Assurances

Assurances

Application de l'analyse de crédibilité à l'assurance des accidents de travail

Michel Beuthe et Philippe Godfroid

Volume 66, numéro 2, 1998

URI : https://id.erudit.org/iderudit/1105209ar
DOI : https://doi.org/10.7202/1105209ar

Aller au sommaire du numéro

Éditeur(s)

HEC Montréal

ISSN

0004-6027 (imprimé)
2817-3465 (numérique)

Découvrir la revue

Citer ce document

Beuthe, M. & Godfroid, P. (1998). Application de l'analyse de crédibilité à l'assurance des accidents de travail. *Assurances*, *66*(2), 285–295. https://doi.org/10.7202/1105209ar

ÉTUDES TECHNIQUES

par Michel Beuthe et Philippe Godfroid

Application de l'analyse de crédibilité à l'assurance des accidents de travail

I INTRODUCTION

En assurance, il est important de bien évaluer les charges qu'implique chaque contrat ou type de contrat. En plus de sa prime, un contrat peut être caractérisé, en première approximation, par l'espérance mathémathique des débours qu'il peut entraîner ainsi que par leur variance. L'espérance mathémathique des paiements ou indemnités représente ce que la compagnie doit s'attendre à payer en moyenne à l'assuré en compensation de ses pertes. La variance indique dans quelle mesure ces pertes peuvent varier. C'est une mesure du risque couvert par le contrat.

Le service rendu par l'assurance au preneur d'assurance est de réduire ce risque en proposant de payer elle-même les coûts des sinistres éventuels tels qu'ils se présentent, en échange d'une prime constante dans le temps, qui est payée par l'assuré qu'il y ait un ou

Les auteurs :

Michel Beuthe est professeur d'économie, Département Techniques Economiques, Facultés Universitaires Catholiques de Mons (FUCaM).

Philippe Godfroid est collaborateur scientifique du Fonds National de la Recherche Scientifique (F.N.R.S) aux Facultés Universitaires Catholiques de Mons (FUCaM).

plusieurs sinistres ou qu'il n'y en ait aucun. Ce service financier et de transfert de risque, l'assureur peut le rendre car sa capacité financière lui permet d'absorber ces variations, alors que l'assuré ne peut pas ou ne désire pas les supporter. Dans cette fonction, l'assureur joue un rôle semblable à celui du banquier qui finance un achat à tempérament. Evidemment, la compagnie d'assurance qui passe de très nombreux contrats avec beaucoup d'entreprises ou d'individus peut compenser ce qu'elle doit payer par les primes qu'elle reçoit de ceux qui n'ont pas encore encouru de sinistre. Il faut donc que la somme des primes perçues couvre à tout le moins en moyenne la somme des indemnités à payer. Pour réaliser cela, il faut que la prime d'un contrat soit au moins égale à l'espérance mathématique des indemnités qu'il peut entraîner: c'est ce qu'on appelle la prime pure.

Ce transfert des risques des assurés aux assureurs et la compensation qu'ils réalisent entre les pertes des uns et les recettes des autres ne réduit cependant pas la variabilité du résultat de chaque contrat ni la variabilité des résultats totaux. Les risques sont inchangés et doivent être assumés par l'assureur sur les réserves qu'il doit constituer par une majoration adéquate de la prime pure. Il détermine ainsi la prime commerciale. L'assureur peut également se réassurer auprès d'autres compagnies de façon à réduire la variance de son propre portefeuille.

Le tarif appliqué à un contrat doit donc tenir compte en premier lieu du coût attendu des indemnités éventuelles, lequel définit la prime pure du contrat. À procéder ainsi, on obtient évidemment des tarifs différenciés et c'est de cette manière qu'il faut procéder. En effet, sous réserve de circonstances particulières d'interdépendance entre contrats, il ne serait pas rationnel sur le plan économique qu'un contrat d'assurance particulier ne soit pas tarifé en fonction de son coût attendu. Dans le cas opposé, certaines activités seraient engagées aux dépens d'autres activités, ce qui ne serait ni équitable, ni efficient. De plus les assurés ne seraient pas incités à la prudence puisque leur comportement n'aurait pas d'influence sur la prime payée.

■ 2 LES MÉTHODES DE CRÉDIBILITÉ

La première tâche de l'assurance consiste donc à calculer le coût attendu d'un contrat. Comme beaucoup de contrats sont semblables, on peut envisager d'analyser séparément les contrats d'une même espèce et de se baser sur le coût moyen de chaque type de contrat. Deux problèmes se présentent cependant. En premier lieu si ces contrats sont vraiment semblables, ils peuvent être trop peu nombreux pour donner une information suffisamment sûre. Dans ce cas, on est amené à tenir compte des résultats des autres contrats. En second lieu, les classes de contrats qui ont été constituées peuvent ne pas être vraiment homogènes car toutes les caractéristiques qui permettraient la constitution de classes homogènes ne sont pas observées.

Il y a trois façons de résoudre ce type de problème. La première est possible lorsque les contrats considérés ne sont pas trop différents et que de nombreuses de leurs caractéristiques sont observées. Dans ce cas, au lieu de répartir les contrats entre classes «homogènes» et de les analyser séparément, on peut faire une analyse globale en tenant compte des caractéristiques suffisamment détaillées des contrats pour trouver une relation qui estime le coût attendu d'un contrat particulier. Cette approche peut certainement être utilisée en assurance automobile où les compagnies disposent de nombreuses informations au sujet des conducteurs et de leurs voitures. Elle correspond à l'utilisation de diverses techniques économétriques bien connues comme l'analyse discriminante et les régressions probit ou logit.

La seconde méthode consiste à maintenir des classes distinctes de contrats mais à les analyser globalement en mesurant statistiquement le poids que l'on peut attribuer au coût moyen observé d'une classe particulière ($\bar{y}_j$) par rapport à celui que l'on peut donner au coût moyen observé de tous les contrats ($\bar{y}$). C'est l'analyse de la crédibilité qui aboutit à évaluer le coût attendu d'un contrat i de la classe j comme

$$\hat{m}_j = Z_j \bar{y}_j + \left(1 - Z_j\right) \bar{y} \qquad (1)$$

où Z_j est le facteur de crédibilité de la classe.

Cette méthode proposée par Bühlmann convient bien lorsque l'on peut facilement créer des classes de contrats mais sans pour autant avoir beaucoup d'information. C'est le cas notamment pour l'assurance des accidents du travail où il n'est pas aisé de caractériser les facteurs qui déterminent le risque, mais où il est possible de regrouper les contrats, par exemple, par leur code industriel NACE. Certes, l'appartenance à un secteur d'activité industrielle peut impliquer des risques d'accidents particuliers, mais la classification NACE n'a pas été établie pour classer les industries selon leurs risques d'accidents de travail et ne peut suffire à cet égard.

Ce modèle d'analyse peut être présenté comme suit.

Supposons que

$$\tilde{y}_{ij} = \tilde{m}_j + \tilde{e}_{ij} \tag{2}$$

soit le coût d'un contrat i de la classe j. Dans cette définition, $\tilde{m}_j = E\,[\tilde{y}_{ij}/\theta_j]$ est l'espérance conditionnelle mais aléatoire du coût du contrat i appartenant à la classe j dont le paramètre de risque est θ_j, lequel n'est pas observé.

Ajoutons que

$$\begin{array}{ll} E\,[\tilde{e}_{ij}] = 0, & E\,[\tilde{m}_j] = m \\ Var\,[\tilde{e}_{ij}] = \sigma_e^2, & Var\,[\tilde{m}_j] = \sigma_m^2 \\ Cov\,[\tilde{e}_{ij}, \tilde{e}_{kl}] = 0, & Cov\,[\tilde{m}_j, \tilde{m}_l] = 0 \text{ pour tous les } i, j, k, l \end{array} \tag{3}$$

Dans ce cas on peut démontrer que l'estimateur non biaisé de $\tilde{m}_r$ qui minimise la moyenne du carré de l'erreur est

$$\hat{m}_r = Z_r \bar{y}_r + (1 - Z_r)\,\bar{y} \tag{4}$$

où

$$Z_r = \frac{n\sigma_m^2}{n\sigma_m^2 + \sigma_e^2} \text{ et } (1 - Z_r) = \frac{\sigma_e^2}{n\sigma_m^2 + \sigma_e^2}$$

et n est le nombre identique d'observations dans chaque classe. Cette formule montre qu'une confiance totale sera accordée à la moyenne observée de la classe r notée $\bar{y}_r$ comme estimation de la moyenne pour le contrat i si la variance σ_e^2 du terme aléatoire est nulle. En effet, dans ce cas, tous les contrats de cette classe sont identiques, en ce sens qu'ils ont exactement le même coût attendu. Par contre si $\sigma_m^2 = 0$, la répartition en classes est bien sûr inutile puisque tous les $\tilde{m}_j$ sont identiques. Dans ce cas, seule la moyenne générale $\bar{y}$ doit être retenue. Ajoutons que si le nombre d'observations n_j est différent dans chaque classe, la crédibilité devient:

$$\hat{m}_r = Z_r \bar{y}_r + (1 - Z_r)\sum_j \bar{Z}_j \bar{y}_j \tag{5}$$

où

$$\bar{Z}_j = \frac{Z_j}{\sum_j Z_j}$$

Une formule semblable est obtenue également lorsque les observations étant des moyennes, $Var\ (e_{ij}) = \sigma_e^2 / n_{ij}$, où n_{ij} est le nombre d'éléments inclus dans la moyenne.

C'est le modèle proposé par Bühlmann et Straub[1].

La troisième méthode est une combinaison des deux premières approches. Si l'on dispose de suffisamment d'information, on peut vouloir «expliquer» quelque peu $\bar{y}_j$ aussi bien que $\bar{y}$ par les caractéristiques connues des contrats, tout en admettant que l'information n'est pas exhaustive au point de pouvoir retenir l'hypothèse que tous les contrats d'une classe sont identiques, au sens où $\sigma_e^2 = 0$. Après avoir estimé les $\bar{y}_j$ par régressions linéaires, par exemple, on combine ces résultats grâce à la même formule de crédibilité.

Pour un contrat i de la classe j dont les caractéristiques particulières sont décrites par le vecteur de variables explicatives X_{ir} on obtient ainsi l'estimateur

$$\hat{m}_{ir} = \hat{\beta}_r' Z_r X_{ir} + \sum_j \hat{\beta}_j' \bar{Z}_j \left(I - Z_r \right) X_{ir} \qquad (6)$$

où $\hat{\beta}_j$ est le vecteur des coefficients de la régression des y_{ij} sur les X_{ij} et Z_r est devenu unematrice, $\bar{Z}_j$ étant définis comme ci-dessus[2]. C'est l'approche proposée par Ch. A. Hachemeister (1975). Par analogie avec les formules (5), on voit qu'elle tient compte du nombre d'observations dans chaque classe. Bien sûr, deux cas extrêmes sont possibles. Si $Z_r = I$, on pourrait se contenter d'analyser séparément les facteurs explicatifs du risque de chaque classe. Par contre, si $Z_r = 0$, on pourrait se limiter à une seule analyse globale car l'information disponible serait détaillée au point où tous les facteurs de risques différents seraient tenus en compte dans l'analyse. On en reviendrait donc à la première méthode.

Dans l'analyse empirique d'un important portefeuille d'assurance des accidents du travail, nous avons utilisé la méthode de Hachemeister, ensuite, à titre de comparaison, celle de Bühlmann et Staub.

■ 3 UNE APPLICATION ILLUSTRATIVE

Le but de cette section est d'expliquer concrètement quelles sont les étapes à suivre pour appliquer la méthode de crédibilité développée par Hachemeister et de montrer la qualité des résultats qu'elle obtient. Ils seront ensuite comparés à ceux obtenus par la méthode de Bühlmann et Straub. Les données mises à notre disposition ne sont certes pas récentes mais elles conviennent parfaitement pour cette illustration du potentiel de l'analyse de crédibilité.

☐ 3.1 Le modèle de crédibilité avec régression linéaire

Le portefeuille assurance accidents du travail analysé comportait 7300 entreprises belges. Les données portaient sur trois années de 1988 à 1990. Chaque donnée annuelle a été considérée comme celle d'une compagnie différente. Seul le «risque ouvrier» a été étudié, le «risque employé» étant d'une nature différente. De même le risque du «chemin du travail» n'a pas été étudié.

Des fichiers mis à notre disposition nous avons extrait la charge totale du risque pour chaque entreprise, c'est-à-dire la somme des indemnités, frais et capitaux payés à la suite d'accidents, la masse salariale totale des ouvriers, le nombre des sinistres, ainsi que le code NACE de l'entreprise. Comme le but était d'estimer le taux de prime pure, c'est-à-dire le rapport de la charge des sinistres au capital assuré, c'est le rapport de la charge totale des entreprises à la masse salariale totale (Y) qui a été analysé. Il pouvait être expliqué à partir de la masse salariale totale (M), considérée comme une mesure de la taille de l'entreprise, ainsi que du nombre de sinistres (n) qui est une mesure de risque.

Au cours de cette période les 22 000 «entreprises» ont enregistré 45 261 sinistres pour une charge moyenne par sinistre de 540 512 francs. Sur base des codes NACE à trois chiffres le portefeuille contenait des entreprises appartenant à 218 classes. Une première étape fut de calculer pour chaque classe le régression de Y par rapport à M et N, en ne retenant que les variables significatives à 0,05. Afin de tenir compte du phénomène d'hétéroskédasticité inhérent à des données qui sont des agrégats, c'est la méthode des moindres carrés généralisés qui fût appliquée avec le nombre de sinistres comme pondération.

À ce stade la possibilité de regrouper certaines classes a été analysée. À cet effet un test en F a été pratiqué (au niveau de 0,05) pour vérifier s'il y avait un changement de structure d'une classe à trois chiffres NACE à une autre, à l'intérieur d'une même classe à deux chiffres. Ce travail a permis certains regroupements qui ont réduit le nombre de classes retenues à 164. Pour chacune de ces classes, une régression linéaire (généralisée) de Y sur les variables M et N a été calculée. Parmi celles-ci, 21 avaient un $R^2 \geq 0{,}8$, 31 avaient un $R^2 \geq 0{,}6$, et, au total, 54 seulement avaient un $R^2 \geq 0{,}2$;

de plus, 50 classes n'avaient que leur moyenne (constante) significative. Ces résultats assez médiocres pour la plupart des classes indiquent qu'il est difficile d'«expliquer» Y avec ces deux variables. Ils montrent donc tout l'intérêt d'appliquer l'analyse de crédibilité.

Dans la plupart des cas, la masse salariale totale M avait un signe négatif, lorsque cette variable était significative au niveau 0,05 de confiance. Ceci indique que les firmes de grande dimension ont un coût moyen de sinistre moins élevé. L'influence du nombre de sinistres (n) est évidemment positive.

Les régressions étant calculées, il restait à estimer les paramètres du modèle de crédibilité à savoir la variance des e_{ij}, notée σ_e^2, la matrice des covariances des coefficients $\hat{\beta}_j$ qui est nécessaire pour calculer la matrice des covariances des moyennes $\tilde{m}_j$ (correspondant à σ_m^2 du modèle simple), ainsi que, pour chaque classe, la matrice des coefficients de crédibilité (correspondant à Z_r). Ce sont des calculs itératifs assez lourds qui ne seront pas détaillés ici.

Sur base de ces éléments il est possible de calculer alors les coûts attendus, ou primes pures, qui correspondent à chaque firme selon sa classe, le nombre de sinistres enregistrés et sa masse salariale. Les primes sont donc individualisées, bien qu'elles tiennent compte de l'information donnée par l'ensemble du portefeuille. La distribution de ces primes individuelles est représentée par le premier histogramme présenté en annexe.

On s'aperçoit avec surprise que certaines des primes sont négatives. Ceci résulte du fait que les fonctions régressées sont linéaires avec un signe négatif pour la variable de la masse salariale en sorte que certains $\hat{\beta}_j Z_j X_{ij}$ sont négatifs. C'est une faiblesse bien connue de la spécification linéaire lorsque la variable dépendante ($\tilde{m}_j$) ne peut être négative. Si ces primes ne peuvent être acceptées, on doit craindre pour la même raison que d'autres primes soient exagérément élevées. Toutefois, il est facile d'y remédier en calibrant ces primes de façon à ce qu'elles soient toutes positives. Il faut veiller évidemment à ce que ces nouvelles primes conservent la position relative des primes estimées qui caractérisent les risques relatifs des différentes entreprises. Notons d'ailleurs qu'un ajustement des primes est également nécessaire pour que la somme totale des primes pures soit égale à la somme des charges payées.

Ultérieurement, toutes les primes négatives ont été ramenées à zéro, tout en abaissant les primes les plus élevées de manière à satisfaire exactement l'équilibre financier. Celles-ci furent ainsi fixées à un maximum de 9,4 %. Cette méthode pratique implique une certaine distorsion de la distribution des primes aux extrémités.

D'autres méthodes qui ne présentent pas cet inconvénient pourraient également être appliquées. Les détails de ces calculs ne sont pas donnés ici. Il n'y a pas lieu d'en discuter davantage car la calibration doit se faire finalement lors de la détermination des primes commerciales, lesquelles tiennent compte de considérations commerciales et financières qui ne sont pas envisagées dans cette analyse.

La qualité de ces résultats a été vérifiée en estimant le modèle beaucoup plus simple de Bühlmann et Straub sur les mêmes données et sur base des mêmes regroupements de secteurs. Le second histogramme (repris lui aussi en annexe) représente la distribution des primes pures calculées de cette façon. Cette méthode exclut évidemment la possibilité de primes négatives.

Bien que les deux histogrammes ne soient pas tellement différents, notez bien qu'ils ne peuvent pas être strictement comparés. En effet, le premier histogramme est fait de primes individuelles en sorte que des firmes appartenant au même secteur NACE peuvent avoir des primes fort éloignées dans l'histogramme. Par contre, les primes de toutes les firmes d'un même secteur sont identiques dans le modèle de Bühlmann et Straub et correspondent à une seule observation dans le second histogramme. Celui-ci ne tient donc pas compte de l'importance relative (en nombre de firmes) de chaque secteur, pas plus que de la dispersion des risques au sein d'un secteur. Une meilleure comparaison des méthodes est donnée pour quelques secteurs particuliers dans le tableau suivant où les résultats moyens et extrêmes de la méthode de Hachemeister sont comparés aux résultats uniques par classe de la méthode de Bühlmann et Straub[3].

La comparaison globale des deux méthodes de crédibilité peut aussi se faire en termes des R^2 calculés sur les valeurs absolues de la variable dépendante, soit $R^2 = \Sigma \hat{n}_j^2 / \Sigma Y_{ij}^2$. Celui des résultats de la méthode de Hachemeister est de 0,42 tandis que celui de la méthode

	Hach.Min	Hach.Moy	Hach.Max	Büh.Str.
Démolition	0,5	19,7	24,7	7,1
Constr. d'imm.	– 3,8	15,7	20,1	12,7
Génie civil	2,1	6,4	9,4	5,5
Installation	4,6	4,9	9	5,9
Aménagement	– 0,7	11,3	12,8	7,9

de Bühlmann et Straub n'est que de 0,089. Ceci montre la supériorité statistique de la première analyse, sous réserve des difficultés résultant de la spécification linéaire. Mais celles-ci devraient pouvoir trouver une solution dans un modèle plus général basé sur des relations non-linéaires comme celui avancé par De Vylder (1986).

L'ajustement statistique supérieur obtenu résulte de deux avantages du modèle de régression: l'introduction de variables explicatives significatives et le calcul de primes individualisées pour chaque entreprise. À ces qualités statistiques correspondent des avantages économiques. En effet, l'individualisation des primes évite le transfert des risques particuliers aux diverses entreprises. Et puisque les primes sont calculées en tenant compte du nombre de sinistres enregistrés, leur application permettrait de réduire les primes des firmes les plus sûres et d'encourager les entreprises à prendre des mesures adéquates de sécurité.

À titre d'exemple, considérons l'entreprise n° 29.440 de la classe 463 du code NACE. Elle avait une masse salariale de 39 millions et avait enregistré 26 sinistres. À ce profil particulier correspond une prime de 4,32 %. Si elle parvenait à limiter le nombre de ses sinistres à 23, la prime se réduirait à 3,72 %. De plus, si cette entreprise se développait pour atteindre une masse salariale de 41 millions, la prime pourrait être diminuée à 3,52 %.

■ CONCLUSIONS

L'objectif de cet article était de donner une explication simple et concrète ainsi qu'une illustration de l'analyse statistique de crédibilité. En effet, cette approche est bien adaptée à la segmentation de certains secteurs d'assurance car elle permet une tarification différenciée qui tient compte malgré une information relativement pauvre. Une approche alternative moins précise, développée par Bühlmann et Straub, permet elle aussi de tarifer les entreprises proportionnellement à leurs risques.

Par la technique de régression multivariée développée par Hachemeister, on obtient ce résultat en combinant par l'intermédiaire d'un coefficient de crédibilité l'information individuelle disponible avec celle de la classe de risque à laquelle elle appartient. L'illustration d'une application au secteur de l'assurance des accidents du travail a montré la qualité des résultats qui pouvaient être obtenus par la méthode de Hachemeister. L'étude a montré également la supériorité de cette méthode par rapport à cette autre variante moins rigoureuse de l'analyse de crédibilité proposée par Bühlmann et Straub.

□ Références

Beuthe M.B. et V. Denuit (1990), «La tarification de l'assurance de responsabilité civile automobile en Belgique», *Risques*, 2.

De Vylder F. (1986), "General Regression in Multidimensional Credibility Theory", in: Insurance and Risk Theory, ed. par M.J. Goovaerts, J. Haezendonck et F. de Vylder, Reidel.

Godfroid Ph. (1992), «Application de la théorie de la crédibilité aux assurances accidents du travail», mémoire d'Ingéniorat, FUCaM.

Goovaerts M.J., R. Kaas, A.E. Van heerwaarden et T. Bauwelinckx (1990), "Effective Actuarial Methods", North Holland.

Hachemeister C.A. (1975), "Credibility for Regression Models with Application to Trend", in: Credibility: Theory and Application, ed. P.M. Khan, Academy Press, N.Y.

□ Notes

1. Un bon exposé des méthodes de crédibilité est donné par M.J. Goovaerts et al. (1990).

2. Notons que β_r et X_{rt} sont des vecteurs d'ordre ($k * 1$). Par contre Z_r est une matrice d'ordre ($k * k$). Les hypothèses du modèle sont données par :

$$
\begin{aligned}
\tilde{y}_{ij} &= X_{ij}\tilde{\beta}_j + \tilde{e}_{ij},\ E\,[\tilde{y}_{ij}/\theta_r] = X_{ij}'\,\tilde{\beta}_j = \tilde{\mu}_{ij} \\
E\,[\tilde{\beta}_r] &= \beta,\ Cov\,[\beta_r] = \Omega \\
Cov\,[\tilde{e}_r] &= V_j \text{ avec les éléments } \sigma^2/\mu_{ij} \text{ dans la diagonale} \\
Cov\,[\tilde{\mu}_{ij}] &= X_j\,\Omega\,X_j' \\
Cov\,[\tilde{\mu}_{ij}, \tilde{\mu}_{kl}] &= 0
\end{aligned}
$$

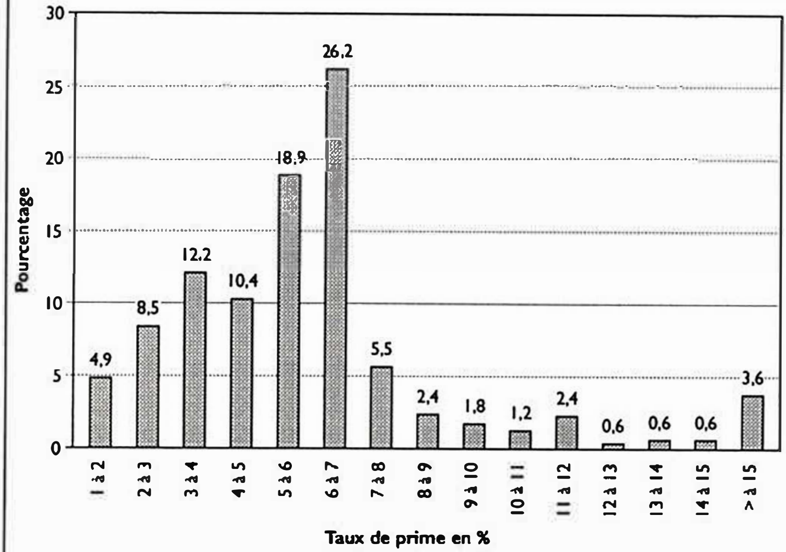
ce qui implique que

$$Z_r = (X_r'\,V_r^{-1}X_r\Omega)\,(I + (X_r'V_r^{-1}X_r\Omega))^{-1}$$

3. N.B. Un rapport de consultance à la confédération de la Construction estimait une prime pure de 5,7 pour l'ensemble du secteur de la construction.

ANNEXE
HISTOGRAMMES DE RÉPARTITION DES TAUX DE PRIME

Méthode de Bühlmann-Straub



Méthode de Hachemeister