

Etude de la corrélation entre l'environnement  
économique et la sinistralité en arrêt de travail

Paul Bonnefoy, Pierre Mousselet, Suzanne Noll

mai 2014

## **Résumé**

Cette étude a pour but de mettre au point un modèle permettant d'étudier la corrélation entre la sinistralité en arrêts de travail et l'environnement économique. Ce modèle pourrait avoir plusieurs applications pour les assureurs, notamment pour la prévision des sinistres à venir, le calcul de provisions...la projection à moyen terme de leurs comptes dans le cadre de Solvabilité II.

# Remerciements

Si un projet de cette ampleur et de cette durée demande une forte implication des étudiants, il réclame aussi l'encadrement et l'appui dont nous avons pu bénéficier. Cette aide nous a permis de mener ce travail à son terme.

Nous remercions en premier lieu Anthony Nahelou, actuariaire chez Sterenn Actuariat. Son soutien, ses conseils et sa grande disponibilité nous ont permis de retrouver le fil de notre travail lorsque nous nous égarions.

Merci à Pierre Ailliot, maître de conférence à l'Université de Bretagne Occidentale (UBO), pour ses conseils graduels. Sa rigueur et son accompagnement lors de la phase de modélisation ont rendu possible l'achèvement de cette étude. Merci à Franck Vermet pour son soutien administratif, d'une efficacité et d'une rapidité rares. Merci à l'ensemble des enseignants de l'Euro Institut d'Actuariat (EURIA) pour leur présence lors de nos passages du vendredi et leurs remarques et suggestions avisées.

Merci à Yann Quéré, directeur actuariat chez le courtier Verlingue. Sa coopération et sa rapidité nous ont redonné l'espoir quand nous l'avions perdu. Merci aux employés du ministère de la santé pour leur disponibilité et leur amabilité. Nous remercions plus particulièrement Renaud Le Gall, Audrey Baillot pour leur efficace aiguillage dans le méandre des services du ministère et Jean-Philippe Perret de la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (CNAMTS) pour sa disponibilité et le grand intérêt qu'il a porté à notre travail.

Merci à nos amis de l'EURIA pour leur solidarité et leur bonne humeur.

# Table des matières

<b>Introduction</b>	<b>7</b>
<b>1 Contexte</b>	<b>9</b>
1.1 La problématique de l'absentéisme au travail . . . . .	9
1.2 Les arrêts de travail . . . . .	10
1.2.1 Définition . . . . .	10
1.2.2 Historique . . . . .	11
1.3 Incapacité et invalidité . . . . .	12
1.3.1 Incapacité temporaire de travail(ITT) . . . . .	12
1.3.2 Invalidité permanente . . . . .	12
1.4 Tables de maintien et de passage . . . . .	13
1.4.1 Le principe des tables en assurance vie . . . . .	13
1.4.2 Les tables usuelles en prévoyance . . . . .	15
1.5 La prise en charge des arrêts de travail . . . . .	17
1.5.1 Prise en charge par la Sécurité Sociale . . . . .	17
1.5.2 Les indemnités complémentaires . . . . .	17
<b>2 Données</b>	<b>19</b>
2.1 Critères de différenciation . . . . .	19
2.2 Difficultés rencontrées pour l'obtention des données . . . . .	20
2.3 Données assurantielles . . . . .	23
2.3.1 L'obtention des données . . . . .	23
2.3.2 Les indicateurs . . . . .	23
2.4 Données globales . . . . .	24
2.4.1 Données économiques et financières . . . . .	24
2.4.2 Données démographiques . . . . .	29
2.4.3 Données médicales . . . . .	30
2.5 Données liées aux entreprises . . . . .	31
2.5.1 Données contractuelles . . . . .	31
2.5.2 Données sectorielles . . . . .	32
2.6 Corrélation entre nos variables à expliquer et nos variables explicatives . . . . .	36

<b>3</b>	<b>Modélisation mathématique</b>	<b>38</b>
3.1	Les limitations imposées à notre étude . . . . .	38
3.2	Le chargement des données avec R . . . . .	39
3.3	La mensualisation des données . . . . .	40
3.3.1	Interpolation . . . . .	40
3.3.2	Répartition . . . . .	40
3.3.3	Actualisation du taux . . . . .	40
3.3.4	Duplication . . . . .	40
3.4	Analyse en composantes principales (ACP) et corrélations . . . . .	41
3.4.1	Analyse en composantes principales . . . . .	41
3.4.2	Corrélations . . . . .	41
3.5	Régressions . . . . .	42
3.5.1	Notions théoriques . . . . .	42
3.5.2	Récapitulatif des caractéristiques de la base de données utilisée . . . . .	43
3.5.3	La modélisation . . . . .	43
3.6	Résultats . . . . .	44
3.6.1	ACP avec les variables sélectionnées . . . . .	45
3.6.2	Analyse de la corrélation entre nos variables à expliquer et les variables sélectionnées . . . . .	45
3.6.3	Inférence statistique . . . . .	46
3.7	Discussions . . . . .	47
<b>4</b>	<b>Applications</b>	<b>48</b>
4.1	Description de la base de données . . . . .	48
4.2	Exploitation possible du modèle sur la base de données fournie . . . . .	48
<b>A</b>	<b>Contacts</b>	<b>55</b>
A.1	Courriel type . . . . .	55
A.2	Interlocuteurs principaux . . . . .	56
A.2.1	CNAMTS . . . . .	56
A.2.2	BCAC . . . . .	56
A.2.3	Verlingue . . . . .	56
<b>B</b>	<b>Code R</b>	<b>57</b>
<b>C</b>	<b>Récapitulatif sur le traitement des données</b>	<b>61</b>

# Table des figures

1.1	Liens entre les différents états : « sain », incapacité, invalidité, décès	13
2.1	Données de la CNAMTS : nombre de cas d'arrêts <b>maladie</b>	22
2.2	Données de la CNAMTS : montant des indemnités journalières pour les arrêts <b>maladie</b>	22
3.1	ACP avec l'ensemble de nos variables explicatives	41
3.2	ACP avec l'ensemble de nos variables explicatives	41
3.3	Corrélation croisée entre le chômage et le nombre de cas d'accidents du travail	41
3.6	ACP uniquement avec les variables conservées après la régression (1)	45
3.7	ACP uniquement avec les variables conservées après la régression (2)	45
3.4	Résultats de la modélisation R pour la variable à expliquer « fréquence ». Les « Estimate » correspondent aux $\hat{\beta}_i$ de notre modèle linéaire.	45
3.5	Résultats de la modélisation R pour la variable à expliquer « fréquence », après utilisation de la fonction « step ».	45
3.8	Normalité des résidus (variable à expliquer : fréquence des accidents du travail).	46
3.9	Intervalles de confiance pour les paramètres $\beta$	46

# Liste des tableaux

1.1	Table de mortalité . . . . .	13
1.2	Exemple de table de maintien . . . . .	14
1.3	Table de maintien en invalidité . . . . .	16
2.1	Statistiques descriptives pour le nombre de cas par mois (en millions) et la fréquence des accidents du travail . . . . .	24
2.2	Statistiques descriptives pour le nombre de jours d'accidents du travail (en millions) et le nombre de jours moyen d'un arrêt . . . . .	24
2.3	Statistiques descriptives pour le PIB en millions d'euros . . . . .	25
2.4	Statistiques descriptives pour l'inflation . . . . .	25
2.5	Statistiques descriptives pour les taux bancaires . . . . .	26
2.6	Statistiques descriptives pour les indices boursiers . . . . .	26
2.7	Statistiques descriptives pour le cours de l'euro dollar . . . . .	27
2.8	Statistiques descriptives pour le cours du baril de pétrole BRENT en dollars US . . . . .	27
2.9	Statistiques descriptives pour la balance des paiements, en millions d'euros, sur données mensualisées . . . . .	28
2.10	Statistiques descriptives pour l'endettement des ménages en milliards d'euros . . . . .	28
2.11	Statistiques descriptives pour la confiance des ménages . . . . .	29
2.12	Statistiques descriptives pour la population française en millions d'habitants . . . . .	29
2.13	Statistiques descriptives pour les taux de natalité et de mortalité . . . . .	29
2.14	Statistiques descriptives pour le nombre de médecins généralistes en milliers . . . . .	30
2.15	Statistiques descriptives pour la consommation médicale annuelle, par personne, en euros relatifs . . . . .	30
2.16	Statistiques descriptives pour l'emploi à temps partiel exprimés en milliers . . . . .	31
2.17	Statistiques descriptives pour le salaire moyen relatif . . . . .	31
2.18	Statistiques descriptives pour l'indice du coût du travail . . . . .	32
2.19	Statistiques descriptives pour le nombre de défaillances d'entreprises . . . . .	32
2.20	Statistiques descriptives pour l'investissement des entreprises (en milliards d'euros pour le montant) . . . . .	33

2.21	Statistiques descriptives pour le taux de marge . . . . .	33
2.22	Statistiques descriptives pour l'indicateur de chiffre d'affaires par secteur . . . . .	34
2.23	Statistiques descriptives pour l'indice de confiance . . . . .	35
2.24	Statistiques descriptives pour le chômage en milliers . . . . .	35
2.25	Statistiques descriptives pour la population active en milliers . . .	36
2.26	Statistiques descriptives pour l'endettement des entreprises non financières en milliards d'euros . . . . .	36
2.27	Corrélation entre les variables à expliquer (données assurantielles) et les variables explicatives . . . . .	37
3.1	Exemple d'interpolation . . . . .	40
C.1	Récapitulatif des indicateurs . . . . .	62



# Introduction

Ce rapport est le reflet du travail que nous avons accompli dans le cadre des Bureaux d'Etudes (BE). Son principal objet d'études est la prévoyance, plus particulièrement les liens entre arrêts de travail et activité économique. Si la relation entre les deux n'est pas mise en évidence par une littérature abondante, elle est au moins supposée comme le montre l'une des analyses<sup>1</sup> de la Direction de l'Animation de la Recherche, des Etudes et des Statistiques (DARES) indique :

Entre 2005 et 2010, le risque d'accident du travail a diminué dans les secteurs concurrentiels, en partie du fait du ralentissement de l'activité économique.

Le coefficient de corrélation de Pearson est défini :  $r_p = \frac{\sigma_{xy}}{\sigma_x * \sigma_y}$  avec  $\sigma_{xy}$  la covariance entre les variables x et y et  $\sigma_x$  et  $\sigma_y$  les écart-types, respectivement de x et de y.

Nous pouvons noter que deux variables ayant une corrélation de +1 sont fortement et positivement corrélées, deux variables ayant une corrélation de -1 sont fortement et négativement corrélées et deux variables ayant une corrélation de 0 ne sont pas du tout corrélées. De plus, il faut remarquer qu'une corrélation n'implique pas un lien de causalité car cette corrélation peut provenir d'une causalité commune.

Notre étude comporte quatre chapitres.

Le premier chapitre définit le phénomène de l'absentéisme au travail. L'environnement dans lequel se place notre étude y est détaillé. Le but est de mieux en comprendre la problématique et les objectifs. Nous exposons la typologie des arrêts de travail ainsi que leurs prises en charge spécifiques. Enfin nous évoquons quelques notions de mathématiques des assurances, via les tables de mortalité, de passage et de maintien.

Le deuxième chapitre expose d'abord la philosophie avec laquelle nous avons préparé la phase de la recherche des données. Puis nous listons et expliquons les nombreux obstacles que nous avons rencontrés pendant cette partie de notre travail : ventilations ou périodes de temps insuffisantes, difficultés à trouver les organismes adéquats. . . **En effet, le principal risque et les principales**

---

1. DARES Analyses, n°10, février 2014

**contraintes de ce projet reposent sur les données** (manque de précision, de cohérence ou de chiffres notamment). La majeure partie de ce chapitre détaille chacun des indicateurs retenus : définitions, méthode d'obtention ou de calcul, justification de notre choix de l'étudier et statistiques descriptives.

Le troisième chapitre présente le coeur de notre sujet : la modélisation. Nous étudions d'abord la corrélation entre chacune de nos variables. Puis nous effectuons la modélisation. Nous utilisons le logiciel R et présentons le code informatique en annexe.

Le quatrième chapitre de ce rapport présente une des applications qui peut être faite à partir de cette étude : la prévision du montant des sinistres à partir de notre modèle et de données d'un assureur.

# Chapitre 1

## Contexte

### 1.1 La problématique de l'absentéisme au travail

L'absentéisme au travail a un coût :

- Pour la Sécurité sociale, c'est-à-dire pour la société
- Pour l'entreprise
  - coût direct : coût de l'indemnisation de l'employé absent, remplacement du salarié (salaire et formation) . . .
  - coût indirect : démotivation, détérioration du climat de travail de l'entreprise, désorganisation, baisse de productivité, perte de temps . . .
- Pour l'employé : baisse de revenu.

Le taux d'absentéisme, lié aux problèmes de santé, est de 5,5% dans le secteur privé, mais il dépasse les 11% dans certaines branches du secteur public. En 2012, il y a eu en France 250 millions de journées d'arrêt de travail en France (pour 48 millions d'actifs). Leur coût est de 9.4 milliards d'euros (dont 6.3 milliards en indemnités journalières)<sup>1</sup>. Le coût total des arrêts de travail, pris du point de vue de l'ensemble des acteurs (salariés, entreprises, système de protection sociale) est généralement estimé à trois ou quatre fois le coût des seules indemnités journalières.

Au delà de l'indemnisation de l'absence, les arrêts de travail entraînent le remboursement de soins et éventuellement le versement d'une rente. Ainsi, la prévention et la prévention<sup>2</sup> des arrêts de travail présentent un grand intérêt pour la société et pour les compagnies d'assurance.

Il est d'ailleurs étonnant que si peu d'analyses aient été menées pour expliquer les arrêts de travail d'un point de vue assurantiel ou quantitatif<sup>3</sup>. En

---

1. Article du Figaro « le coût des arrêts maladie poursuit son dérapage » 24 avril 2013

2. Notre étude bibliographique comportait de nombreux mémoires de sociologie ou de gestion des organisations dont le but était l'étude de moyens de prévention des arrêts de travail

3. Le chapitre XVIII du rapport de 2012 de la Cour des comptes sur la Sécurité Sociale mentionne, à propos de indemnités journalières, « des variations importantes constatées d'une année sur l'autre qui n'ont fait l'objet d'aucune analyse cherchant à les expliquer »

effet, la majorité des études traitent le sujet d'un point de vue purement social. Seules deux études ayant été menées concernent le champs de notre étude. Toutes deux assez anciennes, l'une prend en compte les seuls arrêts de travail et la deuxième ne concerne pas la France.

## 1.2 Les arrêts de travail

### 1.2.1 Définition

Il existe plusieurs causes d'arrêts de travail :

- Les congés maladie (simple, longue maladie, longue durée, pour accident du travail, pour maladie professionnelle)
- Les congés liés à la maternité, à la paternité ou à l'adoption
- Les congés pour événements familiaux
- Les absences pour motifs syndicaux
- Les absences pour révisions et concours.

C'est le premier type de congés qui nous intéresse particulièrement car les autres sont des impondérables, prévus et encadrés par le législateur. En outre ils sont soit prévisibles (motifs syndicaux) soit de fréquence et de durée négligeables (concours, événements familiaux) comparativement aux absences pour maladies.

Un accident est un événement soudain, provoquant une lésion du corps. L'accident du travail couvre tout accident survenant à un salarié du fait ou à l'occasion de son travail : accidents de trajet, accidents ayant lieu pendant les périodes périphériques au travail (habillage, pauses, repas...). Cependant des exceptions sont possibles sous différentes conditions prévues par la loi.

Une maladie est dite professionnelle si elle résulte des conditions d'exercice de l'emploi ou de l'exposition à un risque du fait du travail. Les maladies professionnelles sont listées (par type d'affection ou de risque) dans plusieurs dizaines de tableaux comprenant, pour chaque pathologie, le délai pendant lequel on peut déclarer la maladie après la fin de l'exposition.

Les accidents du travail, les accidents de trajet et les maladies professionnelles sont définis dans le livre 4 du code de la Sécurité sociale (L.411 et L.461 pour les définitions stricto sensu). Le code du travail mentionne également les arrêts maladie et les accidents du travail, mais du point de vue de leurs conséquences sur le contrat de travail.

Les jours de carence sont les jours pendant lesquels le salarié ne sera pas indemnisé au début de son arrêt de travail. Ainsi, s'il n'y a pas de jours de carence, la fréquence d'arrêt de travail sera plus élevée car les petits arrêts seront également pris en compte. Pour la même raison, le nombre de jours total sera plus important alors que le nombre de jours moyen sera plus faible.

Une franchise est un montant restant à la charge de l'assuré : il devra nécessairement le payer lorsqu'il sera pris en charge. Un assuré sera plus réticent à se mettre en arrêt de travail s'il a un reste-à-charge. Ainsi, l'existence d'une franchise entraîne une diminution du nombre et de la fréquence des arrêts de travail. Le nombre de jours et le nombre de jours moyen des arrêts seront eux

aussi impactés à la baisse : pourquoi déclarer un sinistre (et donc s'exposer au risque de voir sa prime d'assurance augmenter ou son contrat résilié par son assureur pour sur-sinistralité - dans le cas des assurances privées) si la franchise est trop élevée par rapport au montant d'indemnisation espérée.

### 1.2.2 Historique

En 1868, une caisse d'assurance contre les accidents du travail est créée. L'assurance contre ce risque est encadrée par la loi à partir de 1884. La responsabilité patronale est reconnue à partir de 1898. En effet, la loi du 9 avril 1898 définit un système de compensation automatique et forfaitaire aux frais de l'employeur pour les victimes d'accident de travail : un accident est présumé d'origine professionnelle s'il survient sur le lieu et le temps de travail. Cette loi a pour but d'empêcher la recherche d'une responsabilité de l'employeur ou de son salarié, étant uniquement fondée sur la présence d'un accident : on parle de responsabilité sans faute.

Dès 1905, l'assurance s'étant généralisée au niveau des entreprises, les employés victimes d'accident du travail sont autorisés à démarcher directement auprès de l'assureur de leur employeur pour obtenir un dédommagement de leur accident.

En 1919, la loi du 9 avril 1898 est étendue aux maladies professionnelles. La loi de 1919 définit l'actuel tableau des maladies professionnelles. La notion d'arrêt de travail apparaît en 1928. A partir du sixième jour d'arrêt justifié, le salarié a droit à un demi salaire par jour ouvrable, jusqu'à guérison ou consolidation, pour un maximum de six mois.

A partir de 1935, l'indemnité est comprise entre trois et dix-huit francs. Elle peut être versée dès le quatrième jour si le salarié a au moins trois enfants à charge.

La loi du 14 juin 1938 marque l'unification du contrôle de l'État sur les entreprises d'assurance de toute nature et ainsi le point de départ de la réglementation française des entreprises d'assurance.

En 1946, la législation évolue, abrogeant la loi d'avril 1898. Le dédommagement des arrêts de travail et des maladies professionnelles est désormais organisé par la Sécurité Sociale récemment créée. Cette dernière travaille en lien avec les assureurs privés dont l'établissement est antérieur. Une assurance obligatoire ainsi qu'un système de solidarité nationale permettent alors la prise en charge des arrêts de travail et des maladies professionnelles.

Les fonctionnaires n'ont jamais eu de jour de carence, sauf en 2012 et 2013 où ils en avaient un et ce jour devrait être supprimé en 2014.

## 1.3 Incapacité et invalidité

### 1.3.1 Incapacité temporaire de travail(ITT)

L'incapacité temporaire de travail correspond à l'impossibilité médicalement reconnue, complète et continue de travailler. C'est un **état transitoire**, qui conduit au passage en invalidité (après consolidation) ou au retour à l'emploi. Il est impossible d'être en incapacité plus de 36 mois : au delà, le passage en invalidité est automatique. En revanche, il est possible de passer en invalidité avant les 36 mois, en cas de consolidation de l'état du malade.

### 1.3.2 Invalidité permanente

La consolidation d'un état de santé correspond à sa stabilisation. Elle intervient souvent quand l'incapacité d'un patient n'est plus susceptible de s'améliorer et se commue donc en invalidité (partielle ou totale). La consolidation est constatée par un médecin. L'invalidité est donc permanente et présumée définitive.

#### Taux d'invalidité

Le taux d'invalidité correspond au pourcentage d'invalidité atteint suite à un accident ou à une maladie. Il détermine le droit à la prise en charge au titre de l'invalidité par la Sécurité sociale et par les assurances (individuelles ou collectives). Le calcul de ce taux est basé sur le barème de la Sécurité sociale ou sur un tableau croisé utilisant :

- le taux d'incapacité fonctionnelle (par rapport aux gestes de la vie quotidienne)
- le taux d'incapacité professionnelle (par rapport à l'activité professionnelle, en prenant en compte les possibilités de reclassement).

Si une activité est toujours possible on pourra parler également d'une invalidité de 1<sup>re</sup> catégorie. Si aucune activité n'est possible on parlera d'invalidité de 2<sup>me</sup>, ou de 3<sup>me</sup> catégorie si en outre le recours à une tierce personne est indispensable.

#### Invalidité permanente partielle (IPP) ou totale (IPT)

L'invalidité permanente partielle (IPP) est une invalidité physique ou mentale partielle survenue avant 65 ans (c'est-à-dire avant l'ouverture des droits à la retraite) et correspondant à un taux d'invalidité compris entre 33% et 66%. L'invalidité permanente totale (IPT) est aussi une invalidité physique ou mentale survenue avant 65 ans, mais celle-ci est considérée comme totale. Le taux d'invalidité doit être supérieur à 66%. Dans les cas les plus graves, lorsque l'incapacité totale et définitive de travailler s'accompagne du recours obligatoire d'une tierce personne pour les gestes habituels de la vie, on parle de perte totale et irréversible d'autonomie (PTIA). La mise en invalidité ne rompt pas le contrat de travail qui lie l'employeur et son salarié.

Le schéma ci-dessous (1.1) présente les mécanismes de transition d'un état à l'autre entre une date  $t$  et  $t+1$ .

FIGURE 1.1 – Liens entre les différents états : « sain », incapacité, invalidité, décès

Pour quantifier le passage d'un état à un autre, notamment dans le calcul des provisions techniques en arrêt de travail, les actuaires utilisent des tables de mortalité et de maintien. Dans le cadre des calculs d'inventaire en norme comptable française, ces tables sont réglementaires et diffusées par l'Institut des Actuaires. Nous présentons dans le paragraphe suivant la typologie de telles tables.

## 1.4 Tables de maintien et de passage

### 1.4.1 Le principe des tables en assurance vie

#### Les tables de mortalité

Une table de mortalité, ou table de survie, est un tableau contenant en général deux ou trois colonnes. Il permet de suivre l'évolution d'une population : le nombre de décès ( $d_x$ ), les probabilités de décès ( $q_x$ ) et de survie ( $p_x$ )... Elle se construit à partir d'une population fictive, comptant généralement, à l'âge 0, 100000 individus.

Age x	Nombre de vivants à l'âge x	nombre de décès entre les âges x et x+1
0	$l_0$	$d_1 = l_1 - l_0$
1	$l_1$	$d_1 = l_1 - l_2$
2	$l_2$	$d_2 = l_2 - l_3$
...	...	...
x	$l_x$	$d_x = l_x - l_{x+1}$
...	...	...
w	$l_w$	$d_w = l_w - l_{w+1} = l_w$

TABLE 1.1 – Table de mortalité

Nous notons  $p_x = \frac{l_{x+1}}{l_x}$  la probabilité qu'un individu d'âge  $x$  soit encore vivant à l'âge  $x+1$ . De même  $q_x = \frac{d_x}{l_x} = 1 - p_x$  est la probabilité pour qu'un individu d'âge  $x$  décède avant d'avoir l'âge  $x+1$ .

#### Transposition aux cas de l'incapacité et de l'invalidité

Une table de mortalité pour les personnes en incapacité ou en invalidité fonctionne exactement de la même manière qu'une table de mortalité classique,

à la différence qu'elle ne concerne qu'un sous-échantillon de la population et non la population dans son ensemble.

Une table de maintien contient :

- Dans la première colonne, l'âge **d'entrée** en incapacité ou en invalidité (selon qu'il s'agisse d'une table de maintien en incapacité ou en invalidité). Les garanties incapacité et invalidité ne concernent que les personnes ayant une activité professionnelle. Ainsi, l'âge dans les tables de maintien, ne commence pas à 0 an comme dans les tables de mortalité classiques, mais autour de 20 ans.
- Dans la première ligne, le temps écoulé depuis l'entrée dans l'état d'incapacité (ou d'invalidité). Ce temps est exprimé en mois pour l'incapacité et en années pour l'invalidité. Après trois ans d'incapacité, toute personne sera considérée invalide : on s'arrête donc à  $3 \times 12 = 36$  mois dans la table de maintien en incapacité. Pour l'invalidité, on va jusqu'à l'âge de la retraite (62 ans).
- Le reste du tableau indique le nombre de personnes restant en incapacité (respectivement en invalidité) parmi la population de départ (souvent 10000 individus).

Par exemple, dans le cas d'une table de maintien en invalidité,

Années \ Âge	0	1	2	3	...
20	10000	9859	9699	9534	...
...	...	...	...	...	...
34	10000	9827	9669	9523	...
35	10000	9818	9663	9509	...
36	10000	9805	9641	9495	...
37	10000	9801	9640	9501	...
...	...	...	...	...	...

TABLE 1.2 – Exemple de table de maintien

Le nombre dans la cellule verte s'interprète ainsi : « Parmi 10000 individus entrant en invalidité à 35 ans, 9663 sont encore invalides à  $35 + 2 = 37$  ans » De même, on peut dire de la cellule rouge : « Parmi 10000 individus entrant en invalidité à 35 ans, 9509 sont encore invalides à  $35 + 3 = 38$  ans » Enfin, pour la cellule jaune : « Parmi 10000 individus entrant en invalidité à 36 ans, 9641 sont encore invalides à  $36 + 2 = 38$  ans »

Une personne en incapacité peut :

- guérir
- atteindre l'âge de la retraite
- devenir invalide (après consolidation de son état)
- mourir.

Une table de passage en invalidité donne, pour un âge d'entrée en incapacité et une durée d'incapacité donnés le nombre de personnes qui passent en invalidité



parmi la cohorte de départ de 10000 individus.

### 1.4.2 Les tables usuelles en prévoyance

Le Bureau Commun des Assurances Collectives a construit :

- Deux tables de maintien en incapacité et en invalidité, qui servent au provisionnement des prestations versées au titre de l'incapacité et de l'invalidité. Elle a été construite en 1993 (à partir d'un portefeuille issu des quatre principaux assureurs de l'époque : AXA, les Assurances Générales de France (AGF)<sup>4</sup>, Groupe des Assurances Nationales (GAN)<sup>5</sup> et l'Union des Assurances Parisiennes (UAP)<sup>6</sup>). Ces tables ont été homologuées et intégrées au code des assurances en 1996. Elles ont été mises à jour en 2012, suite à la réforme des retraites, afin de prolonger l'âge de départ.
- Une table de mortalité pour les personnes en incapacité ou en invalidité (suite à une étude faite en 2000), dont l'utilisation n'est pas imposée par la réglementation. Elle permet de déterminer une probabilité de survie à partir de l'âge et de l'ancienneté de l'incapacité ou de l'invalidité de l'individu.

---

4. En 2007, les AGF ont fait l'objet en 2007 d'une offre publique d'achat d'Allianz, déjà actionnaire majoritaire. Le nom commercial disparaît en 2009.

5. La société a été rachetée par Groupama en 1998, mais la marque commerciale GAN perdure.

6. Cette société, alors 3<sup>me</sup> assureur en France et n°1 en dommages, a fait l'objet d'une offre publique d'achat d'AXA en 1996. La fusion d'AXA et de l'UAP a eu lieu en 1999.

Années	Age																				
	0	1	2	3	4	5	6	7	8	9	...	33	34	35	36	37	38	39	40	41	42
20	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6292	6181	6113	6027	5939	5852	5760	5663
21	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6292	6181	6113	6027	5939	5846	5747	
22	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6292	6181	6113	6027	5933	5833		
23	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6292	6181	6113	6017	5916			
24	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6292	6181	6084	5981				
25	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6292	6193	6089					
26	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6371	6271	6165						
27	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6449	6348	6241							
28	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6426	6318								
29	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...	6394									
30	10000	9859	9699	9534	9331	9163	8994	8874	8761	8696	...										
31	10000	9868	9731	9538	9364	9174	9013	8913	8815	8756	...										
32	10000	9843	9698	9537	9306	9120	8985	8846	8771	8685	...										
33	10000	9844	9705	9562	9328	9131	8997	8872	8789	8695	...										
34	10000	9827	9669	9523	9301	9084	8908	8770	8665	8561	...										
35	10000	9818	9663	9509	9281	9039	8874	8734	8597	8455	...										
36	10000	9805	9641	9495	9258	9038	8852	8724	8573	8456	...										
37	10000	9801	9640	9501	9269	9051	8861	8757	8601	8445	...										
38	10000	9787	9620	9462	9253	9050	8864	8761	8590	8416	...										
39	10000	9751	9566	9414	9214	9018	8861	8762	8586	8397	...										
40	10000	9751	9562	9424	9214	9012	8843	8730	8537	8359	...										
41	10000	9756	9543	9416	9210	8979	8800	8670	8474	8279	...										
42	10000	9772	9580	9432	9243	9010	8817	8673	8444	8250	...										
43	10000	9769	9578	9417	9265	9014	8816	8686	8449	8265	...										
44	10000	9751	9567	9394	9249	8990	8791	8665	8435	8245	...										
45	10000	9774	9619	9458	9303	9049	8835	8693	8471	8293	...										
46	10000	9814	9656	9493	9332	9093	8874	8707	8486	8281	...										
47	10000	9832	9698	9529	9373	9119	8887	8712	8490	8320	...										
48	10000	9846	9713	9534	9360	9138	8900	8709	8492	8317	...										
49	10000	9850	9709	9511	9328	9102	8839	8647	8432	8267	...										
50	10000	9857	9721	9514	9316	9104	8843	8656	8463	8296	...										
51	10000	9869	9721	9526	9297	9056	8772	8565	8382	8202	...										
52	10000	9903	9772	9571	9331	9084	8764	8559	8357	8226	...										
53	10000	9895	9735	9523	9242	8983	8651	8422	8290	8150	...										
54	10000	9895	9752	9549	9268	8968	8650	8515	8371		...										
55	10000	9881	9738	9533	9257	8936	8687				...										
56	10000	9864	9713	9502	9224	9080	8927				...										
57	10000	9872	9721	9509	9360	9202					...										
58	10000	9856	9672	9521	9360						...										
59	10000	9840	9686	9523							...										
60	10000	9843	9677								...										
61	10000	9831									...										

TABLE 1.3 – Table de maintien en invalidité

## 1.5 La prise en charge des arrêts de travail

### 1.5.1 Prise en charge par la Sécurité Sociale

#### En cas d'arrêt maladie

Les soins sont pris en charge selon les règles d'indemnisation (base de Remboursement et taux de prise en charge) du régime de Sécurité Sociale dont dépend le travailleur en arrêt.

Le salaire journalier de base correspond à la moyenne des salaires journaliers bruts des trois derniers mois travaillés (ou, en cas d'activité discontinuée, des 12 derniers mois). Ce montant est plafonné à 1,8 fois le SMIC. C'est ce salaire qui servira de base au calcul de l'indemnité journalière. Cette dernière - versée sous conditions de cotisation - est égale à 50% du salaire journalier de base, soit au maximum 42,77€ pour 2014. Elle peut cependant être majorée aux deux tiers du salaire de base à partir du 31<sup>ème</sup> jour d'arrêt pour les personnes ayant au moins trois enfants à charge.

L'indemnité n'est versée qu'à partir de quatrième jour d'arrêt pour les salariés du privé (cf 3 jours de carence). A noter qu'en Alsace-Moselle, l'employeur doit maintenir l'intégralité du salaire pendant tout le délai de carence.

#### En cas d'accident du travail, d'accident de trajet ou de maladie professionnelle

Le salaire journalier de base correspond au salaire mensuel brut du mois précédent (ou de la moyenne des douze mois précédents en cas d'activité discontinuée) divisé par 30,42. L'indemnité journalière est égale à 60% de ce salaire de base (plafond de 185,3€) jusqu'au vingt-huitième jour d'arrêt. A partir du vingt-neuvième jour d'arrêt, elle est portée à 80% du salaire journalier de base (avec un plafond de 247,07€).

Les soins sont pris en charge à 100%.

### 1.5.2 Les indemnités complémentaires

#### Versées par l'employeur

Ces indemnités complémentaires sont rendues obligatoires par la loi. Elles sont versées par l'employeur aux salariés ayant plus d'un an d'ancienneté. La loi prévoit aussi un délai de carence de 7 jours. Les indemnités complémentaires sont versées en complément des indemnités de la Sécurité Sociale afin d'assurer au salarié un revenu minimum :

- de 90% de la rémunération brute pendant les 30 premiers jours d'arrêt
- puis de deux tiers de la rémunération brute pendant les 30 jours suivants.

Les durées maximales de versement augmentent avec l'ancienneté dans l'entreprise (jusqu'à 90+90 jours à partir de 31 ans).

Les conventions collectives peuvent prévoir le maintien intégral du salaire, réduire le délai de carence ou encore augmenter la durée maximale de versement

des indemnités. Ces conventions peuvent aussi réduire, voire supprimer, la condition d'ancienneté.

### **Versées par une protection individuelle**

L'employé peut aussi se couvrir contre les arrêts de travail par une couverture individuelle. En effet, il existe des contrats de prévoyance couvrant les arrêts de travail. Si ce type de contrats vise particulièrement les indépendants et les professions libérales, il est cependant ouvert à tous.

# Chapitre 2

## Données

### 2.1 Critères de différenciation

Pour être le plus précis possible dans notre modélisation, il nous faut être le plus précis possible dans nos données. Cela passe :

- par la fiabilité des valeurs
- par une bonne segmentation des indicateurs

Ainsi, nous choisissons les sources de données les plus officielles et sérieuses :

- L’Institut National des Statistiques et des Etudes Economiques (INSEE)
- Les sites des ministères de la République Française, notamment le Ministère de l’Economie et des Finances et le Ministère du Travail, de l’Emploi, de la Formation Professionnelle et du Dialogue Social
- La Banque de France
- ...

Nous voulons détailler au maximum nos données. Nous avons donc plusieurs critères de différenciations, qui nous permettront de tenir compte de certaines disparités :

- En différenciant le secteur privé et le secteur public
- Selon la taille des entreprises (TPE, PME, Grandes Entreprises)
- Selon la qualification des personnes (ouvriers, employés, professions intermédiaires, cadres)
- Selon le secteur d’activité. On peut notamment distinguer :
  - Le secteur primaire : activités basées sur l’exploitation de ressources naturelles
  - Le secteur secondaire : activités transformant des matières premières
  - Le secteur tertiaire : activités de services au sens le plus large (commerce, administration, éducation, santé, social, transport, conseil, banque, immobilier, assurance...)

La classification pourrait encore être affinée, notamment en ce qui concerne les secteurs d’activité.

## 2.2 Difficultés rencontrées pour l'obtention des données

Les données médicales et la majorité des données économiques sont facilement accessibles grâce à la base « Eco-Santé » et à l'INSEE. Mais ces chiffres ne sont pas toujours ventilés tel que nous l'aurions souhaité : la plupart des données disponibles sont globales. Des recherches plus poussées nous ont permis de trouver des données ventilées par secteurs pour quelques indicateurs. Nous avons aussi trouvé des chiffres détaillés par taille d'entreprises et par qualification pour deux autres variables. Malheureusement, aucun indicateur n'est détaillé selon tous nos critères de différenciation. Il nous a donc été impossible de prendre en compte cette ventilation des indicateurs dans notre étude.

Autre difficulté, il n'est pas toujours facile de trouver ces données sur une durée suffisante. Cela nous oblige à envisager d'autres méthodes de modélisation, plus adaptées au cas où le nombre de données serait faible face au nombre de variables.

Les plus gros obstacles dans l'obtention de données ont été rencontrés pour les données assurantielles, c'est-à-dire les données concernant les arrêts de travail (durée, fréquence, nombre et indemnités journalières). Suite à de longues et infructueuses recherches sur internet, nous avons décidé de nous tourner vers des organismes publics.

Nous étions convaincus que ces données existaient. Un rapport de la Cour des Comptes à destination de la commission des Affaires Sociales de l'Assemblée Nationale<sup>1</sup> nous en a donné confirmation. Nous avons donc contacté ces deux institutions. Elles nous ont alors conseillé de contacter la DARES. Le secrétariat de cet organisme nous a redirigé vers le Ministère des Affaires Sociales et de la Santé.

A ce moment, nous avons compris qu'il nous fallait changer de méthode. L'administration publique est complexe et ses organisations nombreuses. Les secrétariats avec lesquels nous communiquons ne connaissent pas toujours eux-mêmes la diversité des bureaux existants. Au lieu de nous laisser guider dans un cheminement sans fin, qui nous faisait souvent revenir sur des pistes déjà approchées, nous avons mis en place une démarche positive. Nous avons dressé la liste de toutes les administrations et directions de Ministères susceptibles d'avoir les données que nous cherchions. Ensuite, nous les avons tous contactés, par mail ou par téléphone.

Certains organismes n'ont pas répondu à nos sollicitations. Citons,

- La Direction Générale de l'Offre de Soins (DGOS)
- La Direction de l'Administration Générale de la Modernisation des services (DAGEMO), organe du Ministère du Travail
- La Commission Nationale Paritaire des Techniques de Prévoyance (CNPTP)

---

1. <http://www.assemblee-nationale.fr/14/budget/mecss/cc-juillet2012.pdf>

- La DIRection des Etudes Statistiques (DIRES), organe de l'Union des CAisses Nationales de Sécurité Sociale (UCANSS)
- La Mutualité Sociale Agricole (MSA) et le Régime Social des Indépendants (RSI)

Certaines pistes se sont révélées être « des voies de garage ». Par exemple,

- La Haute Autorité de la Santé (HAS)
- La Cour des Comptes
- L'Assemblée Nationale, Commission des affaires sociales
- travailler-mieux.gouv.fr
- L'UCANSS
- La DARES, organe du Ministère du Travail
- La Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques (DREES), organe du Ministère de la Santé : Sous-direction Observation de la santé et de l'assurance maladie, bureau des professions de santé, observatoire national de la démographie des professions de santé...

Mais plusieurs ont donné lieu à une série de redirections, à chaque fois plus précises, qui nous rapprochaient de notre but :

- Le Bureau Commun des Assurances Collectives (BCAC) et B2V
- L'Institut National de Recherche et de Sécurité (INRS)
- La Société Française des Médecins Généralistes (SFMG)
- La Direction de la Sécurité Sociale (DSS), organe du Ministère de la Santé
- Le Ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social
- Le Ministère des Affaires Sociales et de la Santé
- La CNAMTS, organe de l'Assurance Maladie

C'est ainsi que nous sommes rentré en contact avec Renaud Le Gall, chef de bureau à la sous direction des dépenses de santé et des relations avec la Sécurité Sociale. Il nous a adressé à Audrey Baillot, chargée d'études au Ministère des Affaires Sociales et de la Santé. Elle-même nous a dirigé vers Jean-Philippe Perret de la CNAMTS. Le mercredi 12 février, monsieur Perret nous a fait part de son intérêt pour notre projet. Apparemment, la CNAMTS produit des études macroéconomiques en lien avec les arrêts de travail et il était curieux de l'approche que l'on pourrait avoir. Il souhaitait une demande officielle, qui lui a été transmise par monsieur Vermet dès le 13 février. Après plusieurs relances de notre part, monsieur Perret a contacté monsieur Vermet le 31 mars afin d'établir une convention, préalablement à la communication des données attendues. Suite à un nouveau long silence, nous avons recontacté monsieur Perret mi-avril puis le 30 du même mois. Nous avons finalement reçu le 7 mai un fichier ...qui ne concernait que le nombre de cas et les indemnités journalières versées sur la période 2006-2013. Ces données nous auraient été très utiles, nous permettant de traiter la totalité du champ de notre sujet. Cependant, il ne nous a pas semblé opportun de commencer une nouvelle série de modélisation à seulement une semaine de la clôture de notre projet, d'autant plus qu'elles couvrent moins de la moitié de la période d'étude que nous avons retenue (cf chapitre 3, sur la modélisation).

FIGURE 2.1 – Données de la CNAMTS : nombre de cas d'arrêts **maladie**

FIGURE 2.2 – Données de la CNAMTS : montant des indemnités journalières pour les arrêts **maladie**

Suite aux conseils de Gaël Hiénard, actuaire intervenant à l'EURIA, nous avons contacté le courtier Verlingue, par l'intermédiaire de Yann Quéré. Tout s'est fait très rapidement : le lundi 17 février nous rencontrons monsieur Quéré à l'EURIA à l'occasion de sa venue pour les L3, le 18 février nous envoyons les accords de confidentialité. La base de données anonymisée nous a été transmise le 21 février. Malheureusement, la quasi-totalité des informations disponibles ne concerne que la période 2007-2013 et n'est donc pas exploitable pour la partie modélisation. Par contre, elle pourrait être utilisée pour l'application de notre modèle à un cas concret.

Le troisième et dernier contact ayant abouti est celui que nous avons avec Roberto Wolfrum et Bénédicte Bonnanfant, de B2V<sup>2</sup>. Leur base est déjà agrégée, s'étend sur une douzaine d'années et ne comprend - comme variable assurantielle - que le nombre d'arrêts.

La suite de ce deuxième chapitre est consacrée à nos indicateurs. Pour chacun, après avoir obtenu les données, grâce à diverses sources, nous en avons fait une étude selon la même méthodologie :

- Description (définition) de l'indicateur : pour connaître précisément les données exploitées et la source de ces dernières
- Méthode de calcul : cela permet notamment de s'assurer que la méthode d'obtention des données n'a pas changé pendant la période d'études et donc que les chiffres sont cohérents dans le temps
- Justification : pour rappeler ce qui nous a motivé à garder cet indicateur pour notre étude
- Statistiques descriptives : pour commencer l'exploration des données.

Nous ne donnons les statistiques descriptives que sur la période de notre étude, c'est-à-dire 1999-2012. En effet, même si nous avons trouvé certains indicateurs sur des périodes beaucoup plus longues (à partir de 1949 pour les données sur l'investissement par exemple), 1999 (janvier) - 2012 (décembre) est le meilleur compromis que nous avons trouvé entre la durée de la période d'observation (que l'on voulait d'au moins dix ans) et l'exhaustivité de nos données.

---

2. B2V est un groupe de protection sociale, principalement pour les salariés du secteur de l'assurance et de l'enseignement privé. Il s'agit du même groupe que le BCAC



## 2.3 Données assurantielles

### 2.3.1 L'obtention des données

Ces données ont été les plus difficiles à obtenir. Organismes publics et entreprises privées tiennent secret ces chiffres, qu'ils détiennent pourtant. Nous avons approché une trentaine d'entités, principalement publiques. L'intérêt d'obtenir ces données d'administrations publiques et qu'elles soient de meilleure qualité. En effet, l'affiliation à la Sécurité Sociale est obligatoire. Il n'y a donc pas d'anti-sélection et la population assurée est nécessairement le reflet de la population totale. Les contacts ont été pris soit par mail, soit par téléphone, dans des proportions équivalentes. Nous avons souvent été redirigés d'un service à l'autre, faisant même parfois des boucles. Notre obstination a payé : trois de nos contacts ont donné suite à notre demande :

- Le courtier Verlingue, basé à Quimper, avec qui nous avons signé un accord de confidentialité. La base fournie concerne 52785 arrêts mais ne s'étale malheureusement que sur quelques années (2008-2014). Cette base est naturellement biaisée car elle ne concerne que la population d'assurés du courtier.
- B2V nous a fourni une base agrégée, sur 15 ans. Cette base est aussi naturellement biaisée. Elle l'est même d'autant plus que B2V ne prend en charge que les salariés du secteur des assurances.
- La CNAMTS a répondu à notre demande le 31 mars. Les données transmises ce jour là ne donnait que les indemnités journalières pour les accidents de travail de 1994 à 2013. Ces données ont l'intérêt d'être plus représentatives de la population puisqu'elles sont basées sur l'ensemble des salariés de France métropolitaine<sup>3</sup>. Comme expliqué dans le chapitre précédent, suite à nos nombreuses relances, la CNAMTS nous a envoyé des données début mai. Nous avons jugé que la semaine qui nous restait alors avant la clôture de ce bureau d'étude n'était pas suffisante pour pouvoir exploiter correctement ces informations, d'autant plus que la période couverte était encore insuffisante.

Ces données ne concernent évidemment que la population en activité, et même souvent des sous populations : les salariés, les fonctionnaires territoriaux, les travailleurs médicaux...

### 2.3.2 Les indicateurs

#### Le nombre d'arrêts de travail

Le **nombre d'arrêts de travail** correspond au nombre d'arrêts de travail déclarés sur une période donnée (souvent un an). Cet indicateur a besoin d'être mis en correspondance avec la population active afin de gommer l'effet dû à l'augmentation de la population.

---

3. En 2013 la France comptait 22.8 millions de salariés dans le secteur privé alors qu'un portefeuille de prévoyance chez une mutuelle ne compte que quelques milliers ou dizaines de milliers de personnes.

On se rapproche alors de la **fréquence des arrêts de travail**. Cela se divise en deux indicateurs :

- Le taux de fréquence =  $\frac{\text{nombredecas} * 1000000}{\text{nombre d'heures travaillées}}$ , le nombre d'heures travaillées correspondant au nombre d'emplois équivalent temps plein multiplié par le nombre d'heures du temps plein (35h).
- L'indice de fréquence =  $\frac{\text{nombredecas} * 1000}{\text{effectif salarié}}$

	plage	min	max	moyenne	médiane	ET
Nb cas	1999-2012	2.66	9.18	6.50	6.97	1.98
Année		dec 1999	jan 2011			
Fréquence	1999-2012	101.33	323.73	234.56	252.90	66.44
Année		dec 99	jan 2011			

TABLE 2.1 – Statistiques descriptives pour le nombre de cas par mois (en milliers) et la fréquence des accidents du travail

### La durée des arrêts de travail

La **durée des arrêts de travail** correspond à :

- la **somme de la durée de chaque arrêt de travail** sur une période donnée (souvent un mois ou un an).
- la **durée moyenne des arrêts de travail** sur une période donnée (souvent un mois ou un an).

Cet indicateur reflète naturellement les arrêts de travail.

	plage	min	max	moyenne	médiane	ET
nb jours	1999-2012	2.52	4.15	3.526	3.529	0.47
Année		1999	2011			
Durée moyenne	1999-2012	1.91	2.7	2.39	2.38	0.23
Année		1999	2012			

TABLE 2.2 – Statistiques descriptives pour le nombre de jours d'accidents du travail (en millions) et le nombre de jours moyen d'un arrêt

## 2.4 Données globales

### 2.4.1 Données économiques et financières

#### Le PIB (INSEE)

Le **Produit Intérieur Brut (PIB)** permet de mesurer la croissance économique d'un pays, par le biais de son activité économique interne sur une période donnée (souvent un trimestre, un semestre ou une année). Il est calculé en ajoutant les valeurs des biens et des services produits sur un territoire donné (quelque soit la nationalité des producteurs). On ajoute à cette somme le montant des impôts directs et on la diminue du montant des subventions. Ici nous

utiliserons le PIB annuel français, qui est un indicateur classique pour mesurer la croissance du pays. Le PIB français est de l'ordre de 2032 milliards d'euros. Il a augmenté de 85% en 20 ans et a été multiplié par 35 entre 1962 et 2012.

plage	min	max	moyenne	médiane	ET
1999-2012	1367005	2032297	1756130	1794913	207387.1
Année	1999	2012			

TABLE 2.3 – Statistiques descriptives pour le PIB en millions d'euros

### L'inflation (France-inflation)

L'**inflation** désigne la hausse durable des prix moyens sur tous les biens et services. C'est une perte de valeur de la monnaie, au sens où, avec une même quantité de monnaie, le pouvoir d'achat dont on bénéficie diminue.

L'inflation se mesure grâce aux variations de l'**indice des prix à la consommation** (IPC). Cet indice est la moyenne pondérée (en fonction des quantités consommées) des prix d'un panier de 265 postes (« produits »). Le taux d'inflation annuel peut être calculé en moyenne (comparaison des moyennes annuelles de l'indice) ou en glissement (comparaison des prix de décembre des deux années).

Nous avons choisi d'utiliser l'**indice des prix à la consommation harmonisés**, qui n'est autre qu'un indice des prix dont le panier de biens et de services de référence est le même pour tous les pays de l'Union Européenne.

plage	min	max	moyenne	médiane	ET
1999-2012	0.0083%	0.23%	0.138%	0.145%	0.054%
Année	2009	2008			

TABLE 2.4 – Statistiques descriptives pour l'inflation

### Les taux bancaires (Banque de France et INSEE)

Il ne s'agira pas ici du taux de base bancaire, qui diffère d'une banque à l'autre et qui permet le calcul des taux de prêts aux particuliers et aux professionnels. Nous analyserons plutôt les taux du marché monétaire, auxquels les banques se prêtent entre elles. Nous avons choisi trois taux, dont deux du marché monétaire européen :

- L'**Euro Overnight Index Average (EONIA)** est le taux d'intérêt interbancaire de la zone euro avec une échéance un jour. Il est calculé avec une base de 360. Chaque banque de l'échantillon déclare le taux auquel elle est prête à prêter jusqu'au lendemain ouvré. Les cotations extrêmes sont enlevées et une moyenne est faite avec les données restantes.
- L'**Euro Interbank Offered Rate (EURIBOR)** peut avoir plusieurs échéances (de une semaine à 12 mois). Son mode de calcul est le même

que celui de l'EONIA. Le plus connu est l'EURIBOR 3 mois (E3M, qui sert de base au marché des swaps).

- Le **Taux Moyen des Emprunts d'Etat (TME)**. Il s'agit du taux de rendement moyen sur le marché secondaire des emprunts et des obligations assimilables au Trésor émis par l'Etat français, à taux fixe et pour une durée supérieure à 7 ans. Il est publié par la Caisse des Dépôts et des Consignations, mensuellement. Les banques l'utilisent pour déterminer leurs taux fixes à moyen et à long terme.

Les taux reflètent souvent assez fidèlement la situation économique d'une zone grâce aux indications que cela donne sur la confiance des investisseurs.

	plage	min	max	moyenne	médiane	ET
EONIA	1999-2012	0.07%	5.14%	2.42%	2.47%	1.42
		dec 2012	avr 2001			
E3M	1999-2012	0.19%	5.13%	2.68%	2.59%	1.38
		dec 2012	oct 2008			
TME	1999-2012	2.03%	5.71%	4.10%	4.07%	0.81
		dec 2012	janv 2000			

TABLE 2.5 – Statistiques descriptives pour les taux bancaires

### Les indices boursiers (Yahoo finance)

Les **indices boursiers** peuvent être un indicateur de l'activité et aussi et surtout celui des résultats d'une entreprise. Il faut faire attention, avec cet indice, à ne pas se limiter à une catégorie d'entreprises. Du fait de notre étude, basée sur le modèle français, il nous faut un indice boursier français. Le CAC40 est le plus connu d'entre eux mais il ne convient pas à notre étude, du moins pas pris seul. En effet cet indice ne prend en compte que de très grosses entreprises. Nous avons donc choisi deux :

- le SBF 120. Il comprend 120 valeurs parmi les 200 premières capitalisations boursières de la Bourse de Paris). Sa composition est réétudiée tous les ans par Euronext.
- le CAC mid& small 190 comprend 190 dernières valeurs parmi les petites capitalisations boursières (4.6 milliards d'euros pour la plus grosse capitalisation de l'indice) comme JC Decaux, PagesJaunes, Ubisoft, April... Il est coté à Paris.

	plage	min	max	moyenne	médiane	ET
SBF 120	1999-2012	1843.5	4506.9	3063.4	2915.5	670.9
Année		mars 03	mars 00			
CAC	1999-2012	2574.7	8961.2	5328.7	4997.9	1647.7
Année		mars 03	mai 07			

TABLE 2.6 – Statistiques descriptives pour les indices boursiers

### Le cours de l'euro dollar (BCE, via l'INSEE)

Appelé aussi euro/dollar, il s'agit du nombre de dollars qui « équivaut » à un euro. Sur les marchés monétaires, il y a deux cotations pour un même couple de monnaies : le cours vendeur et le cours acheteur. La BCE publie quotidiennement la moyenne de ces deux cotations, à titre indicatif, pour l'euro par rapport à 34 autres monnaies.

C'est un instrument financier très échangé sur les marchés. Ce taux de change influe beaucoup sur les échanges internationaux : Si l'euro est fort, les exportations sont pénalisées et inversement, favorisées s'il est faible.

plage	min	max	moyenne	médiane	ET
1999-2012	0.84	1.58	1.21	1.25	0.19
Année	oct 2000	mar 2008			

TABLE 2.7 – Statistiques descriptives pour le cours de l'euro dollar

### Le cours du baril de pétrole BRENT (INSEE)

Le pétrole Brent est issu de la mer du Nord. Il est coté en dollars, à la Bourse de Londres. Il sert de prix de référence pour le pétrole au niveau mondial, au même titre que le WTI (coté à New York).

Le cours du pétrole suit généralement les cycles économiques. Par exemple, les cours ont doublé entre 2003 et 2008, pour s'effondrer en 2008-2009, lors de la crise financière.

plage	min	max	moyenne	médiane	ET
1999-2012	10.30	133.4	58.26	53.95	33.15
Année	fev 1999	juil 2008			

TABLE 2.8 – Statistiques descriptives pour le cours du baril de pétrole BRENT en dollars US

### La balance des paiements (Banque de France)

La **balance des paiements** correspond au solde entre les flux sortants et entrants d'un pays.

On considère à la fois les flux d'actifs réels (biens et services) et les flux monétaires et financiers. On y distingue d'ailleurs trois comptes : le compte des transactions courantes, le compte de capital et le compte financier.

On s'intéressera ici au **solde des transactions courantes** de la France, qui est le plus fréquemment cité des trois. Cet indicateur permet d'avoir une vision globale des échanges d'un pays avec le reste du monde.

plage	min	max	moyenne	médiane	ET
1999-2012	-2810	3585	899	1145	1767
Année	2012	2003			

TABLE 2.9 – Statistiques descriptives pour la balance des paiements, en millions d’euros, sur données mensualisées

### L’endettement des ménages (Banque de France)

Un ménage est un ensemble de personnes partageant un même logement.

L’endettement des ménages peut correspondre :

- à la proportion des ménages qui sont endettés
- au taux d’endettement moyen des ménages, c’est à dire à la moyenne du rapport entre la charge financière de l’emprunt et le revenu disponible.
- à l’encours de la dette des ménages : c’est cette définition que nous utilisons.

L’activité économique d’un pays est toujours liée à la consommation des ménages et à leur endettement. C’est d’ailleurs l’un des indicateurs qui a été étudié pour suivre l’évolution de la crise.

plage	min	max	moyenne	médiane	ET
1999 - 2012	432	1126	753.7	726	228.8
Année	fev 1999	dec 2012			

TABLE 2.10 – Statistiques descriptives pour l’endettement des ménages en milliards d’euros

### La confiance des ménages (INSEE)

La **confiance des ménages** est un indicateur synthétique regroupant l’opinion des ménages sur :

- la situation financière passée et future
- la capacité d’épargne actuelle et future
- l’opportunité d’épargner
- l’opportunité d’engager des dépenses importantes
- le niveau de vie
- les perspectives d’évolution du chômage
- les prix...

Il est calculé suite à une enquête téléphonique faite auprès de 2000 ménages. Les résultats sont toujours présentés corrigés des variations saisonnières. Cet indicateur permet d’estimer la consommation future des ménages. Il permet donc d’avoir une idée sur l’activité économique future. Nous sommes en ce moment en dessous de la moyenne (calculée sur la période 1987-2013).

plage	min	max	moyenne	médiane	ET
1999-2012	81	124	98.7	98	11.6
Année	06,07,08,10/08	janv 2001			

TABLE 2.11 – Statistiques descriptives pour la confiance des ménages

## 2.4.2 Données démographiques

### La population française (INSEE)

La **population française** correspond au nombre de personnes ayant leur résidence habituelle en France auquel on ajoute le nombre de sans abris. Les données sont produites par l'INSEE dans le cadre des recensements de la population, strictement encadrés par la loi. Dans le cadre de notre étude, cette donnée nous servira principalement à relativiser d'autres indicateurs, pour les ramener (s'il le faut) à la population totale.

plage	min	max	moyenne	médiane	ET
1999-2012	58.50	63.68	61.24	61.38	1.55
Année	1999	2012			

TABLE 2.12 – Statistiques descriptives pour la population française en millions d'habitants

### Les taux de natalité et de mortalité (INSEE)

Le **taux de natalité** correspond à la division du nombre de naissances vivantes sur un an par la population moyenne de l'année. Les enfants morts-nés ne sont donc pas pris en compte.

Certaines théories lient natalité et activité économique : au début d'une période de forte croissance, le taux de natalité augmente, comme ce fut le cas pendant les Trente Glorieuses et le baby boom.

Le **taux de mortalité** correspond au rapport entre le nombre de décès sur un an et la population moyenne de l'année.

La mortalité peut refléter l'accès aux soins : plus la mortalité est faible, plus l'accès aux soins est performant. Ces taux sont généralement exprimés en pour mille.

	plage	min	max	moyenne	médiane	ET
Natalité	1999-2012	12.6‰	13.3‰	12.9‰	12.9‰	0.19
Année		2011	2000			
Mortalité	1999-2012	8.3‰	9.1‰	8.6‰	8.5‰	0.29
Année		2004	2003			

TABLE 2.13 – Statistiques descriptives pour les taux de natalité et de mortalité

### 2.4.3 Données médicales

#### Le nombre de médecins généralistes (Eco-Santé - DRESS)

Le **nombre de médecins généralistes** est un bon indicateur de l'accès au soin de la population. Il en est plus représentatif que le nombre de spécialistes, notamment depuis l'existence du « parcours de soins coordonné », qui oblige les patients à s'adresser à leur médecin traitant avant d'aller consulter un spécialiste. Pour gommer les variations liées à des modifications du nombre d'habitants, nous pouvons considérer le nombre de médecins généralistes pour 100000 habitants.

plage	min	max	moyenne	médiane	ET
1999 - 2011	193	217	205.3	207.2	6.8
Année	1999	2011			

TABLE 2.14 – Statistiques descriptives pour le nombre de médecins généralistes en milliers

#### La consommation médicale (Eco-Santé - DRESS)

Le **nombre de consultations chez le médecin généraliste** aurait permis de connaître le suivi médical effectif de la population. En effet, c'est le médecin généraliste qui délivre le plus souvent les arrêts de travail. Nous n'avons pas trouvé de données sur cet indicateur donc nous en avons cherché un autre qui l'engloberait.

Nous utiliserons donc la **consommation médicale totale, par personne, en euros relatifs**. Le fait de prendre les données par personne et pas les données globales permet de gommer l'effet dû à l'évolution démographique. Enfin, nous avons choisi les chiffres en euros relatifs et non en euros courants afin de gommer l'effet dû à l'évolution globale des prix.

plage	min	max	moyenne	médiane	ET
2000-2012	2087.3	2545.6	2382.3	2420.1	144.1
Année	2000	2012			

TABLE 2.15 – Statistiques descriptives pour la consommation médicale annuelle, par personne, en euros relatifs



## 2.5 Données liées aux entreprises

### 2.5.1 Données contractuelles

#### L'emploi à temps partiel (INSEE)

Le **taux d'emploi à temps partiel** correspond au rapport entre le nombre de personnes travaillant à temps partiel et le nombre d'actifs.

De la même manière, on peut calculer le **taux d'emploi à temps plein**. On parle de travail à temps partiel lorsque le temps de travail prévu par le contrat de travail est inférieur au minimum entre la durée légale du travail (35 heures) et la durée conventionnelle. Nous n'avons pas trouvé de série longue sur le taux d'emploi à temps partiel.

Nous exploiterons donc les écarts entre deux indicateurs : L'**emploi total en équivalent temps plein** et l'**emploi total (en nombre de salariés)**. Cet indicateur a un double intérêt :

- on peut imaginer qu'en période de crise, le recours au temps partiel est plus important qu'en période de grande croissance.
- un employé à temps partiel a moins l'habitude de son poste et est moins impliqué dans l'entreprise. Il peut donc avoir plus d'accidents du travail et d'arrêts maladie que les autres.

plage	min	max	moyenne	médiane	ET
1999-2012	2014.3	3130.0	2534.5	2457.7	348.7
Année	avr 2001	oct, nov, dec 12			

TABLE 2.16 – Statistiques descriptives pour l'emploi à temps partiel exprimés en milliers

#### Le salaire moyen relatif (INSEE)

Le **salaire moyen relatif** correspond à la moyenne des salaires.

Il aurait été préférable de considérer les salaires bruts car les retenues conduisant au salaire net ne sont pas toujours les mêmes (cf retraite complémentaire, mutuelle obligatoire d'entreprise). Cependant, nous n'avons trouvé que les données concernant les salaires nets. Nous disposons d'une ventilation selon la qualification des salariés pour la période 1950-2011 et d'une ventilation selon la taille de l'entreprise et le secteur d'activité de 1995 à 2010. Cet indicateur nous intéresse car il reflète l'activité économique du pays. En outre, des études ont déjà montré que, plus les salaires étaient élevés, moins il y avait d'arrêts de travail.

plage	min	max	moyenne	médiane	ET
1999-2012	1636	2190	1899.37	1843	175.6
Année	1999	2012			

TABLE 2.17 – Statistiques descriptives pour le salaire moyen relatif

## L'indice du coût du travail (INSEE)

L'indice du coût du travail est publié, par secteurs, chaque trimestre même si les données sont mensuelles. Son but est de refléter l'évolution du coût horaire du travail, en prenant en compte les salaires et les charges. La méthode de calcul de cet indice change souvent :

- Le secteur public est pris en compte depuis 2009 uniquement.
- Cet indice prend en compte la plupart des crédits d'impôts liés à l'emploi. Or ces crédits changent souvent ce qui rend difficile le suivi de ces données dans le temps.
- L'ICHT-TS, indice du coût du travail de 1997 à 2004, ne tenait pas compte des modifications de rémunération dus à des changements de qualification des emplois

Cependant, l'indicateur actuel, l'ICT, que nous étudions ici à partir de 1998 a été révisé par l'INSEE à la demande d'EUROSTAT afin d'être cohérent sur toute la période.

L'utilisation de cet indice nous permettra de prendre en compte l'évolution des charges patronales, mais surtout il est considéré comme un indicateur primordial pour mesurer l'attractivité d'une économie pour les investisseurs et donc sa dynamique potentielle.

plage	min	max	moyenne	médiane	ET
1999-2012	72.7	111.5	92.1	92.6	11.4
Année	jan, fev 99	dec 2012			

TABLE 2.18 – Statistiques descriptives pour l'indice du coût du travail

## 2.5.2 Données sectorielles

### La santé financière des entreprises (INSEE)

La **santé financière d'un secteur** est un des indicateurs permettant de juger de son dynamisme.

On peut notamment la mesurer grâce au nombre de **défaillances d'entreprises** dans le secteur considéré divisé par le nombre total d'entreprises exerçant dans ce domaine.

Il est évident qu'en période de crise, le nombre d'entreprises faisant faillites sera plus important qu'en période de croissance.

plage	min	max	moyenne	médiane	ET
1999-2012	1572	5681	4318.4	4168	622.6
Année	mars 2001	dec 2012			

TABLE 2.19 – Statistiques descriptives pour le nombre de défaillances d'entreprises

### L'investissement des entreprises (INSEE)

L'**investissement** des entreprises non financières correspond aux dépenses d'argent effectuées par ces entreprises dans l'optique d'en gagner plus. Il s'agit ainsi d'un achat de biens et de services qui serviront au cours de l'activité de production. Le **taux d'investissement** correspond au rapport entre le montant investi et la valeur ajoutée (cf compte de résultat) L'investissement des entreprises est un reflet de la santé de l'économie nationale. En effet, les entreprises investissent plus lorsque l'activité économique est importante.

	plage	min	max	moyenne	médiane	ET
M. invest.	1999-2012	10.7	18.2	14.7	14.5	2.2
Année		ja-mr 99	ja-mr 08			
Tx invest.	1999-2012	0.275%	0.290%	0.282%	0.283%	0.004%
Année		2003	2008			

TABLE 2.20 – Statistiques descriptives pour l'investissement des entreprises (en milliards d'euros pour le montant)

### Le taux de marge (INSEE)

Le **taux de marge** est le pourcentage de gain ou de perte réalisée par une entreprise ou, dans notre cas, d'un ensemble d'entreprises.

Il est calculé en divisant l'Excédent Brut d'Exploitation (EBE) par la valeur ajoutée. L'EBE est égale à la valeur ajoutée à laquelle on soustrait la rémunération des salariés, des autres impôts et on ajoute le montant des subventions. Lorsque l'activité économique est fleurissante, les entreprises ont généralement des marges plus importantes.

plage	min	max	moyenne	médiane	ET
1999-2012	0.3254%	0.3372%	0.3324%	0.3333%	0.0032
Année	2012	2007			

TABLE 2.21 – Statistiques descriptives pour le taux de marge

### Le chiffre d'affaires par secteur (INSEE)

Le **chiffre d'affaires du secteur** permet de chiffrer le volume d'activités réalisées par un secteur. On utilise un **indice de chiffre d'affaires** pour avoir des données chiffrées pour cet indicateur. L'INSEE fournit ces statistiques en partenariat avec la direction générale des Impôts, en utilisant les déclarations de TVA des entreprises.

Les indices mensuels de chiffre d'affaire base 100 en 1995 sont calculés à partir de l'ensemble des entreprises du secteur marchand non agricole et non financier actives au 1<sup>er</sup> de l'année n.

L'indice de l'année n est obtenu en multipliant les maillons des années 95 à n, le maillon de l'année n étant le rapport du montant total du chiffre d'affaire de l'année n sur le montant total du chiffre d'affaire de l'année n-1 (le maillon de l'année 1995 vaut donc 100).

$$indice - annuel - annee_n = \frac{[CA(n)]^{ech(n)}}{[CA(n-1)]^{ech(n)}} * \dots * \frac{[CA(96)]^{ech(96)}}{[CA(95)]^{ech(96)}} * 100 \quad (2.1)$$

avec  $[CA(n)]^{ech(n)}$  le montant total du chiffre d'affaire de l'année n calculé à partir des entreprises de l'échantillon de l'année n.

Le principe du calcul de l'indice de chiffre d'affaire du mois m de l'année n consiste à multiplier l'indice de l'année n par le profil du mois m de l'année n. Le profil du mois m de l'année n est le rapport du montant total du chiffre d'affaire du mois m de l'année n sur 1/12 du montant total du chiffre d'affaire de l'année n.

Le chiffre d'affaires est l'indicateur naturel de l'activité d'une entreprise, et donc, par extension, de l'activité économique.

plage	min	max	moyenne	médiane	ET
1999-2012	67.01	105.97	90.09	91.40	11.26
Année	mars 1995	fev 2013			

TABLE 2.22 – Statistiques descriptives pour l'indicateur de chiffre d'affaires par secteur

### L'indice de confiance (INSEE)

L'indice de confiance, pour les entreprises, se subdivise en deux indicateurs :

- L'indice manufacturier. Il est calculé à partir par l'INSEE à partir de données collectées conjointement par l'INSEE, certains Ministères et les organismes professionnels.
- L'indice du climat des affaires. Il s'agit d'une sorte de sondage effectué mensuellement (jusqu'en 2007, il y avait une exception pendant le mois d'août) par voie postale ou par internet auprès des professionnels du secteur des services.

La confiance des entreprises baisse en période de crise. Elle augmente avant ou au moment d'un début de reprise.

	plage	min	max	moyenne	médiane	ET
Manufacturier	1999-2012	92.8	120.9	110.33	115	7.93
Année	avr 2009	fev 2009				
Affaires	1999-2012	70	118	101.73	104	10.34
Année	mars, av 99	dec 1999				

TABLE 2.23 – Statistiques descriptives pour l'indice de confiance

### Le chômage (Ministère du Travail)

Le **chômage** correspond au nombre (ou au taux) de demandeurs d'emplois dans un pays. On distingue 5 catégories de demandeurs d'emploi : notés de A à E. Nous nous intéressons ici au nombre de demandeurs d'emploi des catégories A, B et C : ce sont ceux obligés de chercher réellement un emploi. Les catégories D et E font référence aux demandeurs d'emploi qui ne sont pas disponibles de suite (et donc non tenus de chercher activement un travail) : personnes malades, en stage, en formation, en emploi....

En période de crise, le chômage augmente : il reflète donc l'activité économique. Une étude de 2012<sup>4</sup> assure que si le taux de chômage est élevé, le nombre d'arrêts de travail sera plus élevé car les salariés qui auraient envisagé de démissionner préféreront prendre des congés maladie.

plage	min	max	moyenne	médiane	ET
1999-2012	3056	4619	3700	3714	355.5
Année	mai 2008	dec 2012			

TABLE 2.24 – Statistiques descriptives pour le chômage en milliers

### Le nombre de travailleurs (INSEE)

Le **nombre de travailleurs** ou population active occupée correspond au nombre de personnes (de plus de 15 ans), qu'elles soient salariées, aides, professions libérales, stagiaires rémunérés...), travaillant dans chaque secteur et pour chaque taille d'activité (de l'auto-entrepreneur aux plus grandes entreprises). On l'exprime en « équivalent temps-plein » pour qu'il n'y ait pas de biais du à un recours au temps partiel plus important dans certains secteurs. La population active regroupe la population active occupée (c'est-à-dire ayant effectivement une activité) et les chômeurs.

Nous avons gardé comme indicateurs l'**emploi total** en nombre de personnes et en équivalent temps plein. Ces données auront pour nous une utilité similaire à celles de la population totale : permettre de relativiser les informations fournies par d'autres indicateurs en les ramenant à la population travaillant ou au

4. Etude de l'Institut de Recherche et de Documentation en Économie de la Santé (IRDES), intitulée « Arrêts maladie : comment expliquer les disparités départementales ». Cette étude s'appuie elle-même sur les travaux de Barmby et Treble, publiés en 1991 : « Worker Absenteeism : An Analysis Using Microdata »

nombre d'heures travaillées.

plage	min	max	moyenne	médiane	ET
1999-2012	25984.0	28566.0	27486.1	27535.4	786.7
Année	janv 99	dec 2012			

TABLE 2.25 – Statistiques descriptives pour la population active en milliers

### L'endettement des entreprises non financières (Banque de France)

Le **taux d'endettement des entreprises** reflète la dette d'une entreprise par rapport à ses fonds propres. Les dettes prises en compte sont notamment les emprunts bancaires et obligataires, les avances reçues de la part des associés, les crédits-bails en cours, les découverts sur les comptes courants. . . On peut imaginer qu'en période de crise, la dette des entreprises augmente car elles doivent faire face à des difficultés supplémentaires. Cependant, on peut aussi imaginer qu'en période de reprise économique, la dette augmente du fait de nouveau investissement non encore rentabilisés.

plage	min	max	moyenne	médiane	ET
1999-2012	434.7	820.7	628.7	581.2	128.1
Année	mars 99	juin 2012			

TABLE 2.26 – Statistiques descriptives pour l'endettement des entreprises non financières en milliards d'euros

## 2.6 Corrélation entre nos variables à expliquer et nos variables explicatives

Après examen de ces résultats (visibles dans le tableau 2.27), il apparaît que les variables les plus corrélées sont la consommation médicale, la population active, la population totale, le PIB, le nombre de médecin, l'indice coût du travail, le chiffre d'affaire et le salaire moyen. Il semble évident que la consommation médicale soit le reflet fidèle de la fréquence ou le nombre de jours d'arrêt de travail. En outre, la présence accrue de médecin aura pour effet de consulter plus facilement et donc d'être en arrêt de travail plus souvent et plus vite. L'augmentation de la population totale ou de la population active a bien évidemment pour conséquence une augmentation du nombre de jours d'arrêt de travail par « l'effet nombre » : la variable à expliquer augmente car la variable explicative augmente. Enfin, les indicateurs économiques cités dans cette liste donnent une image fidèle de la conjoncture. Nous pouvons noter que les variables les moins corrélées sont le chômage, l'inflation, l'indice boursier SBF 120 et la production industrielle dans la construction.

Nous remarquons que globalement pour une même variable explicative les coefficients de corrélation obtenus sont similaires pour les quatre variables à expliquer

	INDFREQ	NBC	NBJ	NBJM
PIB	0.9731	0.9781	0.9421	0.9560
NAT	-0.6613	-0.6637	-0.6508	-0.6631
MORT	-0.6790	-0.6719	-0.5872	-0.5620
TXMARGE	-0.4689	-0.4816	-0.4686	-0.5030
TXINVEST	0.4582	0.4750	0.4876	0.5012
INDCOUTRAVAIL	0.9711	0.9786	0.9491	0.9662
TEMPLEIN	0.9132	0.9080	0.9156	0.8996
INVESTENT	0.8811	0.8891	0.8783	0.8882
CONFMEN	-0.9148	-0.9163	-0.8916	-0.9006
ENDENT	0.8956	0.9090	0.9071	0.9340
ENDMEN	0.9347	0.9466	0.9187	0.9460
CLIMAFFAIRES	-0.4129	-0.4176	-0.4658	-0.4846
PRODINDMAN	-0.6518	-0.6686	-0.6514	-0.7012
PRODINDCONST	0.3682	0.3522	0.3457	0.3096
DEFAILLENT	0.7187	0.7343	0.7079	0.7532
EURDOL	0.8099	0.8132	0.7697	0.7719
EONIA	-0.6531	-0.6621	-0.6040	-0.6479
E3M	-0.5709	-0.5768	-0.5183	-0.5561
CA	0.9639	0.9681	0.9502	0.9569
POPACT	0.9821	0.9873	0.9619	0.9753
BRENT	0.8384	0.8494	0.8220	0.8375
POPTOT	0.9768	0.9833	0.9524	0.9692
SBF	-0.3210	-0.3190	-0.3739	-0.3806
INFLATION	0.2119	0.1991	0.3004	0.2217
MIDSMALL	0.6615	0.6663	0.5855	0.5999
CHOM	0.2373	0.2510	0.1800	0.2223
TME	-0.7955	-0.8004	-0.7167	-0.7471
CONSOMED	0.9911	0.9908	0.9632	0.9661
NBMED	0.9717	0.9749	0.9397	0.9486
TEMPARTIEL	0.59838	0.6192	0.5485	0.6070
BALPAI	-0.6149	-0.6410	-0.5900	-0.6461
SALMOYENENS	0.9573	0.9667	0.9415	0.9591

TABLE 2.27 – Corrélation entre les variables à expliquer (données assurantielles) et les variables explicatives

## Chapitre 3

# Modélisation mathématique

Premièrement, ce chapitre énonce les limites que la réalité du terrain et la disponibilité des données a imposé à notre étude. Deuxièmement, nous présentons les démarches ayant permis de charger les données sur le logiciel R et de les mensualiser. Un troisième point traite des liens entre nos variables. Quatrièmement, nous utilisons une méthode de modélisation, notamment la régression linéaire. Finalement, nous présentons nos résultats et nous les discutons.

Les différentes corrélations que nous avons vues précédemment nous permettent d'affirmer qu'il y a effectivement une corrélation entre la sinistralité en arrêt de travail et les variables économiques. Nous allons mettre en oeuvre un modèle linéaire afin de trouver les influences des différentes variables explicatives sur nos données assurantielles.

Nous rappelons que le but de notre travail est d'obtenir des variables explicatives qui expliquent correctement nos variables à expliquer avec un modèle du type :

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_p X_{i,p} + W_i \quad (3.1)$$

avec  $W_i$  iid  $\forall i \in [1, n]$  tels que  $E[W_i] = 0$  et  $var(W_i) = \sigma^2 < \infty$  et avec  $(\beta_0, \dots, \beta_p)$  des paramètres inconnus et avec  $i \in [1, 4]$ , pour nos quatre variables à expliquer et  $p$  le nombre de variables explicatives.

En notant  $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ ;  $W = \begin{pmatrix} W_1 \\ \vdots \\ W_n \end{pmatrix}$ ;  $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$ ;  $X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ \vdots & & \ddots & \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix}$ ,  
le modèle de régression linéaire multiple s'écrit :  $Y = X\beta + W$

### 3.1 Les limitations imposées à notre étude

Au début de ce projet, nous avons envisagé de ventiler au maximum notre projet afin de prendre en compte les disparités liées au secteur, à la taille de



l'entreprise ou encore au statut des salariés. Nous espérons ainsi :

- avoir des résultats plus précis et surtout plus détaillés.
- déterminer les segments les plus sensibles au risque arrêt de travail.

Malheureusement, nous n'avons pu trouvé qu'une minorité d'indicateurs ainsi détaillés. Nous avons donc renoncé à ventiler notre étude.

La deuxième limitation que nous avons dû nous imposer concerne la notion même d'arrêt de travail. En effet, à moins de deux mois de la conclusion du projet, pour nos variables à expliquer, nous ne disposions que :

- D'une base de données concentrée sur cinq années, concernant la population d'assurés de Verlingue et ayant l'avantage de concerner à la fois les arrêts maladie simples et les accidents du travail.
- D'une base de données sur 14 ans, concernant la population d'assurés de B2V et concernant à la fois les arrêts maladie simples et les accidents du travail. Cette base ne comprend que le nombre d'arrêts, les autres colonnes du tableau comprenant des informations sur les salariés (âge, genre, salaire).
- De bases délivrées par la CNAMTS et disponibles sur internet. Ces bases comprennent l'ensemble des variables à expliquer que nous avons imaginées mais sont limitées au seul cas des accidents du travail.

Le peu de temps qui nous restait pour achever notre étude nous empêchait d'attendre davantage que notre contact à la CNAMTS nous envoie les données concernant les arrêts maladie simples. Nous travaillons donc sur les seuls accidents du travail.

## 3.2 Le chargement des données avec R

Nous avons choisi de travailler avec le logiciel R. Il a donc été nécessaire de mettre l'ensemble de nos fichiers de données en forme afin de pouvoir les exporter ensuite.

Il a aussi fallu pour cela :

- enlever les en-têtes, systématiquement présentes dans les fichiers de données de l'INSEE.
- éventuellement mettre toutes les données présentes sur la même feuille de calcul. En effet, les fichiers CSV ne gardent que la première feuille de calcul du fichier XLSX.
- faire parfois quelques calculs, comme sommer ou diviser plusieurs colonnes. Ces opérations ne prennent que quelques secondes sur Excel. Le but est de rendre plusieurs variables comparables entre elles sur une même période. Ceci consiste, par exemple, à diviser certains indicateurs (nombre d'intérimaires. . .) par la population active pour gommer l'effet de l'augmentation démographique.
- enlever les spécifications techniques pour ne garder que les données chiffrées et leurs libellés.
- mettre chaque fichier ainsi obtenu au format CSV.

### 3.3 La mensualisation des données

Nous avons adapté les données en ne sélectionnant que la même période, ici de janvier 1999 à décembre 2012, pour chacune d’entre elles. Nous les avons également ajustées pour qu’elles soient exploitables en même temps : les données trimestrielles ou annuelles ont été mensualisées. La méthode utilisée pour cela a été différente selon le type de données.

#### 3.3.1 Interpolation

Cette méthode de mensualisation concerne par exemple des indices tels que les indicateurs du coût du travail, la population ou le PIB. Pour ce faire, nous avons utilisé Excel et la fonction « prévision », qui permet d’interpoler entre deux dates.

Date	Valeurs	Utilisation de la fct
01/01/1999	88	
01/02/1999	?	90
01/03/1999	92	

TABLE 3.1 – Exemple d’interpolation

#### 3.3.2 Répartition

Certains indicateurs donnent une quantité sur une période, comme c’est le cas pour l’investissement des sociétés non financières ou l’inflation. Ainsi, si nous voulons obtenir un chiffre pour une période plus courte, il suffit de faire une division (par exemple, diviser par 3 s’il l’on veut passer d’une donnée trimestrielle à une donnée mensuelle).

#### 3.3.3 Actualisation du taux

Nous avons mensualisé les taux (natalité, mortalité) qui étaient sous format annuel à l’aide de la formule :

$$tauxmensuel = (1 + tauxannuel)^{1/12} - 1 \quad (3.2)$$

#### 3.3.4 Duplication

Cette méthode est celle qui demande le moins de travail. En effet, elle consiste à recopier la même valeur sur toute la période. Nous l’avons notamment utilisée pour l’emploi total ou encore le salaire moyen relatif.

FIGURE 3.1 – ACP avec l'ensemble de nos variables explicatives

FIGURE 3.2 – ACP avec l'ensemble de nos variables explicatives

## 3.4 Analyse en composantes principales (ACP) et corrélations

### 3.4.1 Analyse en composantes principales

L'Analyse en Composantes Principales (ACP) est une méthode d'analyse de données. Il est parfois difficile de mettre en évidence les relations existant entre un nombre important de variables. L'ACP permet de mieux visualiser ces liens. En effet, en réalisant une ACP, on va condenser l'information présente dans les données afin d'en retirer les principaux liens entre variables et entre individus. Bien sûr, cela se fait avec une perte d'informations minimum. Le principe de l'ACP est donc de déterminer un sous espace de dimension  $q$  (avec  $q$  petit et inférieur au nombre de variables) sur lequel on projette les données afin que le graphique-résultat soit lisible et déforme le moins possible la réalité des chiffres.

Le nombre important de nos variables rend difficile la lecture des graphiques produits par l'ACP. L'utilisation de cette méthode avec l'ensemble de nos indicateurs nous a toutefois permis de visualiser la tendance commune à nos variables.

Nous utilisons de nouveau l'ACP après la régression, lorsque nous n'avons plus qu'une quinzaine de variables.

### 3.4.2 Corrélations

Nous avons dans un premier temps pris en compte la variable temps en étudiant les différentes corrélations croisées qu'il pouvait y avoir entre nos données qui serviront de variables explicatives et celles qui serviront de variables à expliquer. Cela nous a permis de voir si certaines variables à expliquer agissaient ou non avec un décalage dans le temps. Par exemple, le chômage de juin 2005 pourrait agir sur la sinistralité en arrêt de travail de février 2006. Mais nous n'avons pas réussi à concrétiser ces décalages dans notre modélisation.

Au début de notre travail, nous avons également songé à faire une première sélection de variables à l'aide de la corrélation. Par exemple, ne garder que les variables corrélées à plus de 80% avec nos variables à expliquer. Cependant, nous n'avons pas su quel seuil (0.8, 0.9, 0.95 etc.) choisir et avons décidé de garder toutes nos variables explicatives pour la régression.

FIGURE 3.3 – Corrélation croisée entre le chômage et le nombre de cas d'accidents du travail

## 3.5 Régressions

### 3.5.1 Notions théoriques

#### La p-value

La p-value nous donne le résultat d'un test que voici :

$$H_0 : \beta_i = 0 \text{ contre } H_1 : \beta_i \neq 0.$$

Ainsi, si la p-value du test est  $> 0.05$ , l'hypothèse  $H_0$  est acceptée. Si la p-value est  $\leq 0.05$ , l'hypothèse  $H_0$  est rejetée.

#### $R^2$ et $R^2$ ajusté

Soient :

$SCE = \sum_i (\hat{Y}_i - \bar{Y})^2$  : somme des carrés expliquée par la régression

$SCR = \sum_i (Y_i - \hat{Y}_i)^2$  : somme des carrés résiduelle

$SCT = \sum_i (Y_i - \bar{Y})^2$  : somme des carrés totale

$SCT = SCR + SCE$

Nous avons le coefficient de corrélation multiple (ou  $R^2$ ) qui s'écrit alors :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \quad (3.3)$$

Il s'interprète comme le pourcentage de variance expliquée par le modèle. Un critère de qualité qui apparaît alors encore meilleur est le  $R^2$  ajusté ( $R_{aj}^2$ ). En effet, le  $R^2$  sélectionne souvent les modèles les plus complexes car on projette sur un espace plus grand. Or le  $R_{aj}^2$  tient compte de cette dimension. Il s'écrit, avec  $p$  le nombre de paramètres et  $n$  le nombre d'individus :

$$R_{aj}^2 = \frac{(n-1) * R^2 - p}{n - p - 1} \quad (3.4)$$

Cette influence de la taille de l'espace de projection est alors compensée par le  $p$  (que l'on veut le plus petit possible) qui est soustrait au  $R^2$  (que l'on veut le plus grand possible).

#### AIC

Le critère d'information d'Akaike (AIC) est défini tel que :

$$AIC = 2k - 2\ln(L) \quad (3.5)$$

avec  $k$  le nombre de paramètres et  $L$  la vraisemblance de notre modèle.

Comme on cherche un modèle avec une grande vraisemblance et un petit nombre de paramètres, le meilleur modèle sera celui avec l'AIC le plus faible. C'est cette méthode qui est utilisée avec la fonction `step` de R.

#### Estimation des $\beta$

On obtient, avec la méthode des moindres carrés :  $\hat{\beta} = (X'X)^{-1}X'Y$

### Loi du Chi2

Si  $X_i \sim N(\mu_i, \sigma_i) \forall i \in [1; k]$  alors :

$$X := \sum_{i=1}^k Y_i^2 = \sum_{i=1}^k \left( \frac{X_i - \mu_i}{\sigma_i} \right)^2 \quad (3.6)$$

suit une loi du  $\chi^2(k)$  avec  $k$  le nombre de degrés de liberté.

### Loi de Student

Si on considère  $Z \sim N(0, 1)$  et  $U \sim \chi^2(k)$  alors :

$$T = \frac{Z}{\sqrt{U/k}} \quad (3.7)$$

suit une loi de Student à  $k$  degrés de liberté.

## 3.5.2 Récapitulatif des caractéristiques de la base de données utilisée

Nous disposons de 36 indicateurs mensualisés : 4 pour les variables à expliquer et 32 pour les variables explicatives :

- La durée moyenne d'un arrêt
- Le nombre de jours total des arrêts
- Le nombre d'arrêts
- La fréquence des arrêts, obtenue par la formule indiquée dans la partie sur les données assurantielles, à partir du nombre d'arrêts et de la population active.

Nous réalisons notre étude sur la période où nous avons le plus d'informations : 1999-2012 (inclus), soit 14 ans (ce qui est supérieur aux 10 ans que nous estimions le minimum lorsque nous avons commencé notre réflexion en novembre). Nous avons donc 168 valeurs pour chaque variable.

Nous avons dû faire face à plusieurs **difficultés** dans notre modélisation. En effet, nos variables présentent des tendances marquées (dans le temps). Cela entraîne l'obtention d'un  $R^2$  ajusté d'office très élevé, ce qui rend difficile la lisibilité de l'efficacité.

## 3.5.3 La modélisation

Nous créons d'abord un tableau dont les lignes vont de janvier 1999 à décembre 2012 avec nos 4 variables à expliquer et nos 32 variables explicatives.

Notre première modélisation utilise l'ensemble de nos variables mensualisées sur la période 1999-2012. Nous avons été étonnés d'obtenir un  $R^2$  ajusté très élevé (de l'ordre de 0.98, soit 98%). Cela signifie que notre « modèle explique » 98% de la variance. Comme cela nous semblait être une valeur anormalement importante, nous avons regardé de plus près nos variables, notamment par le

biais de graphiques (classiques ou de type ACP). Nous avons alors remarqué que plusieurs de nos variables, notamment parmi celles ayant le plus de poids dans notre modèle, suivaient une tendance similaire à celle de la variable à expliquer. C'est à cela que nous devons un  $R^2$  ajusté si important.

Nous avons donc exploré différentes pistes, indépendamment les unes des autres ou en regroupant plusieurs. Au final, nous avons :

1. Divisé nos données par la population active pour gommer la tendance liée à celle-ci, pour : le PIB, le nombre de personnes à temps plein, l'investissement des entreprises non financières, l'endettement des entreprises, l'endettement des ménages, les défaillances d'entreprises, le chiffre d'affaire, le chômage, la consommation médicale, le nombre de médecins et le nombre de personnes à temps partiel. . .
2. Normalisé nos données afin de pouvoir mieux interpréter les coefficients de chaque variable. Pour normaliser, nous divisons chaque ligne de nos données par la moyenne de toutes les lignes
3. Créé un data.frame afin de pouvoir nommer les colonnes (données) et les lignes (années).
4. Modélisé avec une régression linéaire (commande « lm » de R) pour chacune de nos variables à expliquer.
5. Sélectionné un modèle avec la commande « step » de R qui utilise le critère AIC.

### 3.6 Résultats

Nous avons utilisé une nouvelle fois la fonction « lm », mais à partir des variables mises en exergue par la sélection de modèle (fonction step). Nous choisissons de conserver, suite à une sélection supplémentaire après « step », les variables explicatives :

- dont la p-value (voir figure 3.4 ci dessous) est inférieure à  $10^{-10}$  dans l'un des modèles
- dont la p-value est inférieure à  $10^{-5}$  dans au moins deux modèles.

On obtient alors :

- Avec une influence positive :
  - Le nombre de personnes à temps plein,
  - L'endettement des ménages,
  - L'EURIBOR 3 mois,
  - Le cours du CAC mid & small,
  - Le taux de chômage,
  - La balance des paiements,
  - Le salaire moyen,
  - L'inflation.
- Avec une influence négative :
  - Le taux de mortalité,
  - Le taux de marge,

[H]

FIGURE 3.6 – ACP uniquement avec les variables conservées après la régression (1)

[H]

FIGURE 3.7 – ACP uniquement avec les variables conservées après la régression (2)

- Le taux d’investissement,
- Le cours du baril de pétrole BRENT,
- Le cours du SBF 120.

Nous entendons, par influence positive, que lorsque l’indicateur croît, les arrêts de travail croissent également.

FIGURE 3.4 – Résultats de la modélisation R pour la variable à expliquer « fréquence ». Les « Estimate » correspondent aux  $\hat{\beta}_i$  de notre modèle linéaire.

FIGURE 3.5 – Résultats de la modélisation R pour la variable à expliquer « fréquence », après utilisation de la fonction « step ».

Nous pouvons noter que certaines données étaient attendues (PIB ...) mais, même si elles ont un impact (presque intuitif) sur nos données, nous avons décidé de préférer le critère objectif de la p-value.

### 3.6.1 ACP avec les variables sélectionnées

Nous voyons que le premier axe est bien expliqué par nos variables à expliquer. En outre, la tendance du graphe des individus nous montre que les valeurs de nos variables à expliquer n’ont jamais été aussi importantes qu’actuellement, ce qui montre bien qu’il y a une réelle problématique par rapport au coût des arrêts de travail.

### 3.6.2 Analyse de la corrélation entre nos variables à expliquer et les variables sélectionnées

Nous utilisons la corrélation de Pearson :

$$r_p = \frac{\sigma_{xy}}{\sigma_x * \sigma_y} \quad (3.8)$$

avec  $\sigma_{xy}$  la covariance entre les variables x et y et  $\sigma_x$  et  $\sigma_y$  les écart-types, respectivement de x et de y.

[H]

FIGURE 3.8 – Normalité des résidus (variable à expliquer : fréquence des accidents du travail).

Nous utilisons la commande « cor » de R qui nous permet d'obtenir les coefficients de corrélations entre nos variables explicatives sélectionnées et nos variables à expliquer avec la méthode de Pearson.

### 3.6.3 Inférence statistique

#### Résidus gaussiens

Nous observons, en comparant nos résidus à une distribution normale avec la fonction plot de R, que ces premiers sont gaussiens (cf figure 3.8). Une estimation non biaisée de la variance de ces résidus est alors :

$$s^2 \equiv \hat{\sigma}^2 = \frac{1}{n - (p + 1)} \sum_{i=1}^n \hat{W}_i^2 \quad (3.9)$$

avec  $\hat{W}_i = Y_i - \hat{Y}_i$ .

Sur la figure 3.5, s correspond à la « Residual Standard Error ». C'est égal à 0.01143.

Nous observons également que les résidus ne sont pas indépendants et identiquement distribués. En effet, on retrouve la forme mensuelle de nos données car la fréquence provient de la division entre le nombre de cas et la population active. Or le nombre de cas a été mensualisé en divisant le chiffre annuelle par 12, ce qui laisse un caractère périodique au vecteur fréquence. Nous retrouvons cela ici.

Comme nos résidus sont gaussiens, nous pouvons avoir quelques résultats d'intervalles de confiance que nous allons exposer dans le paragraphe suivant.

#### Intervalles de confiance

**Pour les  $\beta_i$  :**  $\frac{\hat{\beta}_i - \beta_i}{S\sqrt{h_{ii}}} \sim T_{n-p-1}$  Loi de Student avec  $H = (X'X)^{-1}$

$$IC_{95\%} = [\hat{\beta}_i - t_{n-p-1,0.975} * S * \sqrt{h_{ii}}; \hat{\beta}_i + t_{n-p-1,0.025} * S * \sqrt{h_{ii}}]$$

avec  $t_{n-p-1,0.975}$  le quantile à 97.5% d'une loi de Student à n-p-1 degrés de liberté.

FIGURE 3.9 – Intervalles de confiance pour les paramètres  $\beta$

**Pour  $\sigma$  :**  $(n - p - 1) * \frac{S^2}{\sigma^2} \sim \chi_{n-p-1}^2$

$$IC_{95\%} = \left[ \sqrt{\frac{(n-p-1)*S^2}{x_{n-p-1,0.975}^2}}; \sqrt{\frac{(n-p-1)*S^2}{x_{n-p-1,0.025}^2}} \right]$$

avec  $x_{n-p-1,0.975}^2$  le quantile à 97.5% d'une loi du  $\chi^2$  à n-p-1 degrés de liberté. Avec la fréquence des accidents du travail comme variable à expliquer,



nous obtenons  $[\sqrt{145 * \frac{0.01143^2}{180.2291}}; \sqrt{145 * \frac{0.01143^2}{113.5557}}]$  comme intervalle de confiance à 95% pour  $\sigma$ . Ceci est cohérent avec la valeur de 0.01143 que nous avons obtenue comme estimation de  $\sigma$  dans la partie précédente ( $0.010252 < 0.01143 < 0.012916$ )

### 3.7 Discussions

Il s'avère finalement que, oui, il y a un lien entre les données économiques et la sinistralité en arrêt de travail. Le fait que certaines de nos données expliquent bien les différents modèles en est la preuve. Notre  $R^2$  ajusté supérieur à 0.9 ne pose pas forcément de problème. En effet, nos résultats semblent cohérents avec une analyse macroéconomique succincte.

Tout d'abord, notons que nous avons enlevé de notre analyse la population active et la population totale dont l'influence sur les arrêts de travail est triviale. Le salaire moyen va influencer positivement sur la sinistralité en arrêt de travail car la personne dont le salaire est important sera plus enclin à prendre des jours d'arrêts de travail. En effet, elle bénéficie d'une sorte de coussin de sécurité de par son salaire et ses économies. A l'inverse, quelqu'un qui a un salaire bas fera en sorte de ne perdre aucun jour de paie (délai de carence etc.) et, de ce fait, ne se mettra pas si facilement en arrêt de travail.

Il faut également noter que certaines données influent positivement parce qu'elles ont la même tendance que nos variables à expliquer. C'est le cas de l'endettement des ménages, qui ne cesse de croître au fil des années comme peut le montrer l'ACP faite auparavant.

## Chapitre 4

# Applications

### 4.1 Description de la base de données

L'application s'effectue sur une base de données fournie confidentiellement par l'assureur Verlingue. Elle comporte 52785 individus anonymes et indemnisés par l'assureur : 9198 au motif d'un accident du travail, 43487 pour des maladies, 100 en raison de maternité. 7547 dossiers sont toujours en cours de traitement par l'assureur. L'échantillon d'individus atteint par une maladie s'étale de 2005 à 2014, celui touché par des accidents du travail de 2007 à 2014. Cependant, certaines variables de la base de données présentent des valeurs aberrantes et des valeurs manquantes, ce qui rend la base difficilement exploitable sans un nettoyage préalable.

### 4.2 Exploitation possible du modèle sur la base de données fournie

On remarque que nous avons une censure à gauche car des personnes ont débutés leur arrêt de travail en dehors de la période d'observation mais leur arrêt cours encore en 2013. Ainsi, certains individus en arrêt de travail en 2013 n'apparaissent pas dans les données sélectionnés de cette même année car ils ont été répertoriés dans les années précédentes. Nous observons aussi une troncature à droite car les individus dont la durée de l'arrêt était inférieure à la franchise ne sont pas répertoriés. Ainsi, le seuil de temps défini par la franchise n'est pas atteint pour ces individus en arrêt de travail et n'apparaissent pas dans la base de données fournie par Verlingue.

A partir des valeurs des indicateurs sélectionnés dans le modèle pour l'année 2013, on peut déterminer grâce à notre modèle le nombre moyens de jours et le nombre total de jours d'arrêt de travail pour cette même année. On peut aussi souligner la difficulté d'exploiter ces données avec notre modèle pour les raisons suivantes. Les valeurs manquantes, les aberrations, la censure à gauche et

la troncature remettent en cause la fiabilité des données fournies pour exploiter correctement notre modèle. A cela s'ajoute l'obligation de conserver les individus atteints de maladie afin de conserver un nombre de cas conséquent pour la comparaison avec les résultats obtenus par notre modèle et de ne pas s'appuyer sur une base fortement biaisée. De plus, la franchise de l'assureur qui définit un seuil minimal de jours avant indemnisation entraîne une surestimation du nombre de jours moyens par rapport au nombre de jours moyens obtenus par notre modèle avec les valeurs des indicateurs sélectionnés de l'année 2013, seuls les arrêts de travail d'une durée supérieur à ce seuil étant traités par Verlingue. De la même façon, cette franchise a pour conséquence une sous estimation du nombre de cas. Enfin, le nombre total de jours d'arrêt de travail de la base de données fournie n'est absolument pas comparable avec le nombre de jours obtenu pour l'année 2013 de notre modèle.

# Conclusion

Ce projet nous a permis de mieux cerner les enjeux et les difficultés liés à l'établissement d'un modèle à partir de données réelles. Auparavant nous n'avions pas conscience du poids des obstacles qui pouvait se présenter pour réaliser une modélisation en dehors des « cas d'école ».

Dans un premier temps, nous avons attaqué la partie primordiale de notre travail qui consistait à obtenir des données via internet ou via des organismes sociaux. Nous avons alors été confrontés à des bases de données totalement biaisées et peu représentatives. Cela a considérablement ralenti notre travail. En effet, notre bureau d'étude est très axé sur cette recherche de données puisque pour expliquer les arrêts de travail il ne faut laisser aucun indicateur économique de côté.

Ensuite, une fois toutes les données récoltées, nous avons dû appréhender le traitement de ces données assez fastidieusement car une telle quantité d'informations demande beaucoup de travail à chaque modification effectuée dans notre bureau d'étude. De plus, le traitement des données avec R n'a pas toujours été simple car ce logiciel doit être guidé précisément lors du chargement et de l'exploitation de ces données. Il nous a alors fallu remédier à ces erreurs.

Après cela, le travail de modélisation a été plus simple, R étant très pratique pour les régressions linéaires. Nous avons alors pu adopter différentes méthodes pour modéliser notre problème. En divisant par la population active et après avoir normalisé les données pour mieux lire les Beta du modèle, nous avons alors pu mettre en oeuvre le cours de modèles linéaires pour conclure que, en effet, il y a un lien entre la sinistralité en arrêt de travail et certains indicateurs économiques.

Cependant, notre  $R^2$  est anormalement proche de 1, probablement du fait de la tendance commune à nos indicateurs, et nécessiterait une analyse complémentaire.

La réalisation d'un projet d'une telle envergure était une première pour nous et constitue une bonne introduction à la rédaction du mémoire de fin d'étude. La mobilisation de connaissances scolaires sur une étude concrète a été très enrichissante. Cela nous a permis de voir une des nombreuses applications professionnelles possibles des modèles étudiés lors de notre enseignement. Les démarches effectuées auprès de divers organismes nous ont permis d'exécuter les étapes administratives de la demande officielle de données. Ainsi, nous avons

pu nous rendre compte de leur complexité.

En effet, les nombreuses difficultés rencontrées au cours de projet ont constamment remis en cause le travail possible. Il a donc été obligatoire de s'adapter continuellement et rapidement afin de pouvoir continuer la progression de cette étude.

L'acquisition de nouvelles compétences comme l'assimilation du langage LaTeX pour la rédaction du rapport ou encore le traitement de nombreuses séries de données présente une utilité certaine pour la suite de nos études et notre avenir professionnel.

Les remarques de nos différents interlocuteurs nous ont offert la possibilité d'améliorer notre réflexion et de comprendre l'importance d'être guidé dans l'élaboration d'un projet de cette ampleur et dans la rédaction d'un rapport. Il nous a aussi été nécessaire d'être à l'écoute de tous les commentaires positifs ou négatifs sur notre travail afin de garder un regard critique et objectif sur notre étude.

# Bibliographie

- [1] Comme il y a peu de littérature traitant du fond de notre sujet, cette bibliographie retrace principalement les sites et ouvrages qui nous ont aidé à trouver et à décrire nos indicateurs. On y trouve aussi quelques sources liées à la description du contexte de notre étude ou à l'élaboration de ce rapport. Nous ne mentionnons pas les sources dont nous n'avons retenu qu'une simple citation ou un simple chiffre et dont la mention est déjà faite sous forme de note de bas de page.
  
- [2] INSEE, Définitions et méthodes [en ligne] (page consultée le 08/11/2013). Disponible sur : <http://www.insee.fr/fr/methodes/>.
- [3] INSEE, Base de données [en ligne] (page consultée le 08/11/2013). Disponible sur : <http://www.insee.fr/fr/bases-de-donnees/>.
- [4] Optimind Winter, Ressources, Toolkit [en ligne] (page consultée le 08/11/2013). Disponible sur : <http://www.optimindwinter.com/ressources/toolkit>.
- [5] Administration française, Formation - Travail Santé, sécurité et conditions de travail, Arrêt de travail pour maladie dans le secteur privé [en ligne] (page consultée le 5/11/2013). Disponible sur : <http://vosdroits.service-public.fr/particuliers/N526.xhtml/>.
- [6] INRS, Repères historiques [en ligne] (page consultée le 11/02/2014). Disponible sur : <http://www.inrs.fr/accueil/inrs/panorama-prevention/>.
- [7] Eco-Santé, Etat de santé, Morbidité, Accidents travail, mal. Professionnelles [en ligne] (page consultée le 21/02/2014). Disponible sur : <http://www.ecosante.fr/>.
- [8] Légifrance, Rechercher un thème Moteur de recherche [en ligne] (page consultée le 24/02/2014). Disponible sur : <http://www.legifrance.gouv.fr/>.
- [9] BOUVET, M., YAHOU, N., (2001). Le risque d'accident du travail varie avec la conjoncture économique. *DARES Analyses (n°31)*.
- [10] BITOUZE, D., CHARPENTIER, J.C., (2010). *L<sup>A</sup>T<sub>E</sub>X, l'essentiel : Pour une prise en main rapide et efficace*. Paris : Pearson Education France.

- [11] SERRES, N., (2010). Baisse des accidents du travail sur le long terme : sinistralité et éléments explicatifs par secteur d'activité. *Assurance maladie - Points de repères (n°32)*.
- [12] MOEGLIN, A., (2014). Document de cours préparé par Alain MOEGLIN : Mathématiques des Assurances de personnes.

# Annexes



# Annexe A

## Contacts

### A.1 Courriel type

Voici le message type que nous avons envoyé aux nombreux organismes que nous avons contacté. Comme nous l'adaptions à chacun de nos interlocuteurs, il en existe de nombreuses variantes.

Madame, monsieur,

Etudiante en actuariat à l'institut d'actuariat de Brest (M1 actuariat), mon groupe et moi travaillons, dans le cadre de notre "projet fil rouge", sur la corrélation entre les arrêts de travail et l'environnement économique. Si l'INSEE fournit toutes les variables macro-économiques dont nous avons besoin, nous rencontrons plus de difficultés à trouver des données sur le nombre, la durée et la fréquence des arrêts de travail en France. Idéalement, nous cherchons des données détaillées par secteur et par taille d'entreprises. Cependant, face aux difficultés que nous rencontrons dans nos recherches, même des données moins complètes nous permettraient d'avancer dans notre travail.

J'espère que vous pourrez nous communiquer les données que nous recherchons ou nous adresser à un organisme susceptible de pouvoir nous les communiquer. Je vous remercie, madame, monsieur, pour le temps que vous voudrez bien accorder à notre demande et vous prie d'agréer l'expression de mes salutations respectueuses,

Suzanne Noll

## **A.2 Interlocuteurs principaux**

### **A.2.1 CNAMTS**

Standard : 01.72.60.10.00

Monsieur Jean-Philippe Perret, responsable du département synthèse et prévisions : 01.72.60.25.12

### **A.2.2 BCAC**

Standard : 01.72.98.38.08

Roberto Wolfrum : rwolfrum@b2v.fr

Bénédicte Bonnanfant : bbonnanfant@b2v.fr

### **A.2.3 Verlingue**

Monsieur Yann Quéré : yann.quere@verlingue.fr

# Annexe B

## Code R

```
#####  
###   Chargement des données déjà mensualisées et déjà formatées   ###  
#####   de janvier 1999 à décembre 2012   #####  
#####  
  
TABLEAU=read.csv("C:/Users/Pierre/Documents/BE/Données R/TABLEAUDONNEES.  
csv",header=TRUE,sep=";",dec=".")  
  
#####  
###   Tableau des corrélations entre les données assurantielles   ###  
#####   et les données économiques   #####  
#####  
  
TAB1=data.frame(cor(TABLEAU))  
  
write.table(TAB1, file = "C:/Users/Pierre/Documents/BE/Données R/  
TABLEAUCOR2.csv", append = FALSE, quote = TRUE, sep = ";",row.names=TRUE,  
eol = "\n", na = "NA", dec = ".",col.names = TRUE, qmethod = c("escape",  
"double"),fileEncoding = "")  
  
#####  
###Division par la population active###  
#####  
  
TABLEAU[,5]=TABLEAU[,5]/TABLEAU[,24]   #PIB  
TABLEAU[,11]=TABLEAU[,11]/TABLEAU[,24]   #TEMPLEIN  
TABLEAU[,12]=TABLEAU[,12]/TABLEAU[,24]   #INVESTENT  
TABLEAU[,14]=TABLEAU[,14]/TABLEAU[,24]   #ENDENT  
TABLEAU[,15]=TABLEAU[,15]/TABLEAU[,24]   #ENDMEN
```

```

TABLEAU[,19]=TABLEAU[,19]/TABLEAU[,24] #DEFAILLENT
TABLEAU[,23]=TABLEAU[,23]/TABLEAU[,24] #CA
TABLEAU[,26]=TABLEAU[,26]/TABLEAU[,24] #POPTOT
TABLEAU[,30]=TABLEAU[,30]/TABLEAU[,24] #CHOM
TABLEAU[,32]=TABLEAU[,32]/TABLEAU[,24] #CONSOMED
TABLEAU[,33]=TABLEAU[,33]/TABLEAU[,24] #NBMED
TABLEAU[,34]=TABLEAU[,34]/TABLEAU[,24] #TEMPARTIEL

```

```

#####
###Normalisation des données###
#####

```

```

TAB=TABLEAU
TABLEAUNORMALISE=matrix(c(rep(0,8064)),nrow=168)
for (i in 1:ncol(TABLEAU)){
TABLEAUNORMALISE[,i]=TAB[,i]/mean(TAB[,i])
}

```

```

TAB=data.frame(TABLEAUNORMALISE)

```

```

names(TAB)[1:36]=c("FREQ", "NBC", "NBJ", "NBJM", "PIB", "NAT", "MORT",
"TXMARGE", "TXINVEST", "INDCOUTRAVAIL", "TEMPLEIN",
"INVESTENT", "CONFMEN", "ENDENT", "ENDMEN", "CLIMAFFAIRES",
"PRODINDMAN", "PRODINDCONST", "DEFAILLENT", "EURDOL", "EONIA",
"E3M", "CA", "POPACT", "BRENT", "POPTOT", "SBF", "INFLATION", "MIDSMALL",
"CHOM", "TME", "CONSOMED", "NBMED", "TEMPARTIEL", "BALPAI", "SALMOYENENS")

```

```

row.names(TAB)[1:168]=c("01/99", "02/99", ..., "09/12", "10/12", "11/12",
"12/12")

```

```

#####
###Modélisation avec division par la population totale###
#####

```

```

# VARIABLE A EXPLIQUER : fréquence d'accidents du travail

```

```

# rég linéaire multiple
fitfreq=lm(TAB[,1]~c(1:168)+.,data=TAB[,5:36])
summary(fitfreq)

```

```

step(fitfreq) #sélection de modèle avec le critère AIC

```

```

fitfreqaj=lm(formula = TAB[, 1] ~ c(1:168) + MORT + TXMARGE +
TXINVEST + TEMPLEIN + INVESTENT + CONFMEN + ENDMEN + CLIMAFFAIRES +

```

```
PRODINDMAN + PRODINDCONST + EONIA + E3M + CA + BRENT + POPTOT + SBF +  
INFLATION + MIDSMALL + CHOM + BALPAI + SALMOYENENS, data = TAB[, 5:36])  
#régression linéaire avec le modèle sélectionné avec "step"
```

```
summary(fitfreqaj)          #meilleur R2aj=0.9984
```

```
plot(fitfreqaj)           #graphiques des résidus
```

```
confint(fitfreqaj)        #intervalles de confiance à 95% pour nos Betas
```

```
# VARIABLE A EXPLIQUER : nombre de cas d'accidents du travail
```

```
fitcas=lm(TAB[,2]~c(1:168)+.,data=TAB[,5:36])
```

```
summary(fitcas)
```

```
step(fitcas)
```

```
fitcasaj=lm(formula = TAB[, 2] ~ c(1:168) + PIB + MORT + TXMARGE +  
TXINVEST + TEMPLEIN + INVESTENT + CONFMEN + ENDMEN + CLIMAFFAIRES +  
PRODINDMAN + PRODINDCONST + EONIA + E3M + CA + BRENT + POPTOT +  
SBF + INFLATION + MIDSMALL + CHOM + CONSOMED + NBMED + BALPAI +  
SALMOYENENS, data = TAB[, 5:36])
```

```
summary(fitcasaj)        #meilleur R2aj=0.9987
```

```
# VARIABLE A EXPLIQUER : nombre de jours d'accidents du travail
```

```
fitnbj=lm(TAB[,3]~c(1:168)+.,data=TAB[,5:36])
```

```
summary(fitnbj)
```

```
step(fitnbj)
```

```
fitnbjaj=lm(formula = TAB[, 3] ~ c(1:168) + PIB + NAT + MORT + TXINVEST +  
TEMPLEIN + INVESTENT + ENDMEN + CLIMAFFAIRES + DEFILLENT +  
EURDOL + EONIA + CA + POPACT + BRENT + POPTOT + SBF +  
INFLATION + MIDSMALL + TME + CONSOMED + NBMED + BALPAI +  
SALMOYENENS, data = TAB[, 5:36])
```

```
summary(fitnbjaj)        #meilleur R2aj=0.9915
```

```
# VARIABLE A EXPLIQUER : nombre de jours moyen d'accidents du travail
```

```
fitnbjm=lm(TAB[,4]~c(1:168)+.,data=TAB[,5:36])
```

```

summary(fitnbjm)

step(fitnbjm)

fitnbjmaj=lm(formula = TAB[, 4] ~ c(1:168) + PIB + MORT + TXMARGE +
TXINVEST + INDCOUTRAVAIL + TEMPLEIN + INVESTENT + CONFMEN + ENDMEN +
CLIMAFFAIRES + EURDOL + EONIA + CA + POPACT + POPTOT + SBF +
INFLATION + MIDSMALL + CHOM + CONSOMED + NBMED + BALPAI +
SALMOYENENS, data = TAB[, 5:36])

summary(fitnbjmaj)

#####
###Matrice avec nos variables sélectionnées et les variables à expliquer###
#####

O=matrix(c(TAB[,1],TAB[,2],TAB[,3],TAB[,4],MORT,TXMARGE,TXINVEST,TEMPLEIN,
ENDMEN,E3M,BRENT,SBF,MIDSMALL,CHOM,BALPAI,SALMOYENENS,INFLATION),ncol=17)

#####
###ACP avec nos variables sélectionnées###
#####

Q=data.frame(O)

names(Q)[1:17]=c("FREQ", "NBC", "NBJ", "NBJM", "MORT", "TXMARGE", "TXINVEST",
"TEMPLEIN", "ENDMEN", "E3M", "BRENT", "SBF", "MIDSMALL", "CHOM", "BALPAI",
"SALMOYENENS", "INFLATION" )

res.pca = PCA(Q, scale.unit=TRUE, ncp=5, graph=T)

plot.PCA(res.pca, axes=c(1, 2), choix="ind", habillage="ind")

#####
###Corrélations###
#####

cor(O,method = "pearson")

cor(TABLEAU)

```

## Annexe C

# Récapitulatif sur le traitement des données

Le tableau suivant présente l'ensemble de nos variables explicatives de manière agrégée. On y trouve, par type d'indicateur, le nom de la variable, sa source, la ventilation avec laquelle nous l'avions à l'origine (mensuelle, trimestrielle ou annuelle) et, le cas échéant, la méthode que nous avons utilisé pour la mensualiser. La mention « SO » dans la dernière colonne signifie « ans objet » et concerne les données que nous avons déjà sous forme mensuelle.

	Nom	Source	Format initial	Méthode de mensualisation
Med.	Nb de médecins	Eco-Santé	annuel	Interpolation
	Consommation santé	Eco-Santé	Annuel	Répartition
Economiques	PIB	INSEE	annuel	Interpolation
	Inflation	France-inflation	Mensuel	SO
	EONIA	Banque de France et INSEE	Mensuel	SO
	E3M	Banque de France et INSEE	Mensuel	SO
	TME	Banque de France et INSEE	Mensuel	SO
	SBF 120	Yahoo Finances	Mensuel	SO
	CAC mid small	Yahoo Finances	Mensuel	SO
	Euro-dollar	BCE	Mensuel	SO
	Pétrole BRENT	INSEE	Mensuel	SO
	Balance des paiements	Banque de France	Annuel	Répartition
	Endettement des ménages	Banque de France	Mensuel	SO
	Confiance des ménages	INSEE	Mensuel	SO
Entreprises	Emploi à temps partiel	INSEE	Trimestriel	Interpolation
	Salaires moyen	INSEE	Annuel	Duplication
	Coût du travail	INSEE	Trimestriel	Interpolation
	Défaillances	INSEE	Mensuel	SO
	Investissement	INSEE	Trimestriel	Répartition
	Taux d'investissement	INSEE	Annuel	Actualisation taux
	Taux de marge	INSEE	Annuel	Actualisation taux
	Chiffre d'affaires	INSEE	Mensuel	SO
	Confiance des entreprises	INSEE	Mensuel	SO
	Climat des affaires	INSEE	Mensuel	SO
	Chômage	Ministère du travail	Mensuel	SO
	Population active	INSEE	Trimestriel	Interpolation
Endettement des entreprises	Banque de France	Mensuel	SO	
Dém.	Population totale	INSEE	Annuel	Interpolation
	Natalité	INSEE	Annuel	Actualisation taux
	Mortalité	INSEE	Annuel	Actualisation taux

TABLE C.1 – Récapitulatif des indicateurs