour enfants et adolescents	61 566	52 926	94
Ophtalmologie	72 056	67 817	84
Psychiatrie et psychothérapie	17 014	16 508	282
Total	516 813	369 894	1420

Au total, le bloc de données individuelles contient 1420 cabinets de sept groupes de médecins spécialisés. Le nombre de patients est égal à 369 894. Comme un patient peut être traité par plusieurs médecins, le nombre d'observations est légèrement supérieur, avec 516 813.

Source: données de décompte de trois assureurs maladie, calculs internes.

9.2 Effet spécifique au cabinet et calcul d'indice avec données individuelles

9.2.1 Comparatif entre modèles avec données individuelles

Dans la première étape, nous avons testé des modèles avec diverses variables explicatives avec les données individuelles. Dans chacun des modèles, la variable cible correspond aux coûts logarithmisés. Le Tableau 31 synthétise les résultats. D'une part, nous avons utilisé un modèle avec uniquement les GAS (M1), le deuxième modèle contient en plus des GAS les indicateurs de morbidité franchise élevée, hospitalisation pendant l'année précédente ainsi que le PCG comme quantité DDD par patient (M2). Dans le troisième modèle, nous avons utilisé la quantité DDD par patient et médecin à la place de la quantité DDD par patient (M3).

Les résultats montrent que le modèle est meilleur pour les groupes de médecins spécialisés quand les indicateurs de morbidité sont pris en compte (coefficient de détermination corrigé adj. R^2 en hausse ainsi que critères AIC ou BIC en baisse). De plus, le modèle avec le PCG au niveau médecin et patient (M3) fournit de meilleurs résultats que le modèle avec le PCG au niveau du patient (M2). Pour les évaluations suivantes, nous avons donc toujours utilisé le modèle avec GAS, hospitalisation pendant l'année précédente, franchise élevée et PCG avec quantité DDD au niveau médecin et patient.

Tableau 31 Valeur explicative des modèles examinés avec données individuelles

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants / adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	266 203	11 640	70 106	17 457	61 433	72 000	16 948
Adj. R2: coefficient de détermination corrigé							
▪ M1	0,24	0,16	0,10	0,14	0,12	0,31	0,08
▪ M2	0,34	0,16	0,12	0,16	0,15	0,39	0,21
▪ M3	0,39	0,20	0,13	0,24	0,17	0,40	0,25
AIC: critère d'information d'Akaike							
▪ M1	820 961	30 310	180 086	34 954	167 270	161 231	50 108
▪ M2	782 667	30 248	178 697	34 728	164 719	152 797	47 696
▪ M3	762 225	29 777	177 901	32 830	163 511	151 465	46 775
BIC: critère d'information bayésien							
▪ M1	821 370	30 597	180 416	35 257	167 567	161 589	50 402
▪ M2	783 349	30 660	179 255	35 139	165 233	153 394	48 183
▪ M3	762 907	30 189	178 405	33 241	163 990	151 961	47 239

- M1 = GAS uniquement
- M2 = GAS, hospitalisation pendant l'année précédente, franchise élevée, PCG au niveau patient
- M3 = GAS, hospitalisation pendant l'année précédente, franchise élevée, PCG au niveau patient et médecin

Parmi les trois modèles avec variable cible logarithmisée examinés, le modèle M3 avec GAS, hospitalisation pendant l'année précédente, franchise élevée et PCG avec quantité DDD au niveau du patient et du médecin fournit le meilleur résultat pour tous les groupes de médecins spécialisés. Dans ce modèle, on obtient le coefficient de détermination le plus élevé (adj. R2) ainsi que les valeurs les plus basses pour les deux critères d'information AIC et BIC.

Source: données de décompte de trois assureurs maladie, calculs internes.

9.2.2 Comparatif des modèles avec données individuelles et données agrégées

Pour vérifier si la constitution d'un indice basé sur les données individuelles apporte de meilleurs résultats, nous avons agrégé les données individuelles au niveau médecin et GAS et comparé les résultats avec les données individuelles. Pour y arriver, nous avons utilisé les variables explicatives du Tableau 32.

Tableau 32 Comparatif entre données individuelles et données agrégées

Données individuelles	Données agrégées
▪ GAS ▪ Franchise élevée (0/1) ▪ Hospitalisation pendant l'année précédente (0/1) ▪ PCG avec quantité DDD par médecin et patient	▪ GAS ▪ Part de patients avec franchise élevée par GAS et médecin ▪ Part de patients avec hospitalisation pendant l'année précédente par GAS et médecin ▪ PCG avec quantité DDD moyenne par patient, GAS et médecin

Source: interne.

Dans le modèle avec les données individuelles, la part de cabinets identifiés comme suspects est de 14 %, c'est-à-dire supérieure à celle du modèle avec les données agrégées avec ses 10 % (Tableau 33). Dans le modèle agrégé, près de 50 cabinets en moins sont donc identifiés comme suspects. Comme nous l'avons déjà observé à la section 7.3, les valeurs d'indice de cabinets avec peu de patients (1er quartile par groupe de médecins spécialisés) sont largement supérieures que pour les autres cabinets. La part de cabinets classés comme suspects est près de deux fois plus élevé dans cette catégorie que pour tous les autres groupes confondus. Cet effet s'observe aussi pour tous les groupes de médecins spécialisés. La chirurgie est une exception à cet égard. Ici, la part est plus élevée dans le groupe avec les plus grands cabinets (non illustré dans le Tableau 33).

Tableau 33 Valeur d'indice et part de cabinets suspects selon la taille de cabinet pour les données individuelles et les données agrégées

	1. quartile	2. quartile	3. quartile	4. quartile	Total
Nombre de cabinets suspects (valeur d'indice supérieure à 130)					
▪ Données individuelles	29,9 %	8,7 %	8,2 %	9,4 %	14,1 %
▪ Données agrégées	21,8 %	5,6 %	5,1 %	10,2 %	10,7 %
▪ Dans les deux modèles	20,9 %	4,5 %	4,8 %	7,1 %	9,37 %
Part de cabinets suspects avec limite inférieure (valeur d'indice supérieure à 130)					
▪ Données individuelles	15,9 %	1,4 %	2,5 %	3,4 %	5,8 %
▪ Données agrégées	5,6 %	0,8 %	1,4 %	3,4 %	2,8 %
▪ Dans les deux modèles	5,3 %	0,6 %	1,4 %	2,0 %	2,3 %
Valeur d'indice moyenne (pondérée avec le nombre de patients)					
▪ Données individuelles	112	97	102	98	100
▪ Données agrégées	105	94	101	101	100
Nombre moyen de patients					
▪ Médecine interne générale	149	282	405	684	380
▪ Chirurgie	90	163	236	336	204
▪ Gynécologie	233	370	493	799	473
▪ Cardiologie	118	239	359	590	323
▪ Médecine pour enfants et adolescents	253	480	741	1174	654
▪ Ophtalmologie	347	673	939	1469	857
▪ Psychiatrie et psychothérapie	20	36	56	131	60

Avec 14 %, la part de cabinets suspects est supérieure avec les données individuelles qu'avec les données agrégées (10 %). Dans les deux modèles, les cabinets avec peu de patients (1er quartile par groupe de médecins spécialisés) obtiennent des résultats nettement moins bons. La part de cabinets suspects baisse largement quand on utilise la limite inférieure de la valeur d'indice. Cet effet se fait plus ressentir avec les données agrégées.

Source: données de décompte de trois assureurs maladie, calculs internes.

Si nous utilisons la limite inférieure de la valeur d'indice (voir section 4.4), la part de cabinets suspects baisse nettement dans tous les groupes. Cet effet est même plus fort avec les données agrégées (baisse de 10,7 à 2,8 %) qu'avec les données individuelles (14,1 à 5,8 %). Comparé au chapitre 7.3, on se rend compte que le fait de prendre en compte la limite inférieure avec les

données agrégées des assureurs a une plus grande influence que pour les données de la Sasis. La raison réside probablement dans la couverture nettement inférieure: dans les données des assureurs, le nombre de patients par cabinet n'est pas aussi grand. Le calcul de l'effet spécifique au cabinet est donc moins précis.

9.3 Simulation

La simulation a été réalisée à la manière de la démarche décrite dans le chapitre 8, puis évaluée avec les mêmes chiffres clés. La génération des données se base dans ce cas sur les données individuelles absolues. Pour le calcul d'indice, nous avons calculé des modèles avec des bases de données différentes:

- Toutes les observations
- Winsorisation des coûts au-dessus du 95^{e} centile par groupe de médecins spécialisés
- Exclusion des observations avec des coûts supérieurs au 95^{e} centile par groupe de médecins spécialisés

Pour la winsorisation, ainsi que l'exclusion d'observations, nous avons à chaque fois utilisé le 95^{e} centile par groupe de médecins spécialisés tiré des données de l'année précédente.

En complément aux données individuelles, nous avons agrégé les coûts après correction au niveau du médecin et du GAS. Tant pour les données individuelles que pour les données agrégées, nous avons logarithmisé les coûts après la correction ou l'agrégation.[17]

9.3.1 Résultats de la simulation

Les résultats de la simulation montrent que quand nous utilisons un modèle avec des données individuelles (I2) le nombre de faux négatifs est très faible (0,2 %, Tableau 34). Nous obtenons donc une valeur élevée pour le chiffre clé de la sensibilité (part des inefficients qui sont justement identifiés). Le chiffre clé de la spécificité aussi a une valeur élevée (part des efficients qui sont justement identifiés) même si la part des faux positifs est supérieure, avec 3,1 % (les cabinets identifiés comme suspects bien qu'ils ne le soient pas). Mais comme cette valeur peut être mise en relation avec les justement négatifs, qui représentent 90 % dans l'idéal,[18] elle a donc moins d'importance. Le chiffre clé du PPV est nettement plus bas, ce qui signifie que 76 % des suspects sont réellement inefficients. Cette valeur ne peut être que légèrement améliorée par la winsorisation (I2). Si en revanche, on exclut les observations avec des coûts élevés, la valeur baisse même à 73 %. Il en est tout à fait autrement pour les valeurs agrégées (A1). Les valeurs faussement négatives sont supérieures (2,1 %) et les valeurs faussement positives inférieures (0,7 %). Ceci se répercute aussi sur les chiffres clés. Ainsi, avec 92 %, le chiffre clé du PPV est nettement supérieur à celui des données individuelles. Une winsorisation ou une exclusion des coûts élevés réduit certes les faux positifs, mais augmente aussi les faux négatifs.[19]

[17] Les résultats de la régression avec les données individuelles correspondraient aux résultats d'une régression pondérée avec les données agrégées si les variables explicatives dans les groupes ne variaient pas. Un modèle uniquement estimé avec le GAS et les coûts absolus fournit donc les mêmes coefficients. Mais comme les coûts ne sont logarithmisés qu'après l'agrégation, la valeur moyenne logarithmisée par groupe ne correspond plus à la valeur moyenne des coûts logarithmisés.

[18] Avec la simulation, près de 10 % se sont vu affecter des coûts 30 % supérieurs aux coûts moyens selon la structure de leur patientèle.

[19] Il faut tenir compte du fait que dans ce cas, la winsorisation a été réalisée sur les données originales. Les données ne sont donc pas directement comparables à la winsorisation avec les données de la Sasis.

Les résultats de la simulation montrent que les données individuelles ne s'en sortent pas mieux pour l'identification des médecins inefficients. De manière générale, dans le modèle avec les données individuelles (I1), 48 cabinets sont mal classés (faux positif ou faux négatif) contre 39 pour les données agrégées (A1). Une winsorisation ou l'exclusion de certaines observations n'améliore donc pas la fiabilité dans tous les cas.

Tableau 34 Chiffres clés par modèle

		Faussement négatif	Faussement positif	Sensibilité	Spécificité	PPV
I1	Données individuelles	0,2 %	3,1 %	97,7 % (1,5 %)	96,5 % (0,7 %)	76,0 % (4,8 %)
I2	Données individuelles winsorisées 95-P	0,3 %	2,7 %	96,8 % (1,8 %)	97,0 % (0,6 %)	78,6 % (4,5 %)
I3	Données individuelles exclusion 95-P	0,1 %	3,8 %	98,9 % (1,0 %)	95,8 % (0,7 %)	72,7 % (4,6 %)
A1	Données agrégées	2,1 %	0,7 %	79,8 % (3,9 %)	99,2 % (0,3 %)	92,2 % (3,1 %)
A2	Données agrégées winsorisées 95-P	2,8 %	0,4 %	72,7 % (4,2 %)	99,6 % (0,2 %)	95,0 % (2,4 %)
A3	Données agrégées exclusion 95-P	2,9 %	0,4 %	71,8 % (4,3 %)	99,5 % (0,2 %)	94,2 % (2,6 %)

La valeur moyenne de 250 simulations est indiquée pour chaque chiffre clé (déviation standard entre parenthèses). Les chiffres clés permettent de mesurer la capacité des modèles à bien classer les cabinets. Concernant la fiabilité, les modèles avec des données agrégées fournissent des résultats légèrement meilleurs à ceux avec les données individuelles.

Source: données de décompte de trois assureurs maladie, calculs internes.

Entre les différents groupes de médecins spécialisés, le PPV varie entre 70 (psychiatrie) et 83 % (cardiologie) pour le modèle avec les données individuelles. Au sein des groupes de médecins spécialisés aussi, les valeurs varient relativement beaucoup avec les opérations simulées. À cause du nombre partiellement restreint de médecins par groupe de médecins spécialisés (voir Tableau 30), quelques médecins classés autrement peuvent en effet influencer fortement le chiffre clé. Pour les données agrégées (A1), le PPV varie entre 83 (psychiatrie) et 96 % (médecine interne générale) en fonction du groupe de médecins spécialisés.

Nous avons également analysé la manière dont les chiffres clés variaient en fonction de la taille du cabinet. On s'aperçoit que le PPV pour les petits cabinets (1er quartile) est à peine plus bas que pour les autres cabinets dans le cas des données agrégées. Cet effet est le même pour tous les modèles. Pour les petits cabinets, nous n'obtenons donc pas systématiquement plus de faux positifs à cause de la variation aléatoire que pour les autres cabinets.

Tableau 35 PPV par taille de cabinet pour les données individuelles et agrégées

PPV	1. quartile	2. quartile	3. quartile	4. quartile	total
Données individuelles (I1)	61,8 %	65,0 %	66,3 %	67,8 %	65,0 %
Données agrégées (A1)	87,1 %	88,5 %	88,2 %	90,3 %	88,6 %

Source: données de décompte de trois assureurs maladie, calculs internes.

10 Synthèse

L'objectif de la présente étude était de développer les procédés statistiques, qui constituent la première partie des évaluations d'économicité des cabinets médicaux. Cinq aspects étaient essentiels à cet égard: tout d'abord, nous devions débattre et analyser le procédé mathématique actuellement utilisé sous l'éclairage de la littérature spécialisée internationale. Le deuxième objectif consistait à proposer une meilleure prise en compte de la morbidité de la patientèle et de l'intégrer de manière empirique. Troisièmement, le but était de vérifier si les caractéristiques de l'emplacement d'un cabinet contribuaient significativement à expliquer les coûts par cabinet. Quatrièmement, nous devions évaluer la fiabilité du procédé et proposer des options pour réduire le nombre de cabinets injustement classés comme positifs. Cinquièmement, l'étude visait à savoir si un calcul avec des données au niveau des patients individuels apporterait une amélioration nette de la fiabilité. Ci-après un bref résumé des principaux résultats obtenus pour les cinq aspects:

Sur le principe, nous considérons que le *procédé à deux niveaux* (composé d'une estimation «fixed effects» pour le calcul d'un effet spécifique au cabinet au premier niveau et d'une correction de cet effet au deuxième niveau) représente une bonne spécification de modèle. L'estimation «fixed effects» convient particulièrement bien pour séparer l'influence spécifique d'un cabinet des facteurs d'influence de la patientèle (indicateurs de morbidité), ainsi que de la dispersion due à la variation aléatoire.

Selon nous, les *indicateurs de morbidité* que nous avons vérifiés (niveaux de franchise, hospitalisation pendant l'année précédente et groupes de coûts pharmaceutiques) peuvent être intégrés dans le modèle au premier niveau. Pour la plupart des groupes de médecins spécialisés, ils ont une influence statistiquement significative sur les coûts et améliorent la qualité générale du modèle estimatif par rapport à une estimation avec les seuls facteurs de l'âge et du sexe. La correction de la morbidité devrait apporter une évaluation améliorée surtout pour les cabinets qui s'écartent de la moyenne pour ces facteurs.

Nous avons ensuite examiné la possibilité d'intégrer les *caractéristiques de l'emplacement du cabinet* comme *autres facteurs d'influence au deuxième niveau*. Concrètement, nous avons émis l'hypothèse selon laquelle le taux d'aide sociale, la densité démographique ou le nombre d'étrangers pouvait influencer l'effet de cabinet calculé. Dans les analyses, ces facteurs se sont toutefois avérés comme insignifiants statistiquement et d'importance limitée, de sorte qu'il n'est pas obligatoirement nécessaire de les intégrer dans le modèle. Une raison possible à l'influence limitée réside dans le fait que ces indicateurs n'existent qu'au niveau des communes et que ce n'est pas assez précis. Les grandes villes ont ainsi par exemple des quartiers avec des taux d'aide sociale très élevés et d'autres quartiers avec des taux faibles. La moyenne pour la commune est alors peu pertinente.

Nous avons ensuite analysé la *qualité du procédé statistique* (nombre de cabinets faussement positifs et faussement négatifs) en nous basant sur des simulations, en nous concentrant principalement sur deux thèmes: nous avons dans un premier temps testé deux alternatives à la transformation logarithmique actuellement utilisée des variables cibles. Les simulations ont toutefois montré que ces variantes sont inférieures à la transformation logarithmique pour des tests de qualité. Deuxièmement, nous avons développé un «indicateur d'incertitude» (à la manière d'une déviation standard) pour l'effet spécifique au cabinet. Cet indicateur peut servir à calculer une «plage de confiance» (à la manière d'un intervalle de confiance) pour l'effet spécifique au cabinet. Si à la place de l'estimation ponctuelle (moyenne), nous utilisons la limite inférieure de la plage de confiance pour calculer l'indice, le nombre de cabinets identifiés comme suspects baisse de

près de la moitié. L'évaluation de nos simulations a par ailleurs prouvé que cette démarche permettait de nettement réduire le nombre de cabinets faussement positifs. Mais cette baisse a un prix: comme le critère est plus strict, le nombre de cabinets faussement négatifs augmente. Il faut donc se demander s'il vaut mieux identifier des cabinets qui ne sont pas inefficients en réalité (faux positifs) ou ne pas identifier les cabinets qui seraient à la vérité inefficients (faux négatifs).

Les *calculs avec les données individuelles des patients* ont montré que dans les conditions présentes, ces données n'apportaient pas d'amélioration au test par rapport aux données agrégées utilisées actuellement. Même si l'agrégation entraîne la perte de données, nous ne pensons pas qu'il soit vraiment nécessaire de passer sur un calcul avec des données individuelles. En revanche, il faudrait réévaluer la situation si la base de données pouvait être élargie, par exemple en collectant toutes les données individuelles de patients ou en la complétant par des indicateurs relatifs au diagnostic, qui sont souvent utilisés à l'étranger pour analyser le caractère économique de cabinets médicaux.

11 Annexe

11.1 Transformation de la variable cible

11.1.1 Problème de la retransformation

Kaiser (2016) explique que la relation fonctionnelle entre les facteurs d'influence et la variable cible s'estime mieux dans le modèle logarithmisé. L'inconvénient réside toutefois dans le fait que les coefficients calculés indiquent l'influence d'une variable explicative sur la valeur attendue de la *variable cible logarithmisée* et pas l'influence sur la variable cible en elle-même. Pour le dire formellement, le modèle évalue:

$$ln(y) = x\beta + \epsilon \text{ avec } E(\epsilon|x) = 0$$
$$E(\ln(y)\,|x) = x\beta \tag{16}$$

Sur l'échelle des logarithmes, il faut que les valeurs résiduelles soient égales à zéro en moyenne et ne dépendent pas des variables explicatives ($E(\epsilon|x) = 0$). Cela ne veut toutefois pas dire que la valeur attendue des valeurs résiduelles une fois retransformées soit égale à un et ne dépende pas des variables explicatives ($E(\exp(\epsilon)|x) \neq 1$). La valeur dépend en particulier des variables explicatives lorsque la variance des valeurs résiduelles est en corrélation avec les variables explicatives sur l'échelle logarithmique (hétéroscédasticité). En présence d'une hétéroscédasticité, les coefficients bêta estimés ne correspondent pas directement aux semi-élasticités, mais il faudra également tenir compte de la différence entre les variances de valeurs résiduelles avec une application stricte. (Manning, 1998; Manning et Mullahy, 2001) se penchent de manière détaillée sur ce sujet.

Manning et Mullahy (2001) recommandent une estimation selon le procédé de Generalized Linear Model (GLM) en présence d'une hétéroscédasticité. Le modèle «fixed effects» décrit dans l'équation (1) devrait dans ce cas être spécifié avec une variable [0/1] individuelle pour chaque cabinet à la place de l'estimation Within proposée par Kaiser (2016). Dans le cas des grands groupes de médecins spécialisés, les modèles estimatifs nécessitent beaucoup de calculs et ne convergent quelquefois pas. Nous estimons que ce procédé manque donc de stabilité pour une utilisation opérationnelle dans les évaluations d'économicité. Nous préconisons l'utilisation du modèle logarithmisé comme le décrit Kaiser (2016), dans l'idéal avec l'indicateur d'incertitude décrit dans la section 4.4.

11.1.2 Retransformation approximative

La valeur d'indice calculée dans le chapitre 4.2 $\widehat{U}_i$ est une grandeur relative. Elle indique approximativement de combien de pour cent les coûts moyens par patient diffèrent pour un cabinet i de la moyenne de tous les cabinets. Si l'on veut calculer une grandeur absolue à partir de cette valeur, on peut multiplier les coûts moyens observés pour un cabinet par la réciproque de l'indice (divisée par 100 si l'indice est exprimé en %). Un cabinet avec une valeur d'indice de 145 et des coûts observés de CHF 1000 aurait donc des coûts attendus de 1000 / 1,45 = 690 en raison du rapport moyen entre les coûts et les variables explicatives. Les coûts supplémentaires spécifiques au cabinet, qui ne s'expliquent pas par la patientèle, s'élèvent donc à CHF 310. La formule dans l'équation (17) décrit ce mode opératoire sur le plan formel:

$$\hat{u}_i^{CHF} = \bar{y}_i \times \left(1 - \frac{1}{\exp(\hat{U}_i)}\right) \text{ et } \bar{y}_i = \frac{\sum y_{ij}}{\sum nombre\ de\ patients_{ij}} \tag{17}$$

$$\hat{\bar{y}}_i = \bar{y}_i - \hat{u}_i^{CHF}$$

Pour obtenir l'effet spécifique au cabinet approximatif en francs $\hat{u}_i^{CHF}$, on multiplie les coûts moyens réels par patient $\bar{y}_i$ par le facteur $\left(1 - 1/\exp\left(\hat{U}_i\right)\right)$. Si en complément, on veut calculer une valeur estimée pour les coûts par patient moyens prédits par le modèle $\hat{\bar{y}}_i$, il faut soustraire les coûts moyens réels $\bar{y}_i$ de l'effet spécifique au cabinet $\hat{u}_i^{CHF}$.

Ce procédé est souvent utilisé pour mesurer l'efficience des entreprises dans le contexte réglementaire. La Deutsche Bundesnetzagentur effectue ainsi par exemple un comparatif d'efficience entre les gestionnaires des réseaux de distribution d'électricité et de gaz. En utilisant un procédé de référence statistique avec les méthodes «Stochastic Frontier Analysis» (SFA) et «Data Envelopment Analysis» (DEA), on obtient une valeur d'efficience en %. Ce pourcentage est ensuite utilisé pour prescrire le montant absolu duquel une entreprise doit baisser ses prix. Cette opération a lieu en multipliant la valeur d'efficience par les recettes totales observées de l'entreprise (voir le ministère fédéral allemand pour la justice et la protection des consommateurs, ordonnance de réglementation des incitations, ARegV).

11.1.3 Calcul d'indice de la variable cible en niveaux

Si au premier niveau, on utilise les coûts sous forme de valeurs absolues au lieu de valeurs logarithmisées, l'effet spécifique au cabinet $\hat{a}_i$ et la valeur résiduelle du deuxième niveau $\hat{u}_i$ doivent alors être interprétés comme une déviation absolue par rapport à la moyenne et pas comme différence en % (voir équations 1 à 3). Pour calculer un indice standardisé pour ce modèle, nous avons d'abord déterminé les coûts attendus sur la base du modèle pour le cabinet (coûts moyens avec la patientèle donnée ($\hat{\bar{y}}_i$)). Pour ce faire, nous avons déduit la valeur résiduelle du deuxième niveau $\hat{u}_i$ des coûts moyens réels par cabinet $\bar{y}_i = \frac{1}{j}\sum y_{ij}$:

$$\hat{\bar{y}}_i = \bar{y}_i - \hat{u}_i \tag{18}$$

La valeur d'indice non standardisée est calculée comme étant le rapport entre les coûts réellement observés et les coûts moyens attendus $(\bar{y}_i/\hat{\bar{y}}_i) \times 100$. Le quotient peut être interprété comme grandeur relative et indique le pourcentage duquel les coûts moyens du cabinet sont supérieurs ou inférieurs aux coûts que le cabinet devrait avoir selon le modèle à deux niveaux avec la patientèle dont il dispose.

À la manière du modèle logarithmisé (voir chapitre 4.2), cet indice aussi est standardisé avec la moyenne de tous les cabinets médicaux d'un groupe de médecins spécialisés:

$$S_f = 100/\frac{1}{N}\sum_{i=1}^{N}\frac{\bar{y}_i}{\hat{\bar{y}}_i} \tag{19}$$

On peut donc aussi calculer des valeurs d'indice standardisées pour ce modèle:

$$\hat{U}_i = S_f \times \frac{\bar{y}_i}{\hat{\bar{y}}_i} \text{ für alle } i \text{ und } f \in \{1,2 \ldots, F\}. \tag{20}$$

11.2 Données du pool de données/tarifs Sasis et préparation

11.2.1 Vue d'ensemble des blocs de données

Aujourd'hui déjà, les données du pool de données de la Sasis SA sont utilisées pour le calcul des évaluations d'économicité. Il s'agit actuellement de la seule source de données regroupant toutes les prestations médicales facturées en Suisse à l'assurance-maladie obligatoire. Ces données constituent le fondement des calculs présentés dans les chapitres 6 à 8. Elles sont en particulier utilisées pour constituer les variables cibles, les groupes d'âge, les niveaux de franchise et l'indicateur de l'hospitalisation pendant l'année précédente (voir Tableau 36). Pour constituer le PCG, nous avons besoin de décomptes de médicaments détaillés. Ces décomptes sont disponibles dans le pool de tarifs de Sasis SA.

Tableau 36 Blocs de données disponibles

	Niveaux de regroupement	Informations disponibles	Utilisation
Pool de données registre de prestations	Groupe d'âge et de sexe, hospitalisation pendant l'année précédente (à partir de 2014), modèle d'assurance, niveau de franchise, inclusion des accidents (oui/non), type de sinistre, type de prestation	Somme des prestations facturées (coûts ou nombre de points tarifaires), somme des consultations	Compteur des variables cibles, variables explicatives âge, sexe, hospitalisation pendant l'année précédente et niveau de franchise
Pool de données registre des malades	Groupe d'âge et de sexe, hospitalisation pendant l'année précédente (à partir de 2014)	Nombre de patients différents par médecin et année	Dénominateur des variables cibles
Pool de tarifs registre des prestations	Groupe d'âge et de sexe, code tarifaire: pour les médicaments: Code pharmaceutique	Nombre de postes tarifaires facturés	Constitution des groupes de coûts pharmaceutiques

Le pool de données de Sasis SA est actuellement la seule source de données enregistrant toutes les prestations facturées à la charge de l'AOS. Les données sont complétées par le pool des tarifs, qui contient des décomptes détaillés de médicaments et permet donc aussi d'analyser le PCG.

Source: interne.

Dans l'évaluation d'économicité, on attribue les données de prestations à l'année de décompte. Cette attribution a pour avantage qu'une «année de décompte» possède une date de début et de fin clairement définie. Une alternative serait de l'attribuer à l'année de début du traitement. Mais comme les patients et les prestataires ont le droit d'envoyer leurs factures jusqu'à cinq ans après le traitement à l'assurance-maladie obligatoire, on ne pourrait réellement clôturer l'année de début du traitement qu'après expiration de ce délai, ce qui serait un réel inconvénient pour l'évaluation d'économicité. Mais comme le montre la littérature, les modèles de régression dans lesquels les variables explicatives se fondent sur la prise en compte antérieure des prestations (p. ex. coûts de l'année précédente, hospitalisation pendant l'année précédente) présentent une meilleure valeur explicative pour l'attribution à l'année de début du traitement qu'avec l'attribution à l'année de décompte (Beck 2013).

Une exception dans cette attribution réside dans les hospitalisations pour former la variable «hospitalisation pendant l'année précédente». On les constitue comme pour la compensation des risques selon l'année de début du traitement. Pour cette analyse, cela promet donc une amélioration de la valeur explicative, puisque les prestations ambulatoires qui ont été réalisées en relation avec une hospitalisation sont en partie décomptées pendant l'année suivante.

Dans le cadre du projet, nous avons vérifié si une attribution supplémentaire des PCG à l'année de début de traitement rendrait les prédictions de meilleure qualité. Cette attribution des PCG sera d'ailleurs utilisée dans la compensation des risques dès 2020. Dans nos analyses, la valeur explicative s'est toutefois dégradée lorsque l'attribution a été faite à l'année de début du traitement. Cela s'explique par le fait que les coûts (variable cible) sont alors aussi attribués à l'année de décompte. Les coûts des médicaments qui conduisent à l'attribution dans les PCG font partie des variables cibles, ce qui améliore la valeur explicative.

11.2.2 Mesures de garantie de l'anonymat des prestataires

Pour l'application pratique des calculs définie dans les objectifs du projet, des données de décompte ont dû être transmises à Polynomics. Ces données devaient toutefois rester anonymes afin que Polynomics ne puisse pas reconstituer l'identité des prestataires. L'une des principales étapes à cet égard a été l'anonymisation des codes RCC. Par ailleurs, les caractéristiques de l'emplacement du cabinet ont été transmises pour que dans chaque entité géographique visible par Polynomics, il y ait au moins cinq médecins par groupe de médecins spécialisés. Cet objectif a été atteint d'une part en transmettant la grande région à la place du canton et d'autre part en transférant une sélection limitée de caractéristiques (communes regroupées par densité démographique et regroupées par taux d'aide sociale).

11.2.3 Agrégation, liens et exclusions

Les données de décompte ayant permis de former les groupes de coûts pharmaceutiques (voir section suivante) sont disponibles par responsable anonyme, âge et sexe. Les autres blocs de données sont également agrégés à ce niveau pour pouvoir créer un lien entre les données. Les plus de 96 ans ont été regroupés dans un seul groupe au lieu des cinq groupes d'âge initiaux.

Les répercussions des agrégations sont visibles dans les deux premières lignes du Tableau 37. Le registre des prestations a été agrégé d'initialement plus de 35 millions de blocs de données à près de 750 000 blocs. Ces blocs proviennent de 22 273 numéros RCC anonymes (année de décompte 2015). En plus du registre des prestations, les données comprennent aussi le «registre des malades». Celui est nécessaire pour déterminer le nombre de malades par médecin (dénominateur de la variable cible) et pour constituer l'indicateur de morbidité «hospitalisation pendant l'année précédente». Le registre des malades est déjà agrégé au numéro RCC anonyme, au GAS et aux hospitalisations pendant l'année précédente. L'agrégation au numéro RCC anonyme et au GAS a donc réduit le bloc de données de près de la moitié.

Dans les procédés actuels, on n'analyse pas les cabinets ayant moins de 50 malades et moins de CHF 100 000 de coûts bruts par an (somme des coûts fournis et prescrits). 4000 numéros RCC anonymes ont ainsi été exclus en plus. Pour calculer les modèles, il a par ailleurs fallu exclure les observations pour lesquelles les prestations (valeur moyenne par GAS et médecin) étaient égales à zéro ou négatives. Les observations pour lesquelles il n'y avait pas d'informations sur la grande région ont également été exclues.

Tableau 37 Agrégations, liens et exclusions

	Nombre d'observations données initiales 2015	Nombre d'observations données agrégées	Nombre de codes RCC anonymes
Registre de prestations	35 321 750	752 277	22 273
Registre de malades	1 075 516	606 442	22 395
Blocs de données associés		602 885	22 137
Exclusions			
Exclusion des cabinets de moins de 50 malades et avec des coûts bruts inférieurs à CHF 100 000		29 252	4212
Exclusion des observations avec coûts nuls ou négatifs		510	0
Exclusion des observations sans informations sur les caractéristiques de la commune		1960	41
Bloc de données corrigé		571 163	17 884

Après les agrégations, mises en lien et exclusions, le bloc de données comptait 17 884 cabinets.

Source: données de Sasis SA, année 2015, calculs internes.

11.3 Diagnostic de régression

Dans la section suivante, nous débattrons de résultats choisis du calcul de régression eu égard aux éventuels problèmes statistiques.

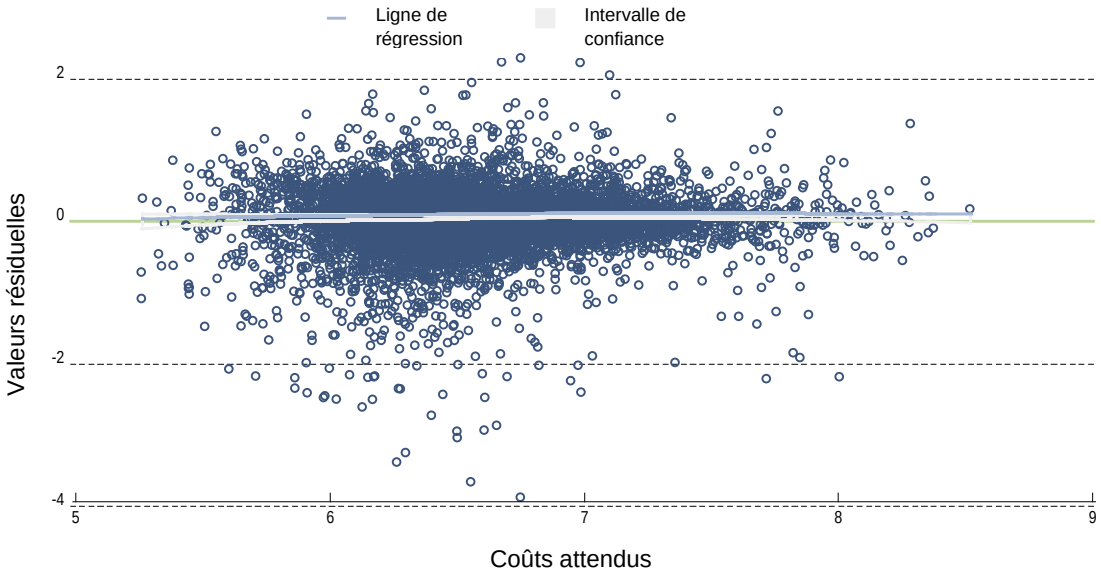
Un aspect important du diagnostic de régression correspond à la distribution des valeurs résiduelles (ε_{it} dans l'équation 1, section 4.1.1). La valeur résiduelle est définie comme étant la déviation de la prédiction du modèle par rapport à la valeur réellement observée. Une valeur résiduelle positive signifie que le modèle sous-estime l'observation concernée. Les observations avec des valeurs résiduelles négatives sont surestimées par le modèle.

11.3.1 Rapport entre valeurs résiduelles et valeurs attendues

L'analyse d'un rapport possible entre les prédictions faites dans le modèle et les valeurs résiduelles est particulièrement significative. Si ce rapport existe, cela signifie qu'une hypothèse essentielle du modèle (indépendance des valeurs résiduelles par rapport aux variables explicatives) n'est pas respectée.

La figure 8 montre l'exemple d'un modèle fonctionnant bien (groupe de médecins spécialisés: cardiologues, modèle logarithmisé). Sur l'axe horizontal, on voit les prédictions de la régression. À gauche, il s'agit donc des groupes de personnes pour lesquels des coûts faibles sont attendus (p. ex. jeunes avec franchises élevées) et à droite des personnes avec des coûts attendus élevés. Les valeurs résiduelles sont représentées sur l'axe vertical. Il n'y a pas de rapport manifeste entre les deux variables.

Figure 8 **Distribution des valeurs résiduelles, variable cible logarithmisée, cardiologues**



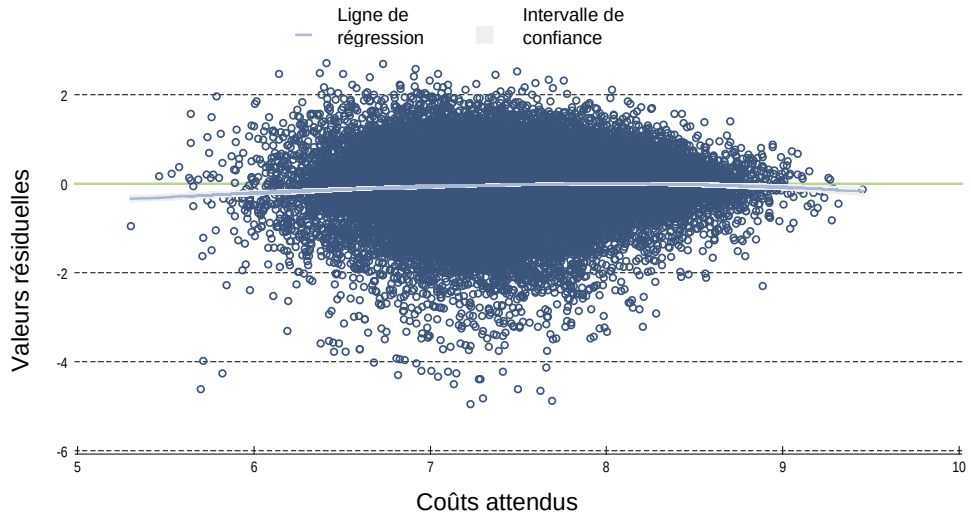
Cardiologues, chaque point correspond à une observation, N = 22 106 (médecins différents = 369).

Sur l'axe horizontal, on voit les prédictions de la régression. À gauche, il s'agit des groupes de personnes pour lesquels des coûts faibles sont attendus (p. ex. jeunes avec franchises élevées) et à droite des personnes avec des coûts attendus élevés. Les valeurs résiduelles (termes d'erreur) sont représentées sur l'axe vertical. La régression a bien fonctionné quand les points sont distribués de manière aléatoire. C'est bien le cas pour les cardiologues du modèle logarithmisé.

Source: calculs internes, Polynomics.

Une deuxième distribution des termes d'erreur du modèle logarithmique est illustrée dans la figure 9. Nous avons délibérément choisi le groupe de la psychiatrie et de la psychothérapie, parce que ce groupe de médecins spécialisés pose problème, en particulier dans le modèle non transformé. La distribution dans le cas transformé est similaire à celle des cardiologues. Dans les deux cas, les termes d'erreur sont largement plus répandus dans la zone négative que dans la zone positive. Le modèle conduit donc plutôt à des surestimations des valeurs.

Figure 9 **Distribution des valeurs résiduelles, variable cible logarithmisée, psychiatrie et psychothérapie**



Dans cette figure aussi, on ne voit pratiquement pas de rapport entre les valeurs résiduelles et les valeurs attendues. Ce qui est toutefois intéressant, c'est que les termes d'erreur sont largement plus répandus dans la zone négative que dans la zone positive.

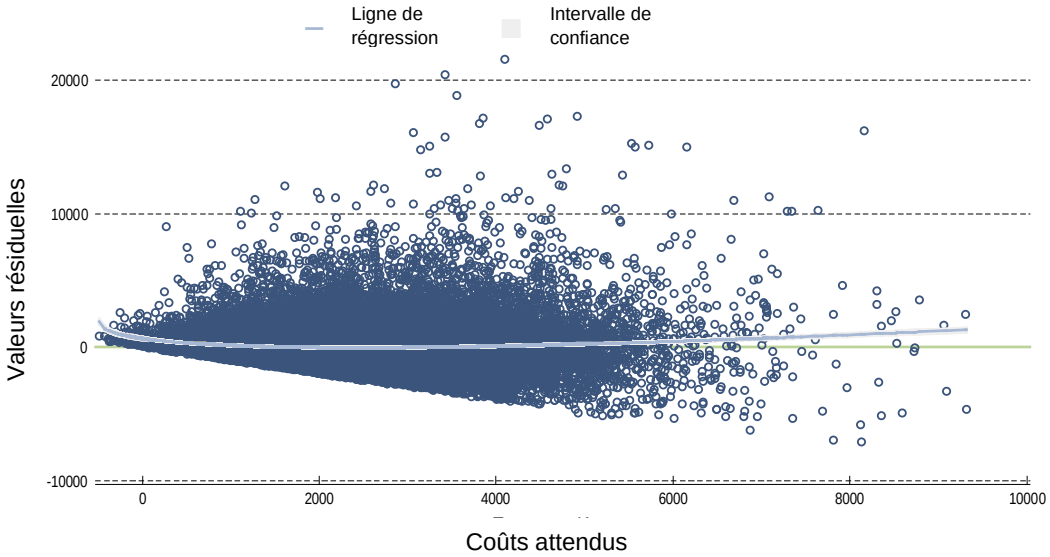
Source: données de Sasis SA, année 2015, calculs internes.

La figure 10 illustre le rapport entre valeurs estimées et valeurs résiduelles dans le modèle non transformé. Dans ce cas, les termes d'erreur positifs prennent nettement plus de place que les négatifs. Il ne s'agit pas d'une surprise parce que les valeurs aberrantes qui n'ont pas été observées dans la statistique descriptive sont difficiles à prédire par le modèle.

La ligne de tendance (trait bleu clair) est égale à zéro dans la plus grande partie de la distribution. On ne peut donc pas parler d'une tendance générale claire. Aux deux extrémités (observations avec coûts attendus très faibles et très élevés), les coûts sont toutefois sous-estimés. Pour une minorité d'observations, le modèle prédit même des coûts négatifs. Ce problème survient souvent lorsqu'on applique un modèle linéaire sur des données dont la distribution est «tronquée» d'un côté. Ce «troncage» est dû au fait que les coûts ne peuvent pas être négatifs, mais que les valeurs basses surviennent souvent. On obtient donc une «distribution avec une extrémité abrupte» comme on peut la voir dans la figure 4. Dans cette situation, les modèles linéaires ne peuvent pas prédire correctement les petits coûts et «débordent» alors dans la zone négative. Le troncage est aussi la raison pour laquelle on peut voir une «arête» dans la zone des valeurs résiduelles négatives. L'arête comprend des observations avec des valeurs relativement faibles qui ont été prédites sur une valeur trop élevée par le modèle. À cause de l'échelle, les valeurs résiduelles apparaissent toutes sur une ligne, même si elles ne sont pas identiques.

Le problème du troncage des données a souvent été abordé dans la littérature sur l'économie de la santé (voir notamment Duan et al., 1982; Buntin et Zaslavsky, 2004; Beck, 2013). Malgré ces problèmes, on utilise souvent des modèles linéaires non transformés en pratique. L'expérience prouve qu'ils autorisent souvent des meilleures prédictions que d'autres méthodes estimatives, comme par exemple les modèles Two Part ou GLM (Buntin et Zaslavsky, 2004; Beck, 2013).

Figure 10 Distribution des valeurs résiduelles, non transformée



Psychiatrie et psychothérapie, N = 52 826, médecins = 2125.

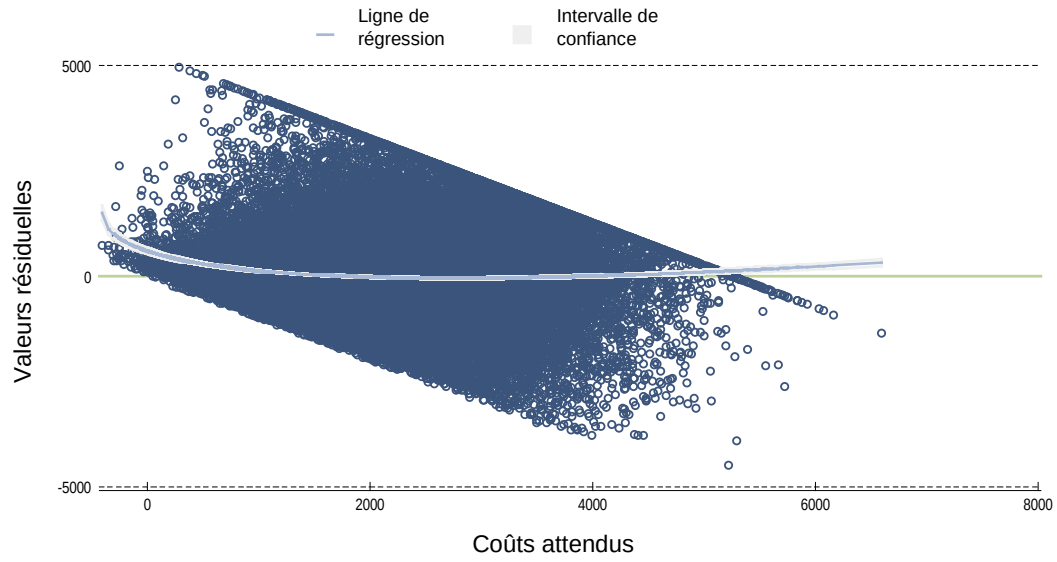
Dans un modèle non transformé, il y a une minorité d'observations pour lesquelles le modèle sous-estime largement les valeurs réelles. Ces observations ont des valeurs résiduelles positives élevées. Au niveau des valeurs résiduelles négatives, on peut voir une arête. Cette arête est due au fait qu'il y a beaucoup d'observations avec de petites valeurs dans les données. Les petites valeurs sont surestimées par le modèle.

Source: données de Sasis SA, année 2015, calculs internes.

Comme nous l'avons déjà évoqué, un modèle non transformé présente un petit groupe d'aberrations élevées comme visible dans la figure 10, qui n'est pas bien expliqué et a donc des valeurs résiduelles élevées. Si on winsorise les données originales, on peut réduire l'influence de ce groupe. La figure 11 représente l'influence de la winsorisation sur les valeurs résiduelles. Elle agit à la manière d'un «troncage» au niveau de l'extrémité inférieure, mais sur les valeurs résiduelles élevées. Ces valeurs sont alors limitées par une barrière visible.

Quand on compare la figure 10 et la figure 11, il faut se rappeler que l'échelle n'est pas la même. L'axe Y de la figure 11 a des valeurs positives moins élevées, ce qui est clairement dû à la winsorisation. On peut remarquer que l'axe X aussi ne comprend pas les mêmes valeurs. Les prédictions s'étendent en effet sur une zone plus large à cause du troncage des aberrations dans l'estimation. Cela indique que certains coefficients, et en particulier certains effets de cabinet, sont nettement supérieurs dans le modèle non winsorisé que dans le modèle winsorisé. Les cinq pour cent de valeurs les plus élevées ont donc une influence sur l'effet spécifique au cabinet.

Figure 11 Distribution des valeurs résiduelles, variable cible winsorisée, psychiatrie et psychothérapie



Psychiatrie et psychothérapie, N = 52 826, médecins = 2125.

On peut réduire l'influence des valeurs résiduelles élevées par winsorisation. Les valeurs résiduelles sont tronquées. Comme on peut le voir en comparant l'axe X à la figure 10, il y a aussi des observations pour lesquelles des valeurs attendues inférieures sont calculées.

Source: données de Sasis SA, année 2015, calculs internes.

11.3.2 Hétéroscédasticité

L'hétéroscédasticité n'entraîne pas de distorsion des estimations, mais les estimations peuvent être moins précises. Le test Breusch Pagan par exemple permet de vérifier l'importance de l'hétéroscédasticité (voir p. ex. Wooldridge, 2016). Ce test consiste à régresser la variance des valeurs résiduelles sur les variables explicatives. Avec un test F, on peut vérifier si les variables sont significatives ensemble.

Dans la suite, nous allons réaliser le test Breusch Pagan pour deux modèles différents. Premièrement pour un modèle avec toutes les variables explicatives et les effets spécifiques au cabinet et deuxièmement pour un modèle avec seulement l'effet spécifique au cabinet. Les résultats sont visibles dans le Tableau 38. Le modèle 1, dans lequel la variance est calculée avec toutes les variables, atteint un R^2 remarquable de 10 à 20 %. Le test F indique que les variables explicatives sont statistiquement significatives ensemble. Il y a clairement une hétéroscédasticité.

Dans le deuxième modèle, qui ne contient plus que les effets spécifiques au cabinet, la valeur explicative est nettement inférieure. Le rapport entre la variance des valeurs résiduelles et les effets spécifiques au cabinet n'est pas très fort, mais il existe. À l'exception des médecins en médecine pour enfants et adolescents, le test F indique que les effets spécifiques au cabinet contribuent ensemble significativement à l'explication des erreurs. Le problème de l'hétéroscédasticité existe aussi en relation avec les effets de cabinet.

Tableau 38 Test Breusch Pagan pour l'hétéroscédasticité dans les termes d'erreur

	Médecine interne générale	Chirurgie	Gynécologie	Cardiologie	Enfants / adolescents	Ophtalmologie	Psychiatrie / psychothérapie
N	196 637	9 687	22 106	12 780	16 135	30 873	52 826
N médecins	5154	278	1137	369	941	770	2125
Adj. R^2 Modèle 1	0,11	0,13	0,18	0,10	0,14	0,21	0,10
Test F Modèle 1	5,5***	5,2***	5,0***	4,3***	3,6***	10,9***	3,7***
Adj. R^2 Modèle 2	0,05	0,06	0,004	0,02	-0,02	0,17	0,02
Test F Modèle 2	3,1***	3,0***	1,1**	1,8***	0,6	9,0***	1,4***

Les coefficients R^2 indiqués se rapportent à la régression auxiliaire pour le calcul du test Breusch Pagan.

Modèle 1: toutes les variables explicatives y compris les effets spécifiques au cabinet; modèle 2: uniquement effets spécifiques au cabinet

Le test F teste la signification commune des effets spécifiques au cabinet.

Niveau de signification: *** p<0.01, ** p<0.05, * p<0.1.

Dans cette régression, la variable cible est la variance des valeurs résiduelles. Ainsi, on peut tester si les variables explicatives (modèle 1) ou les effets spécifiques au cabinet (modèle 2) sont en corrélation avec la variance des valeurs résiduelles (hétéroscédasticité). Dans les deux cas, il y a une hétéroscédasticité, mais pour les effets de cabinet, elle est plutôt faible (à l'exception des ophtalmologues).

Source: données de Sasis SA, année 2015, calculs internes.

On peut remarquer que pour les médecins en ophtalmologie, tant le coefficient R^2 que le test F prennent des valeurs largement supérieures que pour les autres médecins. Dans ce groupe de médecins spécialisés, la variance des valeurs résiduelles est largement supérieure pour certains médecins que pour les autres. Cela indique qu'il ne s'agit pas d'une erreur due au hasard, mais que la raison réside dans les caractéristiques du cabinet que nous n'observons pas. Il s'agit donc d'un effet spécifique au cabinet. L'ophtalmologie est donc un groupe de médecins spécialisés pour lequel il faudrait clarifier, dans le troisième sous-projet, si on peut lui trouver des variables explicatives spécifiques (p. ex. positions Tarmed).

11.4 Intégration de l'emplacement du cabinet

Comme décrit dans la section 7.2.2, nous avons testé diverses spécifications des caractéristiques des communes. Les coefficients correspondants sont illustrés dans le Tableau 39.

Tableau 39 Coefficients des caractéristiques de cabinet

Caractéristiques	Coefficient	Caractéristiques	Coefficient	Caractéristiques	Coefficient
Taux d'aide sociale groupe 2	-0,02 (-0.84)	Densité démographique groupe 2	-0,02 (-0,74)	Part d'étrangers groupe 2	-0,05 (-0,89)
Taux d'aide sociale groupe 3	-0,01 (-0,39)	Densité démographique groupe 3	-0,02 (-0,97)	Part d'étrangers groupe 3	-0,03 (-0,69)
Taux d'aide sociale groupe 4	-0,01 (-0,69)	Densité démographique groupe 4	-0,02 (-0,86)	Part d'étrangers groupe 4	-0,00 (-0,05)
Taux d'aide sociale groupe 5	-0,02 (-0,92)	Densité démographique groupe 5	0,02 (0,89)	Part d'étrangers groupe 5	0,01 (0,13)
Taux d'aide sociale groupe 6	0,04 (1,59)				

Modèle avec grandes régions, erreur standard entre parenthèses.

Les indicateurs testés «taux d'aide sociale», «densité démographique» et «part d'étrangers» n'avaient pas d'influence significative sur l'effet de cabinet. Le groupe de référence est celui avec la valeur la plus faible. Le groupe avec la valeur la plus élevée a toujours un coefficient positif. Un taux d'aide sociale élevé dans la commune entraîne p. ex. une valeur d'indice près de 4 % supérieure.

Source: données de Sasis SA, année 2015, calculs internes.

11.5 Préparation des données des assureurs

Les données individuelles ont été établies à partir des quatre blocs de données «coûts directs», «coûts prescrits», «registre des patients» et «registre des médicaments». Chaque bloc a été préparé séparément, puis ils ont été rassemblés. Nous disposions de données pour les années 2013, 2014 et 2015.

Tableau 40 Blocs de données partiels des données des assureurs et variables y figurant

Coûts directs	Coûts prescrits	Registre de patients	Registre de médicaments
▪ Année de décompte ▪ Canton ▪ Type de prestation ▪ Typepartenaire_LI ▪ ID patient ▪ ID émetteur de facture ▪ Montant de la facture	▪ Année de décompte ▪ Type de prestation ▪ Montant de la facture ▪ Typepartenaire_LS ▪ ID patient ▪ ID émetteur de facture ▪ ID responsable	▪ Année de naissance ▪ Sexe ▪ Franchise ▪ Modèle d'assurance ▪ Hospitalisation ▪ Année de début de traitement ▪ ID patient	▪ Année de traitement ▪ ID patient ▪ ID émetteur de facture ▪ PCG ▪ Nombre DDD ▪ ID responsable

Source: données de décompte de trois assureurs maladie, calculs internes.

Dans les coûts directs, tous les coûts négatifs (montant de facture) ont été remplacés par zéro (1174 cas). Par ailleurs, les coûts du traitement médical ont été standardisés avec la valeur moyenne du point tarifaire. Il n'était possible de ne prendre en compte qu'un type de partenaire (groupe de médecins spécialisés) par émetteur de facture. Pour les émetteurs de facture qui présentent des types de partenaire différents dans la même année pour les mêmes patients et le même type de prestation, le montant total de la facture du patient a été attribuée au type de partenaire avec le montant de facture le plus élevé (102 cas). Lorsque le montant de la facture était égal,

nous avons choisi le groupe des médecins spécialisés en médecine interne générale (2 cas). Si les types de prestation étaient différents, nous avons également choisi le groupe des médecins spécialisés en médecine interne générale (41 cas). Si un émetteur de facture avait des types de prestation différents pour différents patients, nous avons choisi le groupe de médecins spécialisés avec le plus grand nombre de patients (522 cas).

Dans les coûts prescrits, tous les coûts négatifs (montant de facture) ont également été remplacés par zéro (1 147 cas). Puis les coûts prescrits ont été regroupés avec les coûts directs et ajoutés aux coûts totaux par médecin et patient. Nous avons utilisé les variables suivantes pour regrouper les deux blocs de données:

- Coûts directs: année de décompte, ID émetteur de facture, ID patient
- Coûts prescrits: année de décompte, ID responsable, ID patient

Tableau 41 Regroupement des coûts directs et prescrits

Observations	2013	2014	2015
Coûts directs uniquement	273 804	275 499	280 977
Coûts directs et prescrits	217 878	232 144	237 507
Coûts prescrits uniquement	64 602	65 168	61 821
Total utilisé	491 682	507 643	518 484

Source: données de décompte de trois assureurs maladie, calculs internes.

Dans la troisième étape, nous avons ajouté les données du registre des patients aux données de coûts pour chaque patient. Nous avons alors utilisé les variables suivantes:

- Coûts: année de décompte, IDpatient_anonyme
- Registre de patients: année de début de traitement, ID patient_anonyme

Quand nous ne disposions pas de données pour le patient, nous avons utilisé les données de l'année précédente.

Tableau 42 Regroupement des données de coûts et de patients

Observations	2013	2014	2015
Coûts uniquement	5973	9557	14 211
Coûts et registre des patients	485 709	498 086	504 273
Coûts et registre des patients année précédente	4818	7997	12 540
Registre des patients uniquement	173 830	170 113	145 751
Total utilisé	490 527	506 083	516 813

Source: données de décompte de trois assureurs maladie, calculs internes.

Dans la dernière étape, nous avons encore ajouté le registre des médicaments aux données. Nous avons alors utilisé les variables suivantes:

- Coûts: année de décompte, ID émetteur de facture, ID patient
- Registre de médicaments: année de début de traitement, ID responsable, ID patient

Nous disposons donc au total de près de 500 000 observations par an.

Tableau 43 Regroupement des données de coûts avec le registre des médicaments

Observations par patient et médecin	2013	2014	2015
Sans PCG	365 420	375 777	388 951
Avec PCG	125 107	130 306	127 862
Total utilisé	490 527	506 083	516 813

Source: données de décompte de trois assureurs maladie, calculs internes.

12 Sources

Adams, J.L. 2009. The Reliability of Provider Profiling. Product Page. Santa Monica: RAND Corporation.

Adams, J.L., A. Mehrotra und E.A. McGlynn. 2010. Estimating Reliability and Misclassification in Physician Profiling. Santa Monica: RAND Corporation.

Adams, J.L., A. Mehrotra, J.W. Thomas und E.A. McGlynn. 2010. Physician Cost Profiling – Reliability and Risk of Misclassification. Detailed Methodology and Sensitivity Analyses (Technical Appendix). Santa Monica: RAND Corporation.

Beck, K. 2013. *Risiko Krankenversicherung – Risikomanagement in einem regulierten Krankenversicherungsmarkt*. 3. Auflage. Bern: Haupt Verlag.

Beck, N. 2011. Of Fixed-Effects and Time-Invariant Variables. *Political Analysis*, 19(02):119–122. doi:10.1093/pan/mpr010.

Buntin, M.B. und A.M. Zaslavsky. 2004. Too much ado about two-part models and transformation? *Journal of Health Economics*, 23(3):525–542. doi:10.1016/j.jhealeco.2003.10.005.

Cameron, A.C. und D.L. Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Duan, N., W.G. Manning, C. Morris und J.P. Newhous. 1982. A comparision of Alternative Models for the Demand for Medical Care. Santa Monica, CA: Rand Cooperation.

Eijkenaar, F. und R.C.J.A. van Vliet. 2013. Profiling Individual Physicians Using Administrative Data From a Single Insurer: Variance Components, Reliability, and Implications for Performance Improvement Efforts. *Medical Care*, 51(8):731–739. doi:10.1097/MLR.0b013e3182992bc1.

Eijkenaar, F. und R.C.J.A. van Vliet. 2014. Performance Profiling in Primary Care: Does the Choice of Statistical Model Matter? *Medical Decision Making*, 34(2):192–205. doi:10.1177/0272989X13498825.

Gardiol, L., P.-Y. Geoffard und C. Grandchamp. 2005. Separating Selection an Incentive Effects in Health Insurance. CEPR Discussion Paper (5380).

Kaiser, B. 2016. Methodische Weiterentwicklung der Wirtschaftlichkeitsprüfung. Basel: B,S,S. Volkswirtschaftliche Beratung AG.

Kauer, L. 2016. Long-term Effects of Managed Care: Long-term Effects of Managed Care. *Health Economics*. doi:10.1002/hec.3392.

Kristensen, T., K.R. Olsen, H. Schroll, J.L. Thomsen und A. Halling. 2014. Association between fee-for-service expenditures and morbidity burden in primary care. *The European Journal of Health Economics*, 15(6):599–610. doi:10.1007/s10198-013-0499-7.

Manning, W.G. 1998. The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem. *Journal of Health Economics*, 17(3):283–95.

Manning, W.G. und J. Mullahy. 2001. Estimating log Models: To Transform or Not to Transform? *Journal of Health Economics*, 20(4):461–94.

Mihaylova, B., A. Briggs, A. O'Hagan und S.G. Thompson. 2011. Review of Statistical Methods for Analysing Healthcare Resources and Costs. *Health Economics*, 20(8):897–916. doi:10.1002/hec.1653.

Plümper, T. und V.E. Troeger. 2007. Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis*, 15(02):124–139. doi:10.1093/pan/mpm002.

Pope, G.C., R.P. Ellis, A.S. Ash, C.-F. Liu, J.Z. Ayanian, D.W. Bates, H. Burstin, L.I. Iezzoni und M.J. Ingber. 2000. Principal Inpatient Diagnostic Cost Group Model for Medicare Risk Adjustment. *Health Care Financing Review*, 21(3):93–118.

Roth, H.-R. und W. Stahel. 2005. Die ANOVA-Methode zur Prüfung der Wirtschaftlichkeit von Leistungserbringern nach Artikel 56 KVG. Zürich: Seminar für Statistik, ETH Zürich.

von Rotz, S., U. Kunze und K. Beck. 2008. Der Ärzteindex - Ein instrument zur Beurteilung der Wirtschaftlichkeit von Grundversorgern. *Gesundheitsökonomie und Qualitätsmanagement*, 13:142–148. doi:10.1055/s-2007-963626.

Schira, J. 2009. *Statistische Methoden der VWL und BWL: Theorie und Praxis*. Pearson Deutschland GmbH.

Schmid, C. und K. Beck. 2015. Wirken hohe Franchisen kostendämpfend? *Schweizerische Ärztezeitung*, 96(35):1238–1239.

Schmidt, P. und R.C. Sickles. 1984. Production Frontiers and Panel Data. *Journal of Business and Economic Statistics*:367–374.

Schwenkglenks, M. 2010. Vergleich verschiedener Instrumente (Rechnungsstellerstatistik der santésuisse und Praxisspiegelder Trustcenter) zur Beurteilung der von Schweizer Ärzten in der Grundversorgung verursachten Behandlungskosten. Basel: Institute of Pharmaceutical Medicine / ECPM, Universität Basel.

Thomas, J.W., K. Grazier und K. Ward. 2004a. Comparing Accuracy of Risk-Adjustment Methodologies Used in Economic Profiling of Physicians. *Inquiry*: 218–231.

---. 2004b. Economic Profiling of Primary Care Physicians: Consistency among Risk-Adjusted Measures. *Health Services Research*, 39(4):985–1004.

Thomas, J.W. und K. Ward. 2006. Economic profiling of physician specialists: use of outlier treatment and episode attribution rules. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 43(3):271–282.

Trottmann, M., H. Telser, D. Stämpfli, P.K.E. Hersberger, K. Matter und M. Schwenkglenks. 2015. Übertragung der niederländischen PCG auf Schweizer Verhältnisse. Olten: Bundesamt für Gesundheit BAG.

Van de Ven, W. und R.P. Ellis. 2000. Risk adjustment in competitive health plan markets. In: *Handbook of Health Economics*, 14:755–845. Volume 1 A. Elsevier.

Wasem, J., G. Lux und H. Dahl. 2010. Beurteilung des Screening - Verfahrens der Krankenversicherer in der Schweiz zur Identifikation von Überarztung. Gutachten beauftragt von: Verein Ethik und Medizin Schweiz. Essen: ForBig - Forschungsnahe Beratungsgesellschaft im Gesundheitswesen GmbH.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4):817–838. doi:10.2307/1912934.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Bd. 1.

Wooldridge, J.M. 2016. *Introductory econometrics: a modern approach*. 6. ed. Boston, Mass.: Cengage Learning.

Polynomics AG
Baslerstrasse 44
CH-4600 Olten

www.polynomics.ch
polynomics@polynomics.ch

Téléphone +41 62 205 15 70
Fax +41 62 205 15 80