
DU PRÉSENTÉISME AU TRAVAIL

Mesure et facteurs explicatifs

Thèse pour le doctorat de sciences économiques

présentée et soutenue le 3 décembre 2013 par Benjamin HUVER
sous la direction de MM. Nicolas VANEECLOO et Sébastien RICHARD

Rapporteurs

M. Thomas BARNAY, Maître de Conférences HDR à l'Université Paris-Est Créteil
M. Jean-Claude RAY, Professeur émérite à l'Université de Lorraine

Jury de soutenance

M. Thomas BARNAY, Maître de Conférences HDR à l'Université Paris-Est Créteil
M. Eric GOSSELIN, Professeur à l'Université du Québec en Outaouais
M. Hervé LELEU, Directeur de recherche au CNRS - Laboratoire LEM
M. Sébastien RICHARD, Maître de Conférences à l'Université Lille 1
M. Daniel SZPIRO, Professeur à l'Université Paris-Est Créteil
M. Nicolas VANEECLOO, Professeur à l'Université Lille 1

Remerciements

Débuter une thèse en même temps qu'une vie de famille n'a rien d'une évidence, comme Anne-Caroline, Jean et Théophile ont pu en juger au cours de ces dernières années. En premier lieu, c'est donc à eux que je pense au moment d'achever l'écriture de ce document : sans eux aucun projet n'a de sens. Outre de nombreuses heures de présence physique, ce travail, toujours activé dans mon esprit en « tâche de fond », leur a subtilisé une grande part de mon attention : je leur dis pardon et merci.

Grande est ma dette envers Sébastien Richard et Nicolas Vaneecloo, mes directeurs de thèse. Rien n'aurait été possible sans leur confiance et leur disponibilité, sans leur patience et leur pédagogie, et surtout sans le dynamisme intellectuel dont ils m'ont fait bénéficier. Au terme de ce compagnonnage, dont la réussite leur est largement imputable, je leur exprime ma profonde reconnaissance pour leur appui quasi paternel.

Très nombreux sont ceux avec qui des liens ont été tissés au cours de ces années de doctorat. Il serait trop risqué d'en dresser la liste, car celle-ci ne saurait être exhaustive. Aussi, je me contente de citer ici Ingrid Richard, Virginie Delsart et Thomas Delclite, ainsi que Benoît Desmarchelier et Jean Finez, mes collègues de bureau, et l'ensemble des doctorants ou jeunes docteurs du CLERSÉ. Sans eux, ces années n'auraient pas eu la même saveur : qu'ils soient remerciés pour leur présence, ne serait-ce qu'au travers d'une salutation amicale le matin.

Cette thèse est dédiée à mes parents,
en particulier à ma mère,
présentéiste au foyer,
absente trop tôt.

*En souvenir de
Chantal Huver
(1958-2011)*

Introduction générale

Lorsqu'au cours du XIX^e siècle, la France entre dans l'ère industrielle, le salariat devient progressivement le cadre classique de la relation de travail. Le contrat, signifiant au salarié l'investissement temporel que l'employeur attend de lui en échange d'un salaire, se généralise et apparaît alors un nouveau type de comportement, celui de déroger aux attentes horaires prédéfinies. De nombreux documents décrivant la vie ouvrière aux débuts de l'industrialisation l'évoquent et tentent de l'expliquer. Rolande Trespé (1971), dans son travail sur les mineurs de Carmaux décrit ainsi l'exemple des « piqueurs »¹ absents en masse de leur poste de travail en 1865, en particulier lorsqu'ils ont une double activité (agricole et ouvrière) : de 254 travailleurs présents en période normale (mai), on passe à 129 en juillet et 177 en août, mois de la fenaison ou de la moisson.

Dans la mesure où l'ensemble de la production dépend largement de l'activité de ces piqueurs, leurs absences saisonnières sont rapidement identifiées comme de réels obstacles au développement économique de l'entreprise. Dès cette époque, les solutions envisagées par les employeurs reposent sur des sanctions financières : au non-paiement du salaire pour les jours d'absence, s'ajoute une amende forfaitaire de « 3 francs, alors que le salaire moyen journalier est de 1 franc 90 » (Trespé, 1971, p. 205).

1. Ces ouvriers qui manient le pic pour extraire directement le charbon de la veine.

À cette époque, la maladie ne semble pas être identifiée comme un motif majoritaire d'absence. D'après une enquête de l'économiste Blanqui en 1848, la quantité d'absences n'est pas proportionnelle aux mauvaises conditions sanitaires de l'époque : « en dépit d'un travail très dur et d'une nourriture souvent déficiente, le nombre des malades et absents reste peu important » (Démier, 1982, p.17). À moins qu'elle soit très invalidante, la maladie ne constitue pas une raison de s'absenter, notamment à cause de la perte financière qu'elle impliquerait. Le problème va toutefois connaître une transformation radicale après la mise en place de moyens d'indemnisation de l'absence-maladie et le développement des sociétés de secours mutuels. Celles-ci ont en effet pour objectif de subvenir aux besoins primaires des ouvriers et de leur famille en cas de maladie. Face aux premiers abus, il apparaîtra nécessaire d'utiliser des moyens de contrôle comme la diminution graduelle des indemnités ou l'autosurveillance entre travailleurs, « afin que les ouvriers aient un intérêt toujours croissant à abrégier leur convalescence » (Villermé, 1840).

Cet épisode historique contient quasiment tous les éléments qui vont être interrogés au cours de notre thèse : l'absence de la main d'œuvre, la santé des salariés, les politiques d'incitation à l'assiduité, la présence au travail de salariés malades. Il ne s'agit donc pas de comportements nouveaux, puisqu'il sont concomitants à l'existence du salariat². La volonté des employeurs d'en déceler les facteurs explicatifs afin de mieux les maîtriser est cependant plus récente. Les entreprises se sont en effet progressivement dotées de méthodes savantes et d'indicateurs permettant de mesurer l'importance des absences de la main d'œuvre (comme l'emblématique « taux d'absentéisme »). Elles ont aussi ouvert leurs portes à la recherche académique qui, s'emparant de la question, a tenté de mieux comprendre le phénomène et d'y apporter des solutions. De ce fait, un très grand nombre de dispositifs ont été mis en place et évalués dans le cadre de ce que l'on appelle communément la « lutte contre l'absentéisme ». Cependant, la recherche autour de l'absence-maladie a également conduit à redécouvrir un phénomène négligé qui, avant l'indemnisation de l'absence, constituait sans doute un comportement fréquent (Hansen & Andersen, 2008). En effet, certains salariés se rendent à leur travail tout en étant malades, parfois pour échapper à des sanctions financières ou pour bénéficier

2. Et même bien antérieurs. Des vestiges datant de la XIX^e dynastie égyptienne (1250 avant J-C) attestent de l'existence d'un enregistrement des absences des travailleurs avec un détail précis des motifs d'absence (Treble & Barmby, 2011). Mais la systématisation du relevé des absences, ainsi que son traitement, se développe à l'ère industrielle.

d'une prime. Ce comportement, ancien dans les faits mais nouveau dans la littérature, est l'une des manifestations possibles du *présentéisme*.

Il y a beaucoup à dire sur les définitions actuelles de ce phénomène et il nous paraît opportun de le faire dès l'introduction de cette thèse. Un parallèle terminologique s'impose d'emblée : le présentéisme est à la présence ce que l'absentéisme est à l'absence, à condition que ces deux derniers termes ne soient pas considérés comme de simples synonymes. En effet, la littérature parle souvent indifféremment d'absentéisme ou d'absence, ignorant le suffixe « -isme ». Or, en langue française, celui-ci désigne « un mode de vie, de pensée ou d'action, une tendance »³. L'absentéisme possède donc un aspect chronique, répétitif : il se définit d'ailleurs comme un « manque habituel ou systématique d'assiduité à son lieu de travail »⁴. En contraste, l'absence, et en particulier l'absence-maladie, constitue un événement ponctuel et normal de la vie de travail. Peu nombreux sont les auteurs qui respectent cette différence sémantique. Parmi eux, Whitston et Edwards (1990, p. 288) estiment que « la distinction entre absence et absentéisme est importante et négligée » et qu'un niveau d'absence élevé n'implique pas toujours un problème d'absentéisme. Dans la suite de cette thèse, les deux termes seront clairement distingués.

En faisant le parallèle, le présentéisme désignerait donc une présence chronique, systématique, sur le lieu de travail, quelles que soient les circonstances. Le concept de présentéisme étant relativement récent, de nombreuses définitions du présentéisme sont disséminées dans la littérature⁵. Dans un sens large, il caractérise le surinvestissement d'un salarié, exprimé par une présence excessive au travail. Cette thèse, comme la plupart des travaux de recherche récents (Aronsson & Gustafsson, 2005 ; Gosselin & Lauzier, 2011 ; Hansen & Andersen, 2008 ; Johns, 2010), s'intéressera, elle, à une manifestation particulière du présentéisme. Il s'agit de la situation dans laquelle un salarié, tout en estimant qu'une absence-maladie serait justifiée par son état de santé, décide de se rendre sur son lieu de travail⁶. Toutefois, nous n'ignorons pas que le présentéisme peut également transparaître dans l'extension des plages horaires de travail (Simpson,

3. D'après le *Trésor de la langue française* (version informatisée), article « -isme ».

4. *Trésor de la langue française* (version informatisée), article « absentéisme ».

5. Johns (2010) en recense jusqu'à neuf.

6. À l'instar de Gosselin et Lauzier (2011, p. 18), nous estimons qu'il est « inadéquat d'associer le présentéisme à des concepts tels que la démotivation, la paresse ouvrière ou encore la délinquance professionnelle ». En effet, certains auteurs appellent présentéisme le fait de réaliser des tâches non liées à l'emploi sur le lieu de travail (D'Abate & Eddy, 2007), comme la consultation de la messagerie électronique personnelle, l'usage de l'Internet à des fins de loisirs, les appels téléphoniques personnels,

1998), que ces heures supplémentaires soient rémunérées ou non, dans la renonciation à un ou plusieurs jours de congés au profit de tâches professionnelles en cours ou encore dans le sacrifice du temps de loisir au profit du travail (McKevitt et al., 1997). La renonciation à des jours d'absences-maladie constitue néanmoins un fait que nous pourrions caractériser avec plus d'acuité que d'autres manifestations du présentisme.

Comme nous le verrons, l'économie du travail a principalement tenté d'intégrer le présentisme dans les modélisations théoriques de l'absence. D'un point de vue empirique, celui que nous adoptons, les contributions sont peu nombreuses en économie et on doit recourir à la littérature d'autres domaines, comme la gestion des ressources humaines, la psychologie du travail ou la médecine du travail, pour obtenir un premier regard sur le phénomène. L'invisibilité statistique est certainement l'une des raisons expliquant le peu d'intérêt porté aux aspects empiriques du présentisme en économie, et plus largement la méconnaissance de ce comportement dans la recherche académique. Comment savoir en effet, si un salarié présent à son poste de travail devrait être chez lui en train de se soigner ? Seuls des dispositifs d'enquête assez lourds, comme l'usage de questionnaires, permettaient jusqu'ici l'analyse du présentisme. Peu nombreuses sont donc les entreprises qui ont à la fois la détermination et les moyens financiers nécessaires pour mettre en place de telles enquêtes. En comparaison, l'analyse de l'absence est plutôt aisée et repose sur des données déjà existantes. Or, à l'instar de Gosselin et Lauzier (2011, p. 25) nous estimons que « la mesure est la clef de voûte de tout concept [... et qu'il] est impérieux de diagnostiquer le présentisme dans les organisations ». La recherche autour du présentisme a donc besoin de se doter d'instruments facilitant la mesure du phénomène. La description et l'application d'un outil économétrique permettant une telle mesure constituent le cœur de notre thèse.

Ce travail est bâti sur trois grands mouvements. Dans un premier temps, il s'agit d'apporter des éléments empiriques sur un phénomène encore assez méconnu et sur sa mesure (chapitre 1). L'intérêt de cette exploration préliminaire est triple. D'abord, il est opportun de produire un état de l'art à propos d'un phénomène que l'on s'apprête à mesurer. Même si des retours aux travaux d'autres auteurs jalonnent l'ensemble de notre démarche de recherche, une revue de la littérature ouvre cette thèse, proposant au lecteur une synthèse des connaissances disponibles. Ensuite, eu égard à notre question centrale, nous dresserons le bilan des moyens utilisés pour recueillir ces in-

etc. Ces activités semblent plus proches de ce que la littérature économique appelle parfois « flânerie » ou « coulage » et ne relèvent plus d'une baisse de la productivité pour cause médicale.

formations, en soulignant le recours quasi-exclusif à l'auto-déclaration et l'utilité de nouvelles approches plus systématiques. Enfin, se pose la question des conséquences du présentéisme : quels sont les enjeux d'une meilleure connaissance de ce comportement ? *A priori*, la présence d'un salarié, même si sa productivité est limitée par la maladie, semble toujours préférable à l'absence. Dès lors, pourquoi les entreprises devraient-elles s'intéresser au phénomène ? En ce sens, nous aurons l'occasion d'évoquer les implications économiques et sanitaires du présentéisme.

S'ouvre alors le deuxième temps de la thèse, consacré à la question d'une nouvelle méthode de mesure du présentéisme. S'il est dans l'intérêt des entreprises de détecter ce comportement, encore faut-il leur proposer un instrument souple capable d'en fournir une estimation. À ce stade, une donnée communément recueillie par les services de ressources humaines retiendra notre attention, à savoir le nombre annuel de jours d'absence d'un individu. Nous soulignerons la richesse étonnante de cette variable (chapitre 2). En son sein, plusieurs phénomènes coexistent et génèrent une forte hétérogénéité inter-individuelle qu'il serait utile de savoir déchiffrer. Quelle sorte d'instrument permettrait de procéder à un tel décryptage ? En puisant dans les acquis de la modélisation des données discrètes, nous en arriverons à l'outil économétrique qui nous permet de mesurer conjointement l'absence et le présentéisme des salariés (chapitre 3). Outre la description du fonctionnement et des présupposés théoriques d'un modèle binomial négatif à inflation de zéros, nous mènerons une batterie de simulations de Monte-Carlo destinées à tester la fiabilité du modèle pour la mesure du présentéisme. Plusieurs scénarios, dont certains lèvent les contraintes posées par le modèle, valideront notre capacité à fournir une mesure pertinente du présentéisme.

Après avoir fait ses preuves sur des données simulées, le modèle est appliqué à des données réelles. L'analyse des bases de données issues de deux organisations constitue le troisième mouvement de la thèse. Le premier jeu de données provient d'un établissement bancaire privé (environ 2600 salariés), l'autre d'un centre hospitalier public (environ 4000 salariés). Les deux contextes possèdent des originalités intéressantes pour l'analyse, et le second se caractérise par une profondeur temporelle importante (25 années), qui offre la possibilité d'effectuer des croisements longitudinaux. L'application empirique se fera en deux étapes. D'une part, après avoir détaillé les variables explicatives en présence et décrit la structure des populations étudiées, nous présenterons les résultats bruts des régressions (chapitre 4). Sans entrer de suite dans l'analyse des facteurs explicatifs du présentéisme, nous aurons l'occasion de produire une statistique

descriptive très détaillée, intégrant par exemple la distribution des probabilités de présentéisme, une comparaison avec les données longitudinales d'absence ou encore une analyse de l'ajustement des différents modèles de comptage. D'autre part, dans un dernier chapitre, nous nous attacherons à développer successivement l'ensemble des facteurs de présentéisme identifiés grâce à nos modélisations (chapitre 5). Conjointement à la proposition d'un outil de mesure, le second objectif de ce travail est en effet d'examiner les raisons pouvant pousser les salariés à travailler étant malades. Quels éléments ont le plus de poids dans leur décision ? S'agit-il de nécessités individuelles ? De pressions liées à leur emploi ? Les facteurs de présentéisme sont-ils principalement à chercher dans le contexte de travail ? Au cours d'un panorama général sur nos résultats, nous mettrons l'accent sur deux thématiques qui nous semblent essentielles pour la compréhension du présentéisme : l'âge des salariés et la responsabilité managériale.

Chapitre 1

État des lieux sur le présentéisme

Introduction

Au travers des éléments fournis plus haut, il apparaît nettement que la définition du présentéisme n'est pas uniforme selon le domaine de recherche dans lequel elle est exploitée, selon l'époque, voire même selon le chercheur qui la mobilise. Bien que l'approche du présentéisme comme renonciation à une absence-maladie justifiée tende à devenir majoritaire, on recense encore dans la littérature des acceptions très hétérogènes du phénomène. Cette diversité implique naturellement que les méthodes de mesure soient différentes et peut également expliquer en partie les disparités dans les résultats obtenus.

En effet, dans ce premier chapitre consacré à une synthèse des connaissances, on s'apercevra rapidement que les résultats de la littérature sont encore assez imprécis. Même si le nombre d'articles produits au sujet du présentéisme demeure relativement faible, il y a un foisonnement de résultats sur les causes de présentéisme, qui peut produire une impression de flou et peut impliquer des contradictions. Comme toute thématique

de recherche en émergence, le champ du présentéisme reste en quête d’une cohérence interne et d’un agenda de recherche mieux structuré. L’obtention d’éléments plus rigoureux sur les facteurs explicatifs est également entravée par une méthodologie de mesure reposant sur de lourdes enquêtes qualitatives dont les approches du présentéisme sont hétéroclites. L’une des contributions de cette thèse est d’ouvrir une nouvelle voie s’appuyant sur une réflexion globale autour des comportements d’absence et structurée autour d’une méthodologie de mesure unifiée.

Dans ce chapitre, il s’agira essentiellement de parcourir ce qui a déjà été écrit à propos du présentéisme. Pour ouvrir cette revue de littérature, nous décrirons d’abord les conditions dans lesquelles est apparue la question du présentéisme (section 1). Dans une perspective bibliométrique d’abord, pour montrer que cette thématique s’est développée dans le flanc de la recherche sur l’absence-maladie, puis en montrant comment la théorie économique de l’absence a peu à peu intégré le phénomène du présentéisme. Les chapitres qui suivent démontreront l’intérêt de ce retour aux sources : notre démarche est en effet largement fondée sur l’origine commune de l’absence-maladie et du présentéisme et sur le rapport étroit entre ces deux comportements.

Ensuite, nous nous intéresserons aux différentes formes de mesure utilisées dans les articles portant sur le présentéisme (section 2). Avant d’évoquer les résultats de la littérature, il semble en effet opportun d’indiquer comment ils ont été obtenus. Nous aurons alors l’occasion de nous interroger, non seulement sur l’hétérogénéité des méthodes de mesure (qui découle directement de l’hétérogénéité des définitions), mais également sur les limites de ces méthodes, reposant souvent sur l’auto-déclaration. Dans le même mouvement, nous présenterons les résultats obtenus (section 3), en détaillant successivement chaque facteur explicatif du présentéisme.

En dernier lieu, bien que cette thèse n’ait pas pour objectif de montrer les conséquences du présentéisme, nous décrirons les implications importantes du phénomène (section 4). Pour cela nous puiserons largement dans la littérature de la médecine et de l’épidémiologie, puisque les conséquences se manifestent principalement sous la forme d’une dégradation de la santé des présentéistes, d’une augmentation ultérieure de leur nombre d’absences et d’une transmission de la maladie entre l’individu présentéiste et ses collègues. Du point de vue de l’entreprise, ce comportement implique également une réduction de la productivité et une dépréciation de la qualité.

1 Émergence du concept de présentéisme

Pour entamer notre démarche de revue de la littérature, nous proposons une analyse quantitative de la production scientifique sur le présentéisme au travail. En ce sens, la disponibilité (en ligne) de bases de données bibliographiques de plus en plus complètes et le développement de la bibliométrie ouvrent des perspectives étonnantes. L'enjeu de cette première étape n'est bien entendu pas d'entrer dans le contenu des articles, mais d'obtenir un panorama général sur l'intérêt suscité par le présentéisme et par ce que la littérature appelle « absentéisme ».

Dans un deuxième temps, nous voudrions illustrer l'émergence progressive du présentéisme en économie, en convoquant une preuve épistémologique. En effet, on peut montrer comment, dans les dernières décennies, les approches théoriques classiques de l'absence au travail ont peu à peu intégré la problématique de la santé, puis celle du présentéisme dans leurs modèles explicatifs. Cette intégration ne nous semble pas anodine : elle apparaît parallèlement aux premières preuves empiriques des conséquences du présentéisme et met en lumière le souci de compléter les modèles théoriques de l'absence en y incluant un phénomène dont les implications n'avaient été jusque là que soupçonnées.

1.1 Approche bibliométrique

Même si l'objet de ce premier chapitre, et plus largement de cette thèse, est le présentéisme au travail, cette analyse bibliométrique inclut à la fois la production scientifique portant sur le présentéisme et celle concernant l'absence et l'absentéisme. Nous montrerons en effet que la recherche sur l'absence au travail est le terreau dans lequel a germé la thématique du présentéisme. D'autres auteurs ont déjà adopté une telle approche quantitative, soit pour montrer simplement le déficit d'intérêt accordé au présentéisme en comparaison à l'absence (Dew et al., 2005), soit dans la perspective de réaliser une revue systématique de la littérature (Schultz et al., 2009 ; Schultz & Edington, 2007).

1.1.1 Méthodologie

Notre analyse porte initialement sur 294 revues universitaires reconnues pour leur pertinence dans le champ de l'économie et de la gestion. Dans le classement des revues

effectué par le CNRS¹ en octobre 2011, elles occupent les rangs 1 ou 2. Le classement CNRS répartit les revues en 24 domaines de recherche parmi lesquels seuls neuf ont effectivement produit des articles autour de l'absence-maladie ou du présentéisme². L'analyse bibliométrique se concentre donc sur les 112 revues incluses dans ces neuf domaines et en examine la production scientifique entre 1970 et 2011, soit plus de 210 000 articles.

Dans l'objectif de maximiser la couverture temporelle des revues, quatre bases de données bibliographiques sont exploitées : *Web of Science* (Thomson Reuters), *Scopus* (Elsevier), *Business Source Complete* et *Econlit* (EBSCO publishing). Chacune des 112 revues est analysée dans la base de données qui offre la meilleure couverture, ce qui permet d'intégrer la quasi totalité des articles parus depuis 1970 ou depuis le commencement de la publication. Le décompte des articles concernant l'absence au travail et le présentéisme s'est déroulé en deux vagues. Le premier critère de recherche était la présence dans le titre, le résumé ou les mots-clés des expressions « *absenteeism* », « *sickness absence* » ou « *attendance* » d'une part, et des expressions « *presenteeism* », « *sickness presence* » ou « *sickness attendance* » d'autre part. Les titres et les résumés ont ensuite été parcourus pour vérifier que les thématiques de l'absence au travail et/ou du présentéisme étaient réellement l'enjeu principal de la publication. Les articles dans lesquels ces deux thématiques n'étaient que secondaires ou n'apparaissaient qu'en guise d'illustration n'ont pas été retenus.

Enfin, il était nécessaire de choisir une assiette de calcul commune, permettant de tenir compte du volume de publication de chaque revue (certaines d'entre elles étant plus récentes ou publiant moins d'articles) et de comparer les catégories de revues sur des bases similaires. Le rapport utilisé est le nombre d'articles publiés sur le thème pour 1 000 articles publiés dans l'ensemble.

1.1.2 Évolution temporelle

L'intérêt de la littérature académique sur les comportements d'absence de la main d'œuvre est déjà ancien. Dès 1928, Culpin et Smith dépeignent par exemple la situation de plusieurs entreprises et mettent déjà en évidence des facteurs d'absence comme la

1. Comité National de la Recherche Scientifique (section 37).

2. Ces neuf domaines sont : l'économie et la gestion générales, l'économie publique, la gestion des ressources humaines, l'économétrie, la recherche opérationnelle, l'économie de la santé, la stratégie et le management international, la théorie des organisations et l'économie du travail.

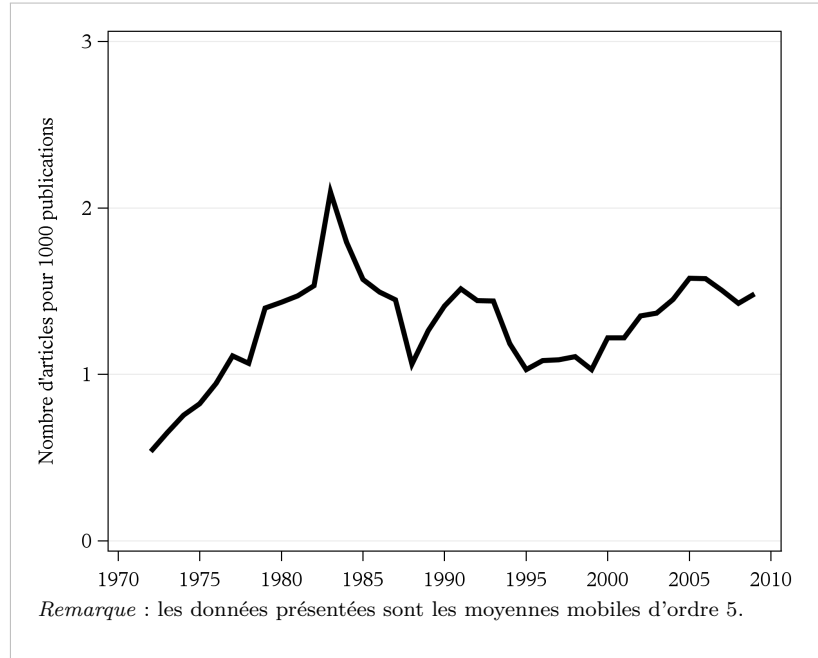
pénibilité du travail ou l'insatisfaction sur le contenu du travail. Les études de ce type vont se multiplier dans la seconde moitié du XX^e siècle et en particulier dans les dernières décennies. L'intérêt porté au présentéisme semble quant à lui plus récent. Même si le terme apparaît déjà en 1955 (Canfield et Soash), il est utilisé dans un sens différent, puisque le présentéisme y est décrit, en contraste avec l'absentéisme, comme un comportement de parfaite assiduité (Gosselin & Lauzier, 2011). Johns (2010) recense même des occurrences antérieures dans le même esprit. Certains auteurs ont cependant conscience de l'existence du présentéisme tel qu'il est entendu dans ce travail, sans le nommer ainsi. Par exemple, Culpin et Smith signalent déjà en 1928 qu'en l'absence d'un suivi médical, certains salariés se rendent à leur travail sans égard pour leur santé, jusqu'à développer de très graves maladies. Steers et Rhodes (1978), qui sont des pionniers de la modélisation de l'absence, perçoivent que certaines absences sont plutôt saines pour l'organisation et que certains dispositifs de lutte contre l'absence peuvent s'avérer contre-productifs. Même dans quelques articles plus récents, le terme de présentéisme n'est pas toujours adopté : c'est le cas dans l'article de McKevitt et al. (1997), le premier spécifiquement consacré au phénomène, où les auteurs décrivent la propension de certaines catégories professionnelles à travailler étant malade, par exemple chez les médecins. Grinyer et Singleton (2000, p. 8) utilisent quant à eux une périphrase pour décrire le présentéisme : « *inappropriate non-use of sick leave* »³.

La fréquence des publications sur l'absence au travail est restée stable lors des quatre décennies étudiées, avec un pic au début des années 1980. Depuis 1970, 285 articles ayant trait directement à cette problématique ont été publiés dans les revues analysées. Les années 1980 et les années 2000 ont été les plus prolifiques, avec respectivement 1,47 et 1,42 articles sur l'absence pour 1 000 articles publiés, en moyenne. En effet, on constate que le nombre de publications a constamment augmenté entre 1970 et le milieu des années 1980. Puis, après un temps de baisse entre 1984 et 1988, la production scientifique revient à un niveau comparable, jusqu'à une rupture au milieu des années 1990. Après 1995, le nombre de publications augmente à nouveau progressivement.

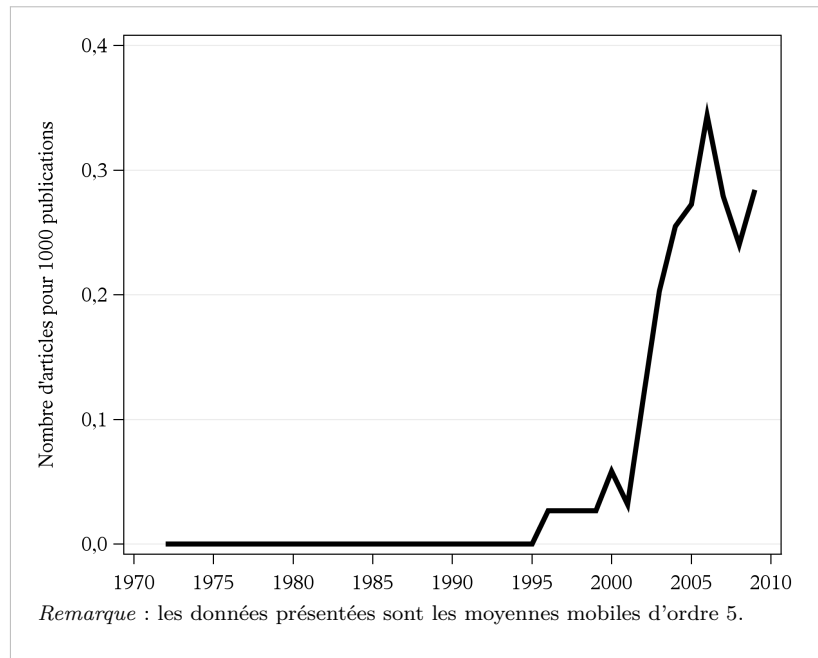
Parallèlement, l'analyse diachronique montre également l'émergence plus tardive de la thématique du présentéisme : sur les 22 articles recensés, la majeure partie a été publiée dans la dernière décennie. Bien qu'il ne soit pas à proprement parler un comportement

3. Littéralement, « la non-utilisation inappropriée de l'absence-maladie ».

Graphique 1.1 – Production scientifique sur l’absence au travail (1970-2011)



Graphique 1.2 – Production scientifique sur le présentéisme (1970-2011)



nouveau, le présentéisme n'est devenu que très récemment un champ de recherche à part entière.

Les graphiques 1.1 et 1.2 (page 13) présentent ces évolutions sous la forme de tendances générales. Celles-ci pourraient s'expliquer assez simplement. On peut penser qu'en premier lieu, la recherche a cherché à comprendre un phénomène perçu depuis longtemps par les entreprises comme un dysfonctionnement aux lourdes implications économiques. Pourquoi les salariés s'absentent-ils ? Comment diminuer le nombre de ces absences ? Jusqu'au milieu des années 1980, paraissent d'ailleurs des travaux pionniers, comme le modèle de Steers et Rhodes (1978) ou d'Allen (1981) en économie. Puis, au fur et à mesure de la mise en place de dispositifs destinés à contrecarrer ces comportements d'absence, a émergé un autre phénomène, potentiellement renforcé par des mesures restrictives visant à diminuer l'absence. Apparaît alors la question du présentéisme : pourquoi certains salariés vont-ils travailler alors qu'ils devraient être absents ?

1.1.3 Domaines de recherche

Comme nous l'avons indiqué plus haut, les revues étudiées se répartissent en différents domaines de recherche. Savoir dans quels domaines sont principalement publiés les articles autour de l'absence et du présentéisme jette une lumière nouvelle sur l'analyse. En effet, quatre domaines de recherche sortent du lot et concentrent à eux seuls 95% des articles sur l'absence et l'ensemble des publications sur le présentéisme. Il s'agit de la gestion des ressources humaines, l'économie de la santé, l'économie du travail et les revues généralistes en économie et en gestion. Le domaine de la gestion des ressources humaines est l'un des plus prolifiques : 37,2% des articles sur l'absence en proviennent. Mais depuis le milieu des années 1980, la production scientifique sur l'absence diminue dans ce champ, tandis que celle d'autres domaines progresse, comme en économie du travail et surtout en économie de la santé. Cet intérêt croissant des économistes de la santé pour l'absence des salariés est sans doute également l'une des raisons de l'émergence d'un intérêt pour le présentéisme. En effet, sur les 22 articles recensés autour de ce thème, 17 proviennent du champ de la santé. Les autres ont été publiés par des revues de gestion (généralistes ou en ressources humaines).

La répartition en domaines de recherche semble donc confirmer le scénario proposé ci-dessus. La recherche autour de l'absence, qui part de l'entreprise et sa volonté de réagir vis-à-vis de ce comportement de la main d'œuvre, a d'abord été d'orientation gestion-

naire. Il s'agissait alors d'entreprendre des actions à l'encontre d'un dysfonctionnement. Puis au fur et à mesure, la thématique se développe davantage dans la perspective de la santé des salariés et de leur bien-être au travail, tout en ayant conscience que des politiques trop restrictives peuvent occasionner un phénomène dysfonctionnel inverse. Le concept de présentéisme naît donc dans le prolongement de la réflexion sur l'absence appréhendée sous l'angle de la santé des salariés.

1.2 Extension de la théorie économique de l'absence

Cette émergence du présentéisme au cœur de la réflexion sur l'absence se manifeste également dans l'ouverture progressive des modèles théoriques de l'absence à cette problématique nouvelle. Avant de décrire cette évolution, rappelons quelques éléments de la théorie standard de l'offre de travail, outil classique de l'analyse des mécanismes de l'absence au travail (Allen, 1981 ; Dionne & Dostie, 2007 ; Dunn & Youngblood, 1986).

Dans ce cadre, l'absence est envisagée comme un choix individuel et rationnel. Le salarié doit arbitrer entre un temps de travail (tel qu'il est planifié par l'entreprise) et un temps de loisir. Le temps de travail donne droit à un salaire qui permet à l'individu de consommer. On considère en effet qu'il existe une combinaison entre travail (salaire) et loisir maximisant l'utilité de l'individu. L'absence au travail est alors envisagée comme un moyen pour le salarié de faire converger son temps de travail effectif (fixé contractuellement par l'employeur) vers le temps de travail qui maximise son utilité.

Cette maximisation se fait néanmoins sous contrainte. Une contrainte budgétaire, d'abord, la consommation ne pouvant excéder l'ensemble des revenus du salarié. Les pénalités par lesquelles l'entreprise compense la perte inhérente à l'absence du salarié, réduisent le salaire. Qu'il s'agisse d'une sanction financière concrète, du risque de ne pas obtenir une prime ou simplement de la perspective d'être licencié, cet item influence donc la façon dont l'individu va maximiser son utilité. Il existe en outre une contrainte temporelle : le salarié dispose d'une quantité de temps limitée qu'il doit partager entre le travail et les loisirs.

L'un des principaux inconvénients de ce modèle classique est que, en dehors de la sanction financière, l'employeur ne peut assister que passivement à « l'absentéisme » des salariés. Pour contourner cet inconvénient deux perspectives prenant des libertés

vis-à-vis du modèle standard, émergent au cours des années 1980 et deviennent prédominantes. Sans entrer dans le détail de ces théories, il est intéressant pour notre réflexion d'observer comment ces deux approches se sont progressivement ouvertes à la question de la santé, puis à celle du présentéisme. D'une part, la théorie du salaire d'efficiency, principalement élaborée par Shapiro et Stiglitz (1984), s'est peu à peu enrichie du constat que la présence de salariés malades n'est pas optimale pour la firme (Brown & Sessions, 2004). D'autre part, l'approche principal-agent, adaptée à la problématique de l'absence par Coles et Treble (1993), a été généralisée par Chatterji et Tilley (2002) pour tenir compte du présentéisme.

1.2.1 Modèle du salaire d'efficiency

Le contexte du modèle de Shapiro et Stiglitz (1984) est comparable à celui du modèle standard, mais l'hypothèse est faite qu'il existe une asymétrie d'information entre l'employeur et les travailleurs sur l'effort que ces derniers sont prêts à déployer au travail. Certains salariés peuvent « tirer au flanc », par exemple par l'absentéisme, avec le risque d'être identifiés et licenciés. L'entreprise a en effet à sa disposition des dispositifs de contrôle, mais ceux-ci ont un coût. En situation de plein emploi, ce risque de licenciement n'est pas efficace pour inciter les salariés à l'effort ; les tire-au-flanc sont rapidement réembauchés et le fait d'avoir été absentéistes ne leur aura valu aucune pénalité. Mais, en offrant à ses travailleurs un salaire supérieur au salaire d'équilibre du marché (le salaire d'efficiency) l'entreprise va leur donner une incitation supplémentaire au travail. En effet, la perspective de ne pas retrouver immédiatement un emploi aussi bien payé représente alors un manque à gagner.

Dix ans plus tard, Barmby et al. (1994) reprennent cette perspective en intégrant la santé comme un nouveau paramètre de la fonction d'utilité des salariés. Cette fois, ils postulent que l'asymétrie d'information porte principalement sur la santé des travailleurs. L'entreprise ne connaît pas l'état de santé des personnes qu'elle embauche. Les salariés eux-mêmes ont d'ailleurs une incertitude à ce sujet. Les décisions qui résultent de la maximisation de leur utilité sont prises *ex post*, après la signature du contrat, juste avant que la production ne commence. Les individus sont neutres au risque et, comme dans la théorie statique de l'offre de travail, partagent une quantité de temps entre travail (générant un salaire) et loisirs.

L'état de santé (δ)⁴ vient pondérer les items de la fonction d'utilité des salariés : plus un salarié est malade, plus le temps de loisir aura un poids important dans sa fonction d'utilité. Mais l'absence n'est acceptable pour l'entreprise qu'à partir d'un état de santé δ^z . Barmby et al. (1994) définissent deux autres niveaux critiques de l'état de santé :

- un seuil δ^b au-delà duquel un travailleur n'est plus en mesure de travailler ;
- un seuil δ^a où il indiffère au salarié d'être au travail ou d'être absent : au-delà, il a une préférence pour l'absence.

Dans l'idéal δ^z , le seuil de tolérance d'absence, doit donc se situer entre δ^b et 1. Entre δ^a , le seuil à partir duquel le salarié préfère s'absenter, et δ^z , le niveau à partir duquel l'absence est acceptable pour l'entreprise, il est possible que les individus simulent un état de santé plus mauvais qu'il ne l'est en réalité. Autrement dit, ils surestiment le paramètre δ et profitent de l'asymétrie d'information pour tirer au flanc. L'entreprise peut contrer ce comportement en mettant en place un dispositif de contrôle. Mais le coût de celui-ci peut s'avérer très élevé. Les conclusions de Barmby et al. (1994) rejoignent alors celles de Shapiro et Stiglitz (1984). Du côté de la firme, la solution pour contourner cette asymétrie d'information concernant la santé, serait d'augmenter les salaires pour créer un manque à gagner en cas de licenciement.

Après dix autres années, le modèle connaît un nouveau développement. Brown et Sessions (2004), reprennent les travaux de Barmby et al. (1994) en intégrant le comportement de certains individus présentéistes. Leur conclusion se concentre davantage sur le niveau critique de δ^z . Si ce seuil de déclenchement de l'absence est trop faible, les salariés sont incités à être absentéistes. Mais Brown et Sessions (2004) soulignent également que lorsque ce seuil est trop élevé, les salariés sont incités à venir travailler, alors qu'ils devraient rester chez eux. Les auteurs estiment que ce comportement est négatif pour l'entreprise (baisse de la productivité), pour le salarié lui-même (indisposition, manque d'attention, etc.) et pour les autres salariés (contagion). Brown et Sessions (2004) adressent ainsi aux entreprises un appel à la vigilance sur l'ajustement de ce seuil de santé au-delà duquel l'absence devient acceptable. Mais dans notre contexte, l'entreprise n'a que très peu d'influence sur la définition d'un seuil critique, cet arbitrage étant la prérogative du médecin, prescripteur de l'arrêt-maladie.

4. Le paramètre δ est compris entre 0 (bonne santé) et 1 (mauvaise santé).

1.2.2 *Approche principal-agent*

La même trajectoire d'évolution peut être observée dans les travaux qui décrivent l'absence des salariés en s'appuyant sur le problème principal-agent. En 1985, Weiss montre que l'absence des salariés est plus coûteuse dans une firme où la production est conditionnelle à la présence d'un nombre donné d'individus : il développe l'exemple de firmes organisées autour d'une chaîne de montage. Les entreprises se trouvant dans cette situation vont être incitées à augmenter leurs salaires pour encourager la présence des salariés. Ceci est vrai même lorsqu'il n'y a pas d'indemnisation de l'absence. Ce premier travail de Weiss (1985) se contente donc d'analyser les mécanismes de l'absence et de déterminer un niveau de salaire optimal qui, conjointement, incite les salariés à être présents et permet à l'entreprise de maximiser son profit.

Les travaux ultérieurs de Coles et Treble (1993) généralisent les constats effectués par Weiss (1985) et intègrent un élément nouveau, à savoir le fait que les salariés peuvent être légitimement absents à cause d'une maladie. La question qui est initialement posée est la suivante : pourquoi les contrats établis entre employeurs et salariés comprennent-ils une indemnisation de l'absence, alors que cela semble aller à l'encontre des intérêts de la firme ? Coles et Treble (1993) y répondent en adoptant le cadre théorique du problème principal-agent. L'employeur (le principal) et les salariés (les agents) cherchent à convenir d'un contrat comportant deux éléments : le salaire (w) et le niveau d'indemnisation de l'absence (s). Le paiement d'une indemnité d'absence a du sens, dans la mesure où les salariés ont une aversion pour le risque (ici le risque de perte de salaire à cause de la maladie). Dès lors, le problème se résume à deux types de contrats : soit l'absence est entièrement indemnisée ($w = s$), soit il existe un manque à gagner ($w > s$). Les salariés vont s'orienter vers l'un de ces deux contrats en suivant leur aversion au risque.

Coles et Treble (1993) comparent deux types d'entreprises. Ils reprennent l'exemple de Weiss (1985), celui d'une firme organisée autour d'une chaîne de montage pour laquelle l'absence d'un salarié peut générer une perte supérieure à la seule production marginale de l'absent. Ils comparent cette situation à celle d'une entreprise pour laquelle l'absence d'un salarié n'entraîne que la perte de la production de celui-ci. Coles et Treble (1993) montrent notamment que les entreprises pour qui l'absence de la main d'œuvre est très coûteuse peuvent renoncer à l'indemnisation de l'absence afin d'attirer des salariés peu enclins à s'absenter. Mais dans ce modèle, seules trois situations sont envisagées :

les salariés peuvent soit être présents, soit être absentéistes (tire-au-flanc), soit être absents pour motif de maladie. Les salariés malades sont donc toujours considérés comme absents : l'hypothèse que certains salariés malades se présentent au travail pour continuer à toucher leur salaire est ignorée.

Dans une troisième étape de développement du modèle, Chatterji et Tilley (2002) tiennent compte de cette possibilité pour les salariés de se comporter en présentéistes. Ils introduisent une nouvelle contrainte : il vaut mieux pour l'employeur, comme pour le salarié, qu'en cas de maladie invalidante ce dernier reste chez lui pour se soigner. En effet, l'hypothèse est faite que la productivité du salarié est une fonction de son état de santé : au-delà d'un certain seuil de santé, l'entreprise n'a donc pas intérêt à encourager la présence au travail des salariés. D'après Chatterji et Tilley (2002), c'est d'ailleurs la seule raison d'être des indemnités d'absence : le principal n'y trouve un intérêt que dans la mesure où il existe un phénomène de présentéisme ayant un effet négatif sur la productivité. L'objectif de la firme n'est plus alors d'atteindre le niveau d'absence le plus faible possible, ni de réduire l'indemnisation de la maladie à la portion congrue. Les auteurs montrent « que le contrat optimal conduit les entreprises à proposer un salaire strictement supérieur aux indemnités de maladie offertes [mais également] des indemnités strictement positives » pour éviter d'inciter les salariés à une assiduité systématique (Chatterji & Tilley, 2002, p. 685).

2 Mesure du présentéisme

Bien que le concept de présentéisme se soit développé dans la continuité des recherches sur l'absence et ait intégré progressivement ses modèles théoriques, ces deux problématiques présentent des différences substantielles en matière de mesure. Tandis que les travaux sur l'absence exploitent souvent des données quantitatives fournies par les employeurs, les articles sur le présentéisme reposent presque tous sur l'interrogation directe des salariés. À tel point que, comme Claes (2011), certains auteurs considèrent que seules les données déclaratives autorisent la mesure du présentéisme, puisque seuls les salariés peuvent savoir s'ils ont effectivement travaillé étant malades ou pas. *A priori*, on voit mal comment il pourrait en être autrement. Il serait certes possible d'envisager des dispositifs lourds qui consisteraient à évaluer chaque matin l'état physique et/ou psychologique d'un échantillon de salariés au travers d'une auscultation

médicale. Pour tout dire, la littérature n'exclut pas de telles enquêtes : Yamamoto et al. (2009) proposent d'associer aux outils existants, la mesure de biomarqueurs liés à l'épuisement. Ils indiquent par exemple que la variabilité de la fréquence cardiaque ou le taux de cortisol salivaire peuvent constituer des indicateurs efficaces de problèmes physiologiques liés au stress et à l'épuisement.

À notre connaissance, une seule étude intègre des données d'absences non-déclaratives recueillies directement auprès de l'employeur pour identifier les présentéistes. Pour Kivimäki et al. (2005), sont présentéistes les individus qui déclarent à la fois être en mauvaise santé⁵ et qui n'ont pas été absents pendant trois ans, d'après les relevés de l'employeur. McKevitt et al. (1997) évoquaient déjà cette possibilité de définir le présentéisme en fonction d'un niveau d'absence « artificiellement faible » pour les salariés concernés. Ils estiment que le niveau d'absence très faible des médecins comparativement à d'autres types de métiers constitue une indication sur leur présentéisme, voire une mesure brute.

Mais le fait est que la seule solution véritablement envisagée jusqu'ici dans la littérature repose sur l'exploitation de données déclaratives. Il nous semble utile de nous attarder sur les méthodes utilisées, non seulement parce que l'objet principal de cette thèse est d'introduire une nouvelle conception de la mesure, mais également parce que l'analyse des méthodologies permet d'éclairer les résultats empiriques que nous synthétiserons dans les sections 3 et 4. Nous aurons l'occasion de détailler les limites de ces méthodes, souvent reconnues par les chercheurs eux-mêmes (cf. 2.3). Mais auparavant, une phase purement descriptive doit aider le lecteur à découvrir les méthodes classiques de production des données : le questionnement direct (2.1) et l'usage d'échelles psychométriques (2.2).

2.1 Questionnement direct

La méthode la plus largement utilisée pour recueillir des informations sur le présentéisme est le questionnement direct. Plus rarement, le présentéisme est déduit du croisement de deux questions différentes, l'une portant sur le nombre d'absences, l'autre sur le nombre d'épisodes de maladie au cours de l'année. Les présentéistes sont alors

5. La santé est évaluée grâce au *General Health Questionnaire*, qui est un outil classique et reconnu. Le résultat obtenu est un score sur 10, la note maximale signifiant un mauvais état de santé. Dans l'étude de Kivimäki et al. (2005), les individus ayant un score supérieur à 4 sont considérés comme étant en mauvaise santé.

ceux qui n'ont pas eu d'absence malgré un épisode de maladie (Agudelo-Suárez et al., 2010). Mais le plus souvent, les salariés sont directement interpellés sur leur propension à venir travailler étant malades, comme dans les deux questions suivantes, utilisées respectivement par Hansen et Andersen (2009) et Johns (2011) :

*How many times during the last 12 months have you gone to work even though it would have been reasonable to take sick leave ?*⁶

*How many days did you go to work in the past six months even though you were sick or not feeling well ?*⁷

Dans les articles étudiés, les questions directes interpellant les individus sur leur présentéisme comportent au moins trois éléments que nous nous proposons de décrire à présent : une description du comportement de présentéisme, une période de rappel et une « unité de mesure ».

Description du comportement. Le choix d'une formulation adaptée est importante pour aider l'interviewé à comprendre l'objet de la question et à y apporter une réponse précise. Deux approches dominant et correspondent à deux définitions du présentéisme. D'une part, dans la plupart des cas et comme dans la première question citée ci-dessus, trois éléments cohabitent : le fait de se rendre à son travail (préféré au verbe direct « travailler »), le fait d'être malade et une règle d'évaluation de la gravité de cette maladie, souvent le seuil de déclenchement de l'absence-maladie. Dans certains cas, cette référence à l'arrêt-maladie est simplement supprimée (Demerouti et al., 2009 ; Johns, 2011). Caverley et al. (2007), puis MacGregor et al. (2008), ajoutent l'idée d'injonction : « *because you felt you had to* »⁸.

D'autre part, certains auteurs abordent le présentéisme sous l'angle des pertes de productivité au travail liées à un problème de santé donné (Koopman et al., 2002 ; Lofland et al., 2004 ; Schultz et al., 2009 ; Zhang et al., 2012). Cette acception inclut par exemple

6. « Combien de fois au cours des 12 derniers mois vous êtes-vous rendu au travail alors qu'il aurait été raisonnable de prendre un arrêt-maladie ? » Il y a six catégories de réponses possibles : « Jamais », « Une fois », « 2 ou 3 fois », « 4 ou 5 fois », « 5 à 10 fois », « 11 fois ou plus ».

7. « Combien de jours au cours des 6 derniers mois vous êtes-vous rendu au travail alors que vous étiez malade ou que vous ne vous sentiez pas bien ? » La réponse donnée est numérique.

8. « Durant les 12 derniers mois, combien de jours avez-vous travaillé malgré une maladie ou une blessure, parce que vous vous y êtes senti obligé ? » (MacGregor et al., 2008).

des affections chroniques (allergies, arthrite, douleurs dorsales, etc.) qui, sans nécessiter un arrêt de travail, diminuent (même marginalement) la productivité d'un salarié. À l'instar de Johns (2010), nous estimons que cette définition, propre au domaine médical, est trop restrictive. Toutefois, les développements proposés par cette frange de la littérature sont parfaitement intégrables à notre travail. Dans cette perspective, les techniques de mesure se focalisent aussi sur la baisse de productivité. Ainsi, par exemple, la question posée par Letvak et al. (2012) :

*How much did your physical or emotional health problems affect your productivity while you were working ?*⁹

Dans le même esprit, Zhang et al. (2012) invitent les interviewés à comparer la durée normale indispensable pour l'exécution de leurs tâches professionnelles quotidiennes à la durée qui a effectivement été nécessaire en cas de maladie.

Période de rappel. Le plus fréquemment, les auteurs invitent les interviewés à se remémorer les épisodes de présentéisme qu'ils ont connus au cours des 12 derniers mois (Hansen & Andersen, 2009 ; Taloyan et al., 2012, et bien d'autres). D'autres fixent des limites plus courtes de 6 mois (Gosselin et al., 2013 ; Johns, 2011) ou de 14 jours (Letvak et al., 2012). Plus surprenantes sont les questions qui ne font référence à aucune limite temporelle, comme McKevitt et al. (1997) ou Krohne et Magnussen (2011)¹⁰. Ponctuellement des questions complémentaires peuvent être utilisées dans la suite de l'enquête à des fins de contrôle : ces questions contournent l'effort de rappel en misant sur une actualisation du comportement dans l'esprit des interviewés : « *If you woke in the morning with a streaming cold and a headache would you take the day off work ?* »¹¹.

Unité de mesure. Dans les cas les plus précis, les auteurs interrogent les salariés sur le nombre de jours travaillés étant malade (Johns, 2011 ; Taloyan et al., 2012). Les données en jours semblent avoir deux avantages. D'une part, elles ouvrent la possibilité de calculer un taux de présentéisme et de le comparer au « taux d'absentéisme » tradition-

9. « À quel point vos problèmes de santé physique ou psychologique ont-ils affecté votre productivité lorsque vous avez travaillé ? » La réponse est donnée sur une échelle de Likert à cinq niveaux.

10. « Have you ever continued to work where it might have been better to take sick leave ? » (McKevitt et al., 1997).

11. « Si vous vous étiez réveillé ce matin avec un rhume et un mal de tête, auriez-vous pris un jour d'arrêt-maladie ? » (McKevitt et al., 1997).

nel, dans les études qui analysent conjointement les deux phénomènes. D'autre part, elles se prêtent aux outils d'analyse quantitative et permettent ainsi d'aller au-delà des régressions logistiques classiquement utilisées. Mais dans la majorité des enquêtes, la question du présentéisme est simplement formulée en termes d'épisodes de présentéisme : « *Has it happened [...] ?* »¹² ou « *How many times [...] ?* »¹³. Les réponses apportées sont soit des catégories détaillant le nombre d'épisodes, soit binaires (présentéiste ou pas). Certaines enquêtes cumulent les deux approches (Johansen, 2012), avec la même double-mesure que la littérature sur l'absence : la gravité du phénomène (durée en jours) et sa fréquence (nombre d'épisodes).

Enfin, la frange de la littérature qui s'intéresse au présentéisme en termes de perte de productivité liée à l'état de santé, cherche principalement à évaluer le nombre d'heures perdues par jour (Zhang et al., 2012). Les salariés sont généralement invités à donner une estimation de la perte de productivité due à leur état de santé (Johns, 2010), autrement dit de leur degré d'inefficacité imputée au présentéisme. La conversion en heures « improductives » se fait dans un second temps (Schultz et al., 2009). Ce degré d'inefficacité peut être mesuré, soit sous la forme d'une échelle à plusieurs paliers (Letvak et al., 2012), soit en interrogeant directement le salarié sur son pourcentage d'improductivité estimé (Schultz et al., 2009).

2.2 Échelles psychométriques

Plus d'une douzaine d'échelles psychométriques sont également disponibles pour le présentéisme (Lofland et al., 2004). Étant généralement issues du domaine médical ou de la psychologie du travail, ces échelles envisagent le présentéisme suivant la définition majoritaire de ces disciplines, c'est-à-dire à l'aune de la perte de productivité relative à un problème de santé. Pour illustrer la nature de ces outils, nous avons sélectionné deux échelles traitant spécifiquement du présentéisme, dont la validité et la pertinence ont été testées (Schultz & Edington, 2007) et qui ont été maintes fois exploitées dans la littérature.

12. « Est-il arrivé au cours des 12 derniers mois que ayez travaillé en ayant l'impression que vous auriez réellement dû prendre un arrêt de travail en raison de votre état de santé ? » (Aronsson & Gustafsson, 2005).

13. « Combien de fois, au cours de 12 derniers mois, avez-vous travaillé bien que votre état de santé nécessitait un arrêt-maladie ? » (Elstad & Vabø, 2008).

Work Limitations Questionnaire. Lerner et al. (2001) proposent un instrument d'auto-évaluation permettant de mesurer à quel point des problèmes chroniques de santé limitent la capacité d'effectuer les tâches professionnelles. Le questionnaire est composée de 25 items (cf. encadré n° 1) évalués sur une échelle de Likert à cinq niveaux. Ces items sont répartis en quatre aspects pour lesquels un état de santé dégradé pourrait être invalidant : le temps d'exécution des tâches, l'état physique, l'état psychologique (ou relationnel) et la baisse de la production. Le résultat final est un score compris entre 0 et 100, indiquant le degré de « limitation » des performances. Le présentéisme n'est donc pas abordé directement, mais est déduit de l'ensemble des inconvénients pratiques inhérents au fait de travailler étant malade, par exemple pouvoir se tenir debout, se concentrer, etc.

1 Items du *Work Limitations Questionnaire*

Time demands	Output demands
Work required hours	Handle workload
Get going beginning of work day	Work fast enough
Start on work soon after arriving	Finish work on time
Work without breaks or rests	Work without mistakes
Stick to routine or schedule	Do all you're capable of
Mental demands	Physical demands
Think clearly	Walk or move around work locations
Work carefully	Sit, stand, stay in one position
Concentrate on work	Repeat motions
Lose train of thought	Bend, twist, or reach
Easily read or use eyes	Use handheld tools, equipment
Speak in person or on phone	Lift, carry, move heavy objects
Help others to work	
Control temper	

Stanford Presenteeism Scale. Après avoir mis en place une échelle de présentéisme à 32 items, Koopman et al. (2002) élaborent une échelle de taille plus restreinte pouvant être utilisée dans un questionnaire plus large. Ils proposent six items (cf. encadré n° 2) destinés à mesurer le degré de présentéisme, et cherchent à évaluer aussi

bien le processus de travail que la production en elle-même. Pour chaque item, l'interviewé se positionne sur une échelle de Likert à cinq niveaux, suivant qu'il est en total désaccord (1) ou en total accord (5) avec l'affirmation avancée. Après calcul, on obtient un score de présentéisme sur 30. Une nouvelle fois, l'idée n'est pas d'évoquer le comportement de présentéisme en lui-même mais d'en mesurer les conséquences dans le travail quotidien.

2 Items de la *Stanford Presenteeism Scale*

Because of my illness, the stresses of my job were much harder to handle
 Despite having my illness, I was able to finish hard tasks in my work
 My illness distracted me from taking pleasure in my work
 I felt hopeless about finishing certain work tasks, due to my illness
 At work, I was able to focus on achieving my goals despite my illness
 Despite having my illness, I felt energetic enough to complete all my work

2.3 Limites des méthodes utilisées

Jusqu'ici, les instruments de mesure du présentéisme exploités dans la littérature ont simplement été décrits. Il convient à présent d'y porter un regard plus critique. D'après nous, et d'après les auteurs eux-mêmes (Elstad & Vabø, 2008 ; Johansson & Lundberg, 2004), cinq limites doivent être soulevées. Les deux premières sont les limites classiquement évoquées des données déclaratives et concernent davantage l'interviewé : le biais de rappel et le biais de désirabilité sociale (Yamamoto et al., 2009). Ensuite, deux questions peuvent se poser respectivement sur le choix des seuils d'identification des présentéistes et sur la grande variabilité des résultats obtenus. Enfin, on peut souligner la lourdeur de ces dispositifs, notamment lorsqu'il s'agit de permettre à une entreprise de mesurer le présentéisme en son sein.

Effort cognitif de l'interviewé. Le questionnement direct sur le présentéisme implique d'abord un véritable effort cognitif pour l'interviewé, ce qui peut être une source de biais dans les résultats. Deux difficultés peuvent être observées. D'une part, on demande à la personne un effort de rétrospection très conséquent (Johns, 2010) : dans la plupart des cas, elle est appelée à se souvenir de faits qui s'étalent sur l'ensemble

de l'année écoulée. Paradoxalement, les créateurs d'échelles psychométriques sont plus attentifs à cela : la plupart d'entre elles n'impliquent qu'un rappel sur quelques semaines, et Lerner et al. (2001) conseillent même de limiter l'effort de mémorisation à deux semaines. Il semble en effet que se rappeler d'une absence-maladie, rendue tangible par le fait d'avoir consulté un médecin ou d'être resté chez soi, est plus aisé que de se remémorer des épisodes de présentéisme (Gosselin et al., 2013). Au final, la réponse dépend plus d'une impression que d'un décompte effectif des épisodes.

Ce biais de rappel est renforcé par une question qui fait reposer la réponse sur le regard subjectif des intéressés quant à leur propre santé (Hansen & Andersen, 2009). Dans la plupart des cas, les questions tentent de réduire cette subjectivité par l'insertion d'un critère indicatif : l'absence-maladie. Si l'état de santé de l'individu légitime un arrêt-maladie et qu'il est au travail, alors c'est un présentéiste. Certains auteurs, tout en étant conscients que le seuil de déclenchement de l'arrêt-maladie est nécessairement différent selon les individus, intègrent ce critère à cause de la fréquence de son usage dans la littérature (Johansen, 2012). Pauly et al. (2008), font le même constat de subjectivité concernant l'évaluation par un salarié de sa perte de productivité liée à la maladie, d'autant plus que l'effort à fournir n'est pas toujours réparti de façon homogène sur une journée de travail.

Biais de désirabilité sociale. Autre risque classique, les salariés peuvent être amenés à déclarer ce que l'on attend d'eux. Hansen et Andersen (2008) admettent que l'une des limites essentielles des données déclaratives est le risque de sous-estimation des absences et d'exagération du présentéisme. Ce risque est particulièrement élevé dans les enquêtes en entreprise. Même s'il dispose de garanties suffisantes sur les raisons de l'enquête ou sur l'utilisation des résultats, un salarié sera sans doute conduit à répondre avec prudence, voire à donner des réponses qu'il sait en accord avec les attentes que l'organisation fait peser sur lui. Mais quel que soit le contexte de l'enquête, les faits déclarés par les interviewés seront toujours tributaires de leur propre perception du phénomène, voire des normes en vigueur dans leur environnement de travail spécifique.

Critères de mesure du présentéisme. Une double remarque peut ensuite être formulée à l'égard de l'analyse des résultats. D'abord, le seuil au-delà duquel un salarié est considéré comme présentéiste dépend largement des choix de l'auteur. Dans certaines études, sont considérés comme présentéistes les salariés qui sont venus travailler

malade au moins une fois (McKevitt et al., 1997), dans d'autres les salariés qui l'ont fait au moins deux fois (Aronsson & Gustafsson, 2005 ; Bergström et al., 2009a), dans d'autres encore ceux qui l'ont fait plus de quatre fois (Nyberg et al., 2008). Certains auteurs ne prennent pas directement parti sur un seuil et parlent davantage d'une augmentation de la fréquence des épisodes de présentéisme (Hansen & Andersen, 2008) ou du nombre de jours de présentéisme (Caverley et al., 2007). Sans doute serait-il nécessaire de clarifier l'objet de la mesure : s'intéresse-t-on simplement aux salariés à qui il est déjà arrivé d'être présentéistes ou à ceux pour qui le présentéisme est devenu une habitude, un trait comportemental ? Les vrais risques du présentéisme semblent plutôt porter sur cette seconde catégorie de salariés.

D'autre part, un problème apparaît dès la formulation de la réponse par l'interviewé. Dans la plupart des cas, la réponse prend une forme catégorielle : le salarié n'a jamais été présentéiste pendant l'année écoulée, l'a été une fois, de deux à cinq fois, etc. Or, le cas des personnes entrant dans la première catégorie, c'est-à-dire celles n'ayant pas été présentéistes pendant l'année, est problématique. Cette catégorie comprend en fait deux types de salariés : d'une part ceux qui ont été malades mais pas présentéistes et d'autre part ceux qui n'ont pas été malades, mais qui auraient été présentéistes s'ils en avaient eu l'occasion. Les individus de cette catégorie sont donc potentiellement des présentéistes, mais ne sont pas considérés comme tels, ce qui biaise les résultats finaux. Johansson et Lundberg (2004) ont conscience de ce biais et contournent le problème en excluant de l'analyse les personnes qui n'ont pas eu à poser le choix du présentéisme.

Quelques enquêtes offrent à l'interviewé la possibilité de répondre qu'il n'a pas été malade pendant l'année écoulée : c'est le cas dans celle qu'exploitent Leineweber et al. (2010) ou dans l'enquête *European Working Conditions Survey* (EWCS). Cependant, un autre problème subsiste, en plus de celui évoqué précédemment. Dans cette enquête EWCS, 37,6% des salariés déclarent avoir été présentéistes au moins une fois au cours des 12 derniers mois, 50,4% déclarent ne pas l'avoir été et 12% déclarent n'avoir jamais été malades. Les individus qui n'ont pas été présentéistes devraient donc avoir été malades au moins une fois, sinon elle répondraient n'avoir jamais été malade. Or, il se trouve que la majorité de ces non-présentéistes (59,7%) déclare, dans une autre question, n'avoir pas eu d'absence au cours de l'année. Le cas de ces personnes est donc paradoxal.

Variabilité de la prévalence. On peut aussi être frappé par la grande divergence, voire l'incohérence (Lofland et al., 2004) des résultats obtenus. Pour illustrer cela, nous avons relevé dans 10 articles différents la part d'individus déclarant avoir été présentéistes au moins deux fois (cf. tableau 1.1). Tous ces articles travaillent sur une population d'un pays nordique (Suède, Norvège, Finlande ou Danemark) et posent une question de forme comparable. On s'aperçoit que les chiffres varient de 30% à 75%, sans se concentrer autour d'une valeur spécifique. Cet exemple n'a certes pas valeur de démonstration, mais il interroge sur la validité des résultats obtenus.

Tableau 1.1 – Prévalence du présentéisme selon différents articles

	Total	Hommes	Femmes
Aronsson et al. (2000)	37%	35%	38%
Aronsson et Gustafsson (2005)	53%	50%	56%
Böckerman et Laukkanen (2009b)	30%	27%	35%
Böckerman et Laukkanen (2010)	33%	29%	40%
Bergström et al. (2009a)	54%	—	—
Bergström et al. (2009b)	55%	—	—
Elstad et Vabø (2008)	75%	—	—
Hansen et Andersen (2009)	58%	55%	60%
Hansen et Andersen (2008)	57%	—	—
Leineweber et al. (2010)	47%	46%	48%

Mesure en entreprise. Le dernier élément que nous voudrions évoquer porte sur la souplesse des méthodes utilisées jusqu'ici. Elles impliquent souvent des procédures lourdes d'échantillonnage, puis de création et d'administration des questionnaires, etc. La majeure partie des articles étudiés dans ce chapitre ont d'ailleurs recours aux résultats d'enquêtes réalisées par les organismes nationaux de statistiques ou par des cabinets indépendants. Lorsque les entreprises s'interrogent sur l'existence du phénomène en leur sein, elles peuvent être découragées par la mise en place de tels dispositifs et par les coûts qu'ils impliquent. L'une des raisons de la profusion d'articles autour de l'absence au travail, est justement la relative simplicité de son analyse. Elle repose sur l'étude des données d'absences disponibles facilement et gratuitement en entreprise. La multiplication des travaux portant sur des micro-données d'entreprise a ainsi contribué à l'extension des connaissances autour de l'absence au travail. Le présentéisme

demeure quant à lui assez méconnu : une raison parmi d'autres est sans doute son invisibilité statistique et la difficulté de sa mesure en entreprise, son lieu d'apparition par excellence.

3 Études empiriques sur les causes du présentéisme

Cette description des méthodes de mesure et le regard critique que nous y avons porté constituent des informations précieuses pour la lecture de la section à venir. Nous y dresserons un état des lieux des connaissances déjà accumulées autour des facteurs de présentéisme, en faisant nôtre la distinction entre facteurs liés à l'individu et facteurs liés à l'emploi, introduite par Aronsson et Gustafsson (2005) puis reprise par d'autres auteurs (Baker-McClearn et al., 2010 ; Hansen & Andersen, 2008 ; Johns, 2010). Bien que les informations à disposition soient encore superficielles (Johns, 2010), voire contradictoires sur certains points, il est essentiel de proposer un panorama des motifs qui conduisent un salarié à prendre la décision d'être présentéiste.

Deux « méta-catégories » de présentéisme apparaissent de façon transversale dans ce qui va être dit. Il est un présentéisme qui résulte d'une injonction, d'une obligation émanant de l'organisation (e.g. charge de travail) ou de l'individu lui-même (e.g. besoin financier). Le salarié se sent alors conduit à venir travailler malade alors qu'il n'aurait pas nécessairement choisi de le faire. Mais le présentéisme peut aussi être perçu par les salariés comme une façon de « donner des gages », de prouver leur investissement, par exemple, lorsqu'étant embauchés à durée déterminée, ils anticipent l'obtention d'un contrat fixe. Ces deux aspects se rapprochent sans doute de ce que Gosselin et Lauzier (2011, p. 21) nomment « présentéisme volontaire » et « présentéisme involontaire ». Cette typologie nous paraît également acceptable, quoique les deux catégories soient difficiles à dissocier. Tout en ayant conscience de la transversalité de ces catégories, nous choisissons toutefois d'aborder les causes du présentéisme selon le découpage désormais traditionnel entre causes liées à l'emploi (que nous appelons également organisationnelles, par simplicité) et causes individuelles.

Hansen et Andersen (2008) ont tenté de déterminer lequel de ces deux types de causes a le plus de poids dans la décision du présentéisme (par comparaison du pseudo- R^2 de leurs régressions logistiques) : aux facteurs relatifs à l'individu et à ceux liés au travail, ils ajoutent les attitudes psychologiques de l'individu. Hansen et Andersen (2008)

avancent que cette dernière catégorie de facteurs a très peu d'importance comparative-ment aux deux autres. Les facteurs organisationnels et individuels auraient quant à eux un pouvoir explicatif similaire : pour comprendre le présentéisme il est donc nécessaire d'étudier successivement ces deux types de causes.

3.1 Facteurs individuels

Les facteurs de présentéisme dits individuels sont directement liés à la sphère privée ou à la personnalité de l'individu et non à des contraintes liées au travail. Ils englobent la situation socio-démographique de l'individu, sa santé et ses traits psychologiques.

3.1.1 *Données démographiques*

Dans la plupart des articles, les variables démographiques sont considérées comme de simples variables de contrôle, et peu nombreux sont les auteurs qui en proposent une interprétation détaillée. Les résultats concernant le sexe, l'âge, le niveau de diplôme et la situation familiale des individus, nous semblent pourtant avoir un enjeu important dans la description du présentéisme.

Sexe. D'un point de vue général, si l'on s'arrête à de simples statistiques descriptives sur la prévalence du présentéisme, la quasi-totalité des articles étudiés indiquent que les femmes déclarent davantage être présentéistes que les hommes. Bien entendu, une approche multivariée est incontournable, puisque cette prévalence peut s'expliquer par d'autres facteurs, comme la surreprésentation des femmes dans les catégories d'emploi à risque. Selon Aronsson et al. (2000), les femmes seraient effectivement plus présentéistes que les hommes, mais dans un article publié cinq ans plus tard (Aronsson & Gustafsson, 2005) et portant sur une population similaire (population active suédoise), cette variable du sexe n'était que faiblement significative. Pour Leineweber et al. (2010), travaillant sur les employés de la police suédoise, les femmes viendraient même très fréquemment travailler étant malade : l'effet est particulièrement prononcé parmi les individus qui déclarent être présentéistes plus de 10 fois par an.

Au total, la littérature se partage entre les travaux ne trouvant aucun effet du sexe (Claes, 2011 ; Johansen, 2012 ; Johansson & Lundberg, 2004), ceux qui concluent que les femmes sont plus présentéistes que les hommes (Böckerman & Laukkanen, 2009a,

2009b) et ceux qui voient dans le présentéisme un comportement typiquement masculin. C'est le cas de Simpson (1998) qui indique que le présentéisme, perçu comme une preuve d'engagement dans le travail, devient un véritable critère de sélection lorsqu'il y a concurrence entre certains salariés. D'après l'auteur, cela expliquerait pourquoi si peu de femmes accèdent à des postes d'encadrement, le présentéisme étant plus volontiers pratiqué par les hommes, voire étant un élément de la culture masculine. Ainsi, les jeunes hommes en voie d'accéder à des postes de management, seraient-ils particulièrement concernés. Enfin, Demerouti et al. (2009), dans le contexte d'un hôpital néerlandais, constatent également que les hommes sont plus présentéistes.

Âge. Une part non négligeable des articles portant sur le présentéisme et intégrant la variable d'âge, concluent à une absence de relation (Böckerman & Laukkanen, 2009b ; Claes, 2011 ; Johansen, 2012 ; Kivimäki et al., 2005). Bien qu'en 2000, Aronsson et al. trouvaient un présentéisme déclaré moins élevé pour les catégories d'âge extrêmes (jeunes et seniors), cet effet ne se retrouve plus dans leurs travaux ultérieurs sur la même population (Aronsson & Gustafsson, 2005) et semble donc assez peu robuste. Hansen et Andersen (2008), mais également Gosselin et al. (2013), estiment quant à eux que les individus les plus jeunes ont généralement une propension de présentéisme plus élevée. Pourtant, sur la base d'une étude qualitative (en Nouvelle-Zélande), Dew et al. (2005) rapportent que certains salariés plus âgés expliquent leur présentéisme par la difficulté de retrouver un travail en cas de licenciement : il leur semble nécessaire de montrer leur investissement professionnel et utilisent le présentéisme comme un argument en ce sens. De même, Leineweber et al. (2010) observent que, si l'on s'arrête à une forme extrême de présentéisme (plus de 10 épisodes par an), les salariés âgés de plus de 54 ans sont plus présentéistes, tandis que les formes plus modérées seraient plutôt caractéristiques des classes d'âge intermédiaires.

Situation familiale. Parmi les facteurs individuels les plus étudiés, se trouvent les conflits entre la sphère privée (famille) et la sphère organisationnelle, avec des résultats surprenants mais plutôt convergents. Intuitivement, le présentéisme devrait être plus faible pour les personnes ayant une charge familiale, par exemple ayant des enfants en bas âge. Mais Aronsson et al. (2000) obtiennent des résultats contre-intuitifs : les salariés ayant des enfants à charge auraient un risque de présentéisme plus élevé. Hansen et Andersen (2008) parviennent à la même conclusion et ajoutent que les salariés

envisageant leurs activités professionnelles comme moins éprouvantes que leurs tâches familiales sont aussi les plus présentéistes. Ils attribuent ce résultat au fait que les tâches professionnelles, surtout lorsqu'elles sont intéressantes et gratifiantes, peuvent être préférées aux activités de la vie familiale. Mais la propension à venir travailler étant malade n'est pas seulement liée à cette différence dans la nature des tâches. Le présentéisme est spécifiquement élevé chez les individus pour lesquels les contraintes privées et les exigences professionnelles entrent souvent en conflit (Hansen & Andersen, 2009), comme si le fait d'être présent étant malade constituait un moyen de compenser les heures de travail sacrifiées à des prérogatives familiales. Simpson (1998) estime en revanche que les femmes ayant des enfants identifient souvent les attentes de leur entreprise (heures supplémentaires, par exemple) comme étant incompatibles avec leur vie de famille, et de ce fait sont moins portées au présentéisme que les hommes, pour lesquels un tel conflit est moins patent.

Le fait d'être marié semble également jouer ici : pour Virtanen et al. (2012), les personnes mariées pratiquent plus souvent des journées de travail extensives. Cocker et al. (2011), qui s'intéressent particulièrement aux salariés présentéistes lors d'un épisode de dépression, signalent que les salariés mariés travaillent davantage étant malades.

Niveau de diplôme. Seuls quelques travaux avancent des éléments sur le lien entre le nombre d'années d'étude et le présentéisme, cette variable étant généralement corrélée à la nature du poste occupé ou au niveau hiérarchique. Aronsson et al. (2000) incluent une variable catégorielle sur le type d'études suivies (primaire, secondaire, universitaire, etc.), mais aucun lien avec le présentéisme n'est identifié. Une étude récente de Taloyan et al. (2012), sur un échantillon représentatif de la population active suédoise, trouve que les salariés ayant un niveau de diplôme intermédiaire sont plus présentéistes que les individus détenteurs d'un diplôme universitaire ou de faible niveau. Par ailleurs, Agudelo-Suárez et al. (2010) effectuent une étude comparative entre travailleurs espagnols natifs et travailleurs immigrés. Chez, ces derniers, le fait d'être détenteur d'un diplôme universitaire est un facteur de présentéisme, en particulier lorsqu'ils sont arrivés en Espagne depuis moins de deux ans.

3.1.2 *Nature de la maladie*

Si le présentéisme affecte les salariés différemment selon leur sexe, leur âge ou leur situation familiale, il se manifeste aussi de façon contrastée suivant la maladie qui touche l'individu. Dès lors, « l'affection du travailleur aurait une incidence sur la fréquence et la propension de ce dernier » à être présentéiste (Gosselin & Lauzier, 2011, p. 20). Dans la mesure où la question du présentéisme ne se pose pas lorsque le salarié est en bonne santé, de nombreux auteurs insistent sur le fait qu'un salarié déclarant un mauvais état de santé est aussi plus présentéiste (Aronsson & Gustafsson, 2005 ; Hansen & Andersen, 2009 ; Kivimäki et al., 2005 ; Leineweber et al., 2010), même si ce constat peut paraître trivial.

La littérature regorge de travaux étudiant le comportement présentéiste d'individus touchés par des maladies spécifiques. Certaines affections semblent en effet prédisposer au présentéisme, tandis que d'autres conduiraient davantage à l'absence (Gosselin et al., 2013). De façon générale, les auteurs conviennent du fait que le présentéisme est plus souvent observé chez les personnes souffrant d'affections chroniques (Caverley et al., 2007 ; Schultz & Edington, 2007), qui impliquent, à cause de leur persistance, un retour au travail malgré un état de santé suboptimal.

Il apparaît également qu'un consensus se forme autour de quatre types de maladies plus propices au comportement de présentéisme. D'abord, le cas de la dépression est très souvent évoqué (Letvak et al., 2012 ; Schultz & Edington, 2007). Cocker et al. (2011) consacrent même un article entier au présentéisme durant un épisode de dépression. D'après eux, parmi les personnes dépressives, celles qui sont touchées de façon plus bénigne seraient plus susceptibles de travailler étant malades. C'est également le constat d'Aronsson et al. (2000). L'entourage des personnes dépressives peut contribuer à un retour anticipé au travail : les personnes mariées (Cocker et al., 2011) ou ayant un suivi psychologique (MacGregor et al., 2008) seraient plus présentéistes. Dans ce cas, le travail est alors envisagé comme un élément favorable à la guérison (Aronsson & Gustafsson, 2005).

Parmi les autres pathologies propices au présentéisme figure la migraine (Goetzel et al., 2004 ; Schultz & Edington, 2007). On peut penser qu'il s'agit typiquement d'une pathologie qui, sans être insignifiante, paraît aux salariés ne pas justifier un arrêt de travail. Mais la migraine peut se révéler très invalidante au poste de travail et les pertes économiques corollaires pourraient être assez élevées (Goetzel et al., 2004). Les aller-

gies, en particulier saisonnières, sont également citées (Gosselin et al., 2013). Schultz et Edington (2007) évoquent par exemple une étude évaluant à 2,3 heures par jour la perte de travail due à une rhinite allergique. Enfin, les douleurs dorsales (Aronsson et al., 2000), thoraciques (Caverley et al., 2007) ou plus largement les troubles musculo-squelettiques (Hansen & Andersen, 2008 ; Letvak et al., 2012), font partie de ces maladies qui vont de pair avec le présentéisme.

Une particularité de ces maladies pourrait expliquer leur relation avec le présentéisme. Il est possible qu'un médecin ne considère pas qu'un arrêt-maladie soit nécessaire, par exemple pour une allergie ou pour une migraine, ou qu'une absence ponctuelle n'améliorerait pas l'état général du patient. De ce fait, l'individu serait amené à travailler étant malade.

3.1.3 *Profil psychologique*

En plus des variables démographiques et des éléments portant sur la santé, la littérature donne quelques indications sur les attitudes, le caractère des individus susceptibles de travailler étant malades, permettant de dresser un « profil psychologique » des présentéistes. Ce profil repose sur trois caractéristiques importantes.

Premièrement, le présentéisme serait plutôt le fait de personnes ne sachant pas opposer un refus aux attentes d'autrui (Aronsson & Gustafsson, 2005). Dans le contexte professionnel, cela se traduit par une incapacité à décliner la prise en charge de nouvelles activités ou de nouveaux projets. Aronsson et Gustafsson (2005, p. 964) expriment cela par le terme anglais *boundarylessness*, c'est-à-dire le fait de ne pas savoir se fixer de limites dans l'acceptation de nouvelles tâches. Bien entendu, en première intention cela peut être lié à l'intérêt intrinsèque des activités et à la satisfaction qu'en tire l'individu (Gosselin & Lauzier, 2011), mais peut conduire à privilégier les tâches professionnelles au détriment de la santé. Ce trait psychologique peut aussi être associé au surinvestissement (*over-commitment*) dans le travail. C'est ce que mesurent Hansen et Andersen (2008) en interrogeant les salariés sur leur capacité à se détacher mentalement de leurs activités professionnelles lorsqu'ils rentrent chez eux ou sur leur impression de se sacrifier pour leur travail. Les individus les plus engagés dans leur travail sont aussi les plus présentéistes. Dans les formes extrêmes, le terme de *workaholism* (Johns, 2011), qui caractérise une forme néfaste de dépendance au travail, correspond très bien à l'attitude décrite dans la littérature.

Deuxièmement, lorsque le contexte professionnel soumet les individus à un fort stress, ceux-ci opteraient souvent pour une présence inconditionnelle au travail (Gosselin et al., 2013; Leineweber et al., 2010; MacGregor et al., 2008). Intuitivement, on aurait pu croire qu'un contexte professionnel de crise augmenterait la propension des salariés à se retirer afin de se protéger. Mais Elstad et Vabø (2008) confirment que lorsque le stress augmente (véhiculé par des éléments organisationnels comme les contraintes de temps ou la charge de travail), le présentisme augmente plus fortement que le niveau d'absence. Les salariés auraient donc tendance à faire face à la charge de travail, même lorsque leur état de santé n'est pas favorable. Mais pour Johns (2011), ceci n'est vrai que dans une certaine limite. Ils étudient un trait de personnalité appelé le « neuroticisme », c'est-à-dire la propension de certains salariés à déclencher une forte réaction émotionnelle dans une situation stressante. Selon eux, les individus les plus sensibles au stress seraient moins susceptibles de venir travailler étant malades, mais s'ils décident de travailler malgré tout, l'impact négatif sur la productivité est bien plus élevé.

Troisièmement, on peut noter le résultat de Hansen et Andersen (2008) sur l'influence de l'opinion personnelle des salariés vis-à-vis de l'absence-maladie, dont ils sont les seuls à évaluer l'effet. Ils souhaitent déterminer si l'avis d'un salarié sur l'absence-maladie est plutôt « conservateur » ou plutôt « libéral ». Dans le premier cas, les individus envisagent l'absence comme un dernier recours et se positionnent plutôt sous l'angle d'une « morale du devoir », tandis que dans le second cas, ils considèrent l'absence comme un droit pour tout salarié. La mesure passe par 7 items, mettant les salariés en situation. On leur demande par exemple s'il serait raisonnable de prendre un arrêt-maladie dans la situation suivante : « X a une température de 38,2°C et ne se sent pas bien. Cependant, X sait qu'il y a trop peu de collègues pour que ses tâches soient prises en charge ». D'après Hansen et Andersen (2008), les personnes ayant un regard « conservateur », plutôt intransigeant sur l'absence, s'appliquent à eux-mêmes leur vision des choses et travaillent davantage étant malades. Ceci conforte l'idée de Johns (2010) que le comportement de présentisme doit être sous-tendu par un caractère persévérant, tenace.

3.2 Facteurs liés à l'emploi

Le présentéisme se matérialise dans le comportement d'une personne elle-même insérée dans un environnement de travail. Jusqu'ici nous avons observé les facteurs de présentéisme caractéristiques de l'individu : un homme ou une femme d'un certain âge, que certains motifs privés ou traits de caractère conduisent à accomplir leurs tâches professionnelles nonobstant la maladie. Mais ces tâches professionnelles se déroulent dans un cadre organisationnel spécifique, sont marquées par des relations inter-individuelles, sont formalisées par un contrat de travail, sont influencées par un contexte conjoncturel, etc. Ces multiples dimensions du travail salarié comportent des facteurs suscitant également des comportements présentéistes chez l'individu. Ceux-ci occupent une place centrale dans la littérature.

3.2.1 *Éléments contractuels*

Le premier contact de l'individu avec son employeur se matérialise par la signature d'un contrat de travail. Trois éléments constitutifs de ce contrat sont mis en évidence par la littérature comme des facteurs possibles de présentéisme.

Nature du contrat. Dans de nombreuses entreprises, l'embauche d'un salarié en contrat à durée indéterminée (CDI) est précédée par un contrat à durée déterminée (CDD). Cette période est souvent appréhendée par les deux parties comme un temps de découverte réciproque et comme l'antichambre d'un poste pérenne. Les salariés dans cette situation sont donc incités à manifester leur motivation et leur efficacité par un réel engagement dans le travail. Ichino et Riphahn (2005) comparent aussi le niveau d'absence de salariés pendant et après leur période d'essai contractuelle et relèvent que le nombre de jours d'absence par semaine augmente significativement une fois que l'emploi est définitivement acquis, c'est-à-dire après la période d'essai, en particulier pour les hommes. Les conclusions de plusieurs articles laissent penser que les salariés considèrent le présentéisme comme un moyen de mettre en avant leur investissement (Caverley et al., 2007) et leur capacité à occuper des postes en responsabilité (Simpson, 1998).

À ce titre, l'analyse de Grinyer et Singleton (2000) paraît judicieuse. Ils envisagent le présentéisme comme un choix entre deux risques : un risque de passer à côté d'oppor-

tunités professionnelles s'ils s'absentent et un risque pour leur santé s'il ne se soignent pas. Le risque de manquer des opportunités étant beaucoup plus fort pour les salariés en CDD, une simple analyse coût-avantage conduit ceux-ci à privilégier le présentéisme, quelles que soient leurs caractéristiques individuelles. Pour ces salariés, s'absenter lorsqu'on est malade est assimilé à une « conduite à risque » dans l'optique de la pérennisation de l'emploi. Les résultats de Hansen et Andersen (2008), Caverley et al. (2007) ou encore Agudelo-Suárez et al. (2010) vont nettement en ce sens : plus l'emploi est précaire, plus le présentéisme est élevé. Hansen et Andersen (2008) ajoutent que ceci est encore plus vrai lorsque le taux de chômage est élevé, c'est-à-dire lorsque les salariés anticipent une difficulté à retrouver un emploi en cas de licenciement. Enfin, Johns (2011) interroge directement les salariés sur leur sentiment à propos de la pérennité de leur emploi. Ceux qui perçoivent leur emploi comme précaire (même à tort) sont aussi les plus présentéistes.

Temps de travail. Les travaux de Böckerman et Laukkanen (2009a, 2009b, 2010) incluent tous la question du temps de travail. Ils soulignent que les salariés pour lesquels le temps de travail hebdomadaire contractuel correspond à leurs souhaits initiaux sont moins présentéistes. Les auteurs perçoivent cette variable comme une sorte de proxy d'un temps de travail équilibré : trouver le quota hebdomadaire qui convient le mieux au salarié diminuerait donc le présentéisme, comme d'ailleurs l'absence.

Aronsson et Gustafsson (2005) s'attendaient intuitivement à trouver que les salariés ayant une plus grande autonomie et une certaine liberté dans l'organisation du temps de travail seraient plus présentéistes. En cas de maladie, ils pourraient par exemple privilégier certaines tâches plus aisées ou écourter leur journée de travail. Mais les auteurs obtiennent le résultat inverse, que confirment Hansen et Andersen (2008). Les salariés exerçant un contrôle sur leur rythme de travail sont moins présentéistes, comme s'ils acceptaient plus aisément de prendre une journée d'absence, sachant qu'ils pourront rattraper leur retard une fois rétablis.

Par ailleurs, les salariés à temps plein seraient davantage susceptibles d'être présentéistes que ceux travaillant à temps partiel (Böckerman & Laukkanen, 2009b), notamment chez les femmes (Böckerman & Laukkanen, 2010). Pour les auteurs, cela serait lié au fait que les postes à temps plein sont moins aisément remplaçables que les postes à

temps partiel. Les salariés, sachant que leur travail ne sera pas pris en charge pendant leur absence-maladie, hésiteraient donc à y consentir.

Enfin, un autre constat est récurrent : les salariés qui font régulièrement des heures supplémentaires sont plus présentéistes (Böckerman & Laukkanen, 2009b ; Caverley et al., 2007). C'est aussi le cas pour ceux qui ont un temps de travail hebdomadaire très élevé : plus de 45 heures (Hansen & Andersen, 2008) ou plus de 48 heures (Böckerman & Laukkanen, 2009b).

Salaire. Quelques travaux intègrent le niveau de salaire des individus comme un facteur de présentéisme. La variable de salaire est alors utilisée comme un proxy du besoin financier de la personne (Aronsson et al., 2000). En effet, l'une des causes intuitives du présentéisme est la nécessité d'avoir un revenu : les personnes qui ont le plus besoin d'argent seront sans doute incitées à venir travailler malgré la maladie, notamment si l'indemnisation de l'absence n'est pas intégrale (sur les sanctions financières de l'absence, cf. 3.2.4). Tout en contrôlant les variables en interaction avec le salaire, quelques articles intégrant cette variable trouvent une relation entre le fait d'avoir un salaire faible ou moyen et le présentéisme (Aronsson & Gustafsson, 2005 ; Johansen, 2012). Pour Hansen et Andersen (2008), en revanche, la relation n'est pas significative.

3.2.2 Contenu de l'emploi

D'autres facteurs de présentéisme peuvent être décelés dans le contenu même de l'emploi. De nombreux articles abordent la question de manière générale en termes de quantité de travail, mais une analyse plus fine permet de mettre en évidence certains types de tâches plus spécifiquement associées au présentéisme.

Charge de travail. Il s'agit presque d'un lieu commun de la littérature sur le présentéisme : les salariés devant faire face à une lourde charge de travail sont plus prompts à venir travailler étant malades (Caverley et al., 2007 ; Elstad & Vabø, 2008 ; Hansen & Andersen, 2009 ; McKevitt et al., 1997). Cette remarque est particulièrement vraie lorsque les tâches s'accumulent en l'absence du salarié (Aronsson et al., 2000) ou lorsqu'elles doivent être réalisées dans une certaine limite temporelle (Caverley et al., 2007). L'existence d'une échéance, par exemple l'obligation de présenter un projet à une certaine date, incite clairement un salarié malade à être présentéiste, en par-

ticulier lorsqu'il reste peu de temps avant le terme fixé (Hansen & Andersen, 2008). Pour Claes (2011), les salariés identifient le présentéisme comme la seule solution pour éviter une augmentation additionnelle de cette pression temporelle. En effet, s'accorder une absence-maladie et délaissier son travail augmenterait encore la difficulté de tenir ses échéances. Dans l'étude de Claes (2011), il s'agit d'un résultat particulièrement robuste et qui se vérifie dans tous les pays étudiés. D'après Elstad et Vabø (2008), mais aussi Leineweber et al. (2010), la coexistence d'une charge de travail importante et de contraintes temporelles strictes génère une atmosphère de travail stressante et contribue à la progression du présentéisme.

Nature des tâches. Peu nombreux sont les travaux portant sur des salariés issus de secteurs d'activité différents ou parvenant à identifier des catégories d'emploi plus présentéistes que d'autres. À en croire Aronsson et al. (2000, 2005), les métiers les plus propices au présentéisme seraient liés aux besoins fondamentaux de la personne : ils citent ainsi les métiers au contact de jeunes enfants (professeurs des écoles, nourrices) ou de malades (infirmières, médecins, sage-femmes, aides à domicile). Les secteurs du soin hospitalier ou de l'aide à la personne semblent être en première ligne : de nombreux auteurs en ont d'ailleurs eu l'intuition, puisqu'un grand nombre de travaux s'intéressent à ces métiers. Les médecins ont aussi fait l'objet d'une des premières études sur le présentéisme. McKevitt et al. (1997) les y décrivent comme fondamentalement convaincus de l'importance de leur tâche : celle-ci ne peut ni être reportée dans le temps, ni être déléguée à des collègues eux-mêmes débordés par le nombre de patients. Cette constatation peut être étendue aux infirmières (Demerouti et al., 2009) ou aux métiers de soin des personnes âgées (Elstad & Vabø, 2008) : tous ces salariés estiment que leur présence et leurs tâches (toilette, nourriture, acte médical) sont vitales pour leurs patients et, par conséquent, travaillent même lorsqu'ils sont eux-mêmes malades. Böckerman et Laukkanen (2009a) confirment un autre résultat d'Aronsson et al. (2000) : le présentéisme serait davantage le fait de salariés du secteur public. Des analyses qualitatives expliquent ceci par le souci des salariés d'assurer la continuité du service public et de sauvegarder son image (Baker-McClearn et al., 2010 ; Dew et al., 2005). C'est notamment le cas pour les soignants qui souhaitent garantir un accueil inconditionnel des patients, même lorsque leur environnement de travail prend des airs de « champ de bataille » (Dew et al., 2005, p. 2278). Dans le même esprit, les salariés dont les métiers impliquent d'importants coûts d'opportunité, c'est-à-dire par exemple lorsque

le chiffre d'affaires du jour d'absence ne peut être rattrapé le lendemain, sont aussi plus susceptibles d'être présentéistes (Pauly et al., 2008).

Enfin, on peut encore noter que certaines circonstances plus spécifiques rendent le présentéisme presque trivial. C'est par exemple le cas sur les plateformes pétrolières maritimes où les salariés travaillent par postes de 14 jours et doivent être héliportés à terre en cas de maladie durant cette période. De ce fait, le seuil de déclenchement de l'absence-maladie est plus élevé et les salariés demeurent à leur poste étant malades (Krohne & Magnussen, 2011).

Responsabilité managériale. La thématique de la responsabilité managériale est à mettre en lien avec ce qui vient d'être dit : les salariés sont davantage présentéistes lorsque leur présence est nécessaire à d'autres personnes, par exemple à leurs subordonnés (Caverley et al., 2007 ; Gosselin et al., 2013). Mais l'influence de la responsabilité d'équipe sur le présentéisme a peu été étudiée. Intuitivement, on pourrait penser que les managers considèrent leur présence comme une condition nécessaire au bon déroulement des activités. En ce sens, les managers optent sans doute davantage pour le présentéisme, par égard pour leurs équipes de travail. Les résultats de trois articles suggèrent qu'il s'agit en effet d'un élément important. Parmi les facteurs identifiés par Caverley et al. (2007) comme étant les principaux contributeurs au présentéisme, figure le sens de la responsabilité, notamment vis-à-vis d'une équipe. Hansen et Andersen (2008), quant à eux, incluent clairement une variable binaire indiquant si le salarié à un statut de manager (*supervisor*) ou pas. Cette population a une propension de présentéisme nettement plus élevée. C'est également le résultat obtenu par Johansen (2012), sur un échantillon de salariés suédois et norvégiens. Ceux qui ont une responsabilité managériale, sont à la fois moins absents et plus présentéistes. Les auteurs expliquent cela par le fait qu'ils se sentent nécessaires au sein de leurs équipes, mais aussi par la satisfaction liée aux activités managériales. Indiquons enfin que dans leur étude sur la police suédoise, Leineweber et al. (2010) obtiennent un résultat contradictoire : les subordonnés opteraient davantage que les managers pour une forme extrême de présentéisme (plus 10 fois par an).

En s'appuyant sur une définition élargie du présentéisme (pratique régulière des heures supplémentaires), Simpson (1998) indique que ce comportement est caractéristique des managers. Il est considéré comme un moyen d'établir une compétition et de confronter

leur investissement et leur capacité de résistance, dans la perspective de promotions professionnelles. Dans le même esprit, Virtanen et al. (2012) précisent que les salariés ayant une responsabilité importante ont aussi un temps de travail quotidien plus élevé.

3.2.3 Contexte de l'emploi

Bien que la nature des activités quotidiennes ait une influence sur la décision de présentéisme, celle-ci peut aussi être infléchie par l'entourage professionnel d'un salarié. Le groupe dans lequel il se situe, avec ses collègues et son manager, développe en effet ses propres normes et règles de fonctionnement.

Relations avec les collègues. Johns (2010) pointe le caractère social du phénomène en termes de « culture de présentéisme » comme il l'avait déjà fait auparavant pour la culture d'absence (Nicholson & Johns, 1985). Dans cette perspective, l'environnement de travail et le comportement des collègues orientent les décisions des individus en matière d'absence : « les travailleurs observent par exemple le comportement d'absence des autres, ainsi que les réactions de diverses personnes devant ce comportement, et adoptent alors une norme ou un niveau d'absence qui reflète ces observations » (Nicholson & Johns, 1985, p. 398). Une telle culture se manifeste souvent dans des équipes de travail très solidaires où l'absence est évitée par sympathie pour les collègues. Très nombreux sont les travaux appuyant cette idée (Baker-McClearn et al., 2010 ; Caverley et al., 2007 ; Grinyer & Singleton, 2000 ; Johansson & Lundberg, 2004 ; McKevitt et al., 1997). Dew et al. (2005) décrivent le contexte professionnel d'une clinique privée néo-zélandaise : d'un commun accord, les salariés décrivent les relations entre collègues comme étant « familiales ». Le fait d'être présentéiste est directement mis en lien avec le souci de ne pas imposer leur charge de travail à des collègues qu'ils apprécient. Dans le contexte très spécifique des plateformes pétrolières en haute mer, c'est aussi ce que montrent Krohne et Magnussen (2011). La vie commune pendant plusieurs semaines crée de tels liens entre les salariés qu'il devient impensable pour certains de rester à l'infirmerie ou de se faire transporter à terre à cause de tracas de santé, pendant que leurs collègues doivent travailler davantage.

Certains auteurs trouvent une relation négative entre le présentéisme et le soutien social provenant de l'équipe de travail (Hansen & Andersen, 2008 ; Leineweber et al., 2010). Autrement dit, dans les équipes où la solidarité est moins marquée, le présentéisme

pourrait aussi être plus important. Ce résultat n'est pas intuitif : on s'attendrait plutôt à ce que, dans une équipe en conflit, le seuil de déclenchement de l'absence soit atteint plus précocement. Un élément explicatif avancé par les auteurs est le report des tâches : si en cas d'absence, celles-ci s'accumulent sans qu'aucun collègue ne s'en soucie, le présentéisme pourrait émerger, même dans une équipe peu soudée. Caverley et al. (2007) trouvent également, que le manque de confiance vis-à-vis des collègues est un facteur d'augmentation du présentéisme.

Par ailleurs, cette culture du présentéisme peut aussi jouer dans des équipes en forte compétition. Le présentéisme peut véritablement être utilisé comme un instrument destiné à montrer son engagement au travail et pour faire ses preuves à certains postes. Il cristallise alors l'émulation entre certains salariés, notamment dans l'accès aux postes de management (Simpson, 1998).

Relations avec le manager. Ce qui vient d'être dit sur la culture de présentéisme peut aussi s'appliquer aux relations avec le manager : lorsqu'il fait partie intégrante de l'équipe, il peut avoir une influence décisive sur ses subordonnés. Cela se manifeste d'abord dans la qualité de la relation entre managers et subordonnés. Plusieurs situations peuvent être relevées. Pour Caverley et al. (2007), avoir de bonnes relations avec son manager peut constituer un motif de présence inconditionnelle. Les salariés qui déclarent pouvoir compter sur leur manager direct en cas de situation problématique sont ainsi les plus présentéistes. Dans ce cas, le présentéisme est sans doute considéré comme un juste retour des choses envers un manager compréhensif. On en revient alors à la nature familiale de certains contextes de travail. Dans l'article de Krohne et Magnussen (2011), les managers sont perçus comme empathiques et attentifs : ils sont en mesure de trouver des compromis avec les salariés en leur proposant éventuellement d'occuper temporairement un poste plus adapté à leur état de santé.

Mais, paradoxalement, un manque de soutien social de la part du manager (Leineweber et al., 2010) ou l'existence de relations tendues avec le manager (Dew et al., 2005) peut également contribuer à l'augmentation du présentéisme. Le manager est identifié comme celui qui impose des exigences de rendement ou d'efficacité (Böckerman & Laukkanen, 2010) ou celui qui fixe des contraintes temporelles (Claes, 2011) et les salariés peuvent répondre aux exigences du manager par leur présence étant malade. Dew et al. (2005) décrivent le contexte d'une entreprise de fabrication, marquée par des

relations conflictuelles entre les salariés et leurs managers. Ces conflits se cristallisent notamment autour de la justification de l'absence-maladie. L'incitation à être présent ou à écourter son absence a une grande influence sur les salariés, surtout lorsqu'elle provient directement du manager : ils se sentent alors clairement forcés à travailler nonobstant leur état de santé.

Les salariés sont parfois particulièrement attentifs aux comportements de leurs managers et peuvent chercher à les reproduire. Concernant l'absence au travail, Nielsen (2008) indique que la fréquence d'absence des managers de proximité influence directement le comportement d'absence de l'équipe de travail dont ils ont la charge. Ce constat pourrait être transférable au présentéisme. Si l'absence-maladie n'est pas considérée comme un droit par le manager, mais plutôt comme un évènement gênant le déroulement des activités, le présentéisme se manifeste comme une réponse aux attentes du manager (Baker-McClearn et al., 2010). En quelque sorte, le manager a la faculté d'institutionnaliser le présentéisme.

3.2.4 *Politiques RH*

L'environnement relationnel direct des salariés et les normes informelles qui s'y établissent jouent un rôle important dans la décision des salariés d'être absents ou non en cas de maladie. Mais cette sphère de l'équipe de travail est tributaire de décisions prises à un autre niveau organisationnel. Les politiques régulant les situations d'absence ou, plus largement, redéfinissant les choix de l'entreprise en matière de ressources humaines, peuvent aussi avoir des conséquences directes sur le présentéisme.

Coût financier et psychologique de l'absence. Plus haut, nous indiquions déjà que les salariés ayant peu de marges financières étaient plus présentéistes que les autres. Cet effet peut être renforcé par les dispositifs sanctionnant financièrement l'absence. Barmby (2002) souligne à quel point le niveau d'absence d'une entreprise est lié au coût financier de l'absence pour les salariés. Plus la perte de salaire est conséquente, plus la probabilité d'être absent diminue. Baker-McClearn et al. (2010) notent que la question de l'indemnisation des absences est évoquée dans la plupart de leurs entretiens qualitatifs avec les salariés. Certains estiment que l'objectif du service des ressources humaines est avant tout de garantir leur présence au travail, sans égard pour leur état

de santé. D'autres perçoivent les primes de présence comme une bonne raison d'être présentéistes.

Dans cet esprit, Böckerman et Laukkanen (2010) évaluent l'effet d'un dispositif de gestion de l'absence moins restrictif. Dans l'échantillon qu'ils étudient, une part des salariés peut s'absenter pendant trois jours dans l'année sans justification et tout en conservant l'intégralité de son salaire. Ces individus sont nettement moins susceptibles d'être présentéistes, comme si les trois jours octroyés leur offraient une marge de manœuvre permettant de se soigner. Enfin, d'un autre point de vue, Claes (2011) compare l'importance du présentéisme dans des pays ayant des systèmes de sécurité sociale plus ou moins généreux, c'est-à-dire indemnisant plus ou moins bien l'absence-maladie. Ses résultats semblent indiquer que dans les pays les plus généreux, le présentéisme est moins important.

Mais le coût financier de l'absence n'est pas seul en cause. Certains dispositifs ont un poids psychologique important pour les salariés. Ainsi, les entretiens de retour, sont-ils parfois perçus comme de bonnes raisons de venir travailler malgré un état de santé détérioré (Baker-McClearn et al., 2010). L'entreprise étudiée par Grinyer et Singleton (2000) a mis en place un tel dispositif : un entretien avec leur manager est systématiquement imposé aux salariés après quatre épisodes d'absence-maladie. Les auteurs décrivent alors les stratégies élaborées par les salariés pour ne pas atteindre ce seuil. Si certains, en cas d'absence, attendent d'être parfaitement rétablis pour retourner au travail par crainte d'une rechute qui nécessiterait un épisode d'absence supplémentaire, d'autres salariés recourent au présentéisme pour éviter d'atteindre le seuil fatidique.

Possibilité de remplacement. Pour prendre leur décision d'absence ou de présence, les salariés prennent souvent en compte la possibilité d'être remplacé. Certaines entreprises facilitent la prise en charge des activités de l'absent que ce soit par une modification de l'organisation du travail à l'interne ou par le recours à une main d'œuvre extérieure à l'entreprise¹⁴. Les travaux d'Aronsson et al. (2000) montrent que l'impossibilité d'être remplacé est clairement un facteur de présentéisme, ce que confirment Böckerman et Laukkanen (2009b) ainsi que Krohne et Magnussen (2011). Böckerman et Laukkanen (2009a) ajoutent que les femmes seraient particulièrement sensibles à cet argument. Pour Elstad et Vabø (2008), l'insuffisance de personnel contribue à l'aug-

14. Dans ce cas, pour Pauly et al. (2008), les pertes liées à l'absence d'un salarié sont compensées selon le degré de substituabilité du remplaçant.

mentation du niveau de stress dans l'équipe de travail et favorise les comportements de présentéisme.

Ces constats peuvent s'expliquer de trois façons, selon la nature du poste. Premièrement, les salariés pour qui les travaux s'accumulent en leur absence et qui auront à effectuer l'intégralité de leurs tâches à leur retour sont fortement incités à travailler malades (Aronsson et al., 2000), lorsque l'entreprise ne propose pas de solution de remplacement. Deuxièmement, les salariés dont les tâches doivent d'une façon ou d'une autre être prises en charge au quotidien (e.g. vis-à-vis d'une clientèle), et qui savent que leur charge de travail va incomber à leurs collègues, sont susceptibles d'être présentéistes (Dew et al., 2005). Troisièmement, certains salariés, à tort ou à raison, se considèrent comme irremplaçables : c'est par exemple le cas des managers, comme nous l'avons vu (cf. 3.2.2) ou des professions médicales (Demerouti et al., 2009 ; McKevitt et al., 1997).

Restructurations et mobilité. Pour un salarié, le fait d'avoir été amené à changer de poste ou d'emploi dans la même entreprise (par exemple dans le cadre d'une fusion ou d'une restructuration) peut conditionner ses décisions de présence ou d'absence. Firms et al. (2006) analysent l'absence des salariés dans une entreprise en cours de restructuration. Ils indiquent que le niveau d'absence est plus élevé pour les salariés qui anticipent un licenciement. À partir de ce constat, on pourrait faire l'hypothèse que l'impression d'être « abandonné » par une organisation dans laquelle il s'était investi pourrait conduire un salarié à choisir l'absence, en cas de maladie. Ce sentiment d'être en face d'une entreprise déloyale conduirait également à une réduction du présentéisme, entendu comme un moyen de montrer son investissement dans l'entreprise. Cette hypothèse n'est pas vérifiée par les deux articles s'intéressant au présentéisme en période de restructuration. Caverley et al. (2007), à la suite de Simpson (1998), indiquent plutôt que la perspective de mobilités professionnelles et de licenciements conduiraient les salariés à privilégier le présentéisme pour maximiser leur chance de conserver un poste. Simpson (1998) explique que, pour ceux qui aspirent à des postes de management, le présentéisme est une stratégie parmi d'autres pour faire montre de sa fiabilité, malgré la maladie.

4 Conséquences du présentéisme

D'autres informations issues de la littérature sur le présentéisme concernent les conséquences de ce comportement. Dans cette dernière section seront donc explorées les multiples implications du présentéisme. Une frange européenne de la littérature (très majoritairement originaire des pays scandinaves), a cherché à en identifier les conséquences pour les salariés. Plusieurs travaux montrent que le fait de venir travailler étant malade est associé, quelques mois ou années plus tard, à une santé dégradée et donc à des absences-maladies plus nombreuses et plus longues, mais également à un risque de contagion au sein du groupe de travail. Au contraire, la littérature nord-américaine s'est particulièrement attachée à évaluer les conséquences d'un état de santé imparfait sur la productivité au travail.

4.1 Conséquences sur la santé des salariés

En cas de maladie, la renonciation à un arrêt de travail peut avoir des effets sanitaires négatifs. La conséquence à laquelle on songe spontanément est le risque de transmission de la maladie entre le présentéiste et ses collègues. Le champ de l'épidémiologie confirme en effet que le lieu de travail est l'un des contextes de propagation des maladies. Mais le présentéisme a également des répercussions, à plus ou moins long terme, sur la santé des présentéistes eux-mêmes.

4.1.1 *Phénomènes de contagion*

Le risque de transmission de la maladie entre le salarié malade et ses collègues est sans doute la conséquence du présentéisme à la fois la plus citée et la moins étudiée par la recherche. Lorsqu'ils sont interrogés, les salariés associent pourtant clairement le présentéisme et la contagion (Baker-McClearn et al., 2010). Krohne et Magnussen (2011) dans le contexte d'une plateforme pétrolière norvégienne, évoquent même des situations où les collègues font pression sur les salariés présentéistes afin qu'ils se soignent. Le présentéisme est donc parfois perçu négativement à cause de la transmission de microbes (Grinyer & Singleton, 2000). Mais, il faut puiser dans la littérature épidémiologique pour trouver des éléments supplémentaires sur ce phénomène de contagion.

Peu d'articles s'intéressent spécifiquement aux phénomènes de contagion sur le lieu de travail. En étudiant les contacts journaliers directs d'environ 7 000 individus¹⁵, Mossong et al. (2008) montrent qu'une grande partie des contacts directs ont lieu au domicile des individus (23% des contacts) ou dans leur environnement de travail (21%). Ces lieux sont donc plus propices à la transmission d'une maladie. Kumar et al. (2013) constatent que pour éviter la transmission du virus de la grippe, l'agence gouvernementale américaine de santé publique¹⁶ préconise aux salariés de rester chez eux pendant leur maladie et jusqu'à 24 heures après la disparition de la fièvre. Ils souhaitent vérifier l'impact d'une telle mesure sur la diffusion de la maladie et cherchent notamment à savoir si le présentéisme y contribue. Leur article, l'un des premiers sur la question¹⁷, s'appuie sur un outil de simulation communément utilisé pour étudier la propagation d'un virus dans une population de grande échelle. Dans ce cadre, les caractéristiques démographiques d'une population réelle de plus d'un million d'individus¹⁸ sont exploités et la propagation du virus de la grippe est simulée sur la base de données épidémiologiques classiques. Ils comparent notamment la progression de la transmission dans une situation normale (aucun changement dans les habitudes) et dans un scénario où les individus malades prennent systématiquement des jours d'absence-maladie. Le nombre d'infections sur le lieu de travail diminue de 25,3% lorsque les salariés prennent un seul jour d'arrêt, et de 39,2% lorsqu'ils en prennent deux. Dans tous les scénarios étudiés, une baisse du présentéisme conduit à une réduction significative du taux d'attaque (rapport entre le nombre de personnes touchées par la maladie et la population totale).

4.1.2 *Dégradation de la santé*

Bien que le présentéisme favorise manifestement la transmission de la maladie, c'est la santé du présentéiste lui-même qui est en première ligne. La plupart des articles parcourus penchent en faveur d'un effet négatif du présentéisme sur la santé. Grinyer et Singleton (2000) pointent le fait que le présentéisme génère une dégradation de la santé et suggèrent que le report répété d'une absence dont l'objet serait de soigner une maladie bénigne, risque de générer une maladie plus grave, qu'elle soit physique

15. Il s'agit d'individus issus de plusieurs pays européens à qui les auteurs ont demandé de reporter dans un carnet de bord, l'ensemble de leur contacts directs (conversation, contact physique, etc.) avec d'autres individus au cours d'une journée donnée (24 heures).

16. *Centers for Disease Control and Prevention*.

17. Publié dans l'*American Journal of Public Health*, revue de référence en épidémiologie.

18. Vivant aux États-Unis, autour de la ville de Pittsburgh.

ou psychologique. Ceci est aussi mis en lumière par les travaux de Kivimäki et al. (2005). Ils s'intéressent à une pathologie bien précise, l'infarctus du myocarde, dans une population de plus de 5 000 fonctionnaires britanniques (hommes en bonne santé au départ). Les fonctionnaires présentéistes sont identifiés en croisant deux critères : il s'agit d'hommes n'ayant pas connu d'absence sur trois ans (absences relevées par l'employeur) et qui, d'autre part, se disent en mauvaise santé (une échelle de santé perçue est utilisée). Kivimäki et al. (2005) mettent en lumière un risque d'accident coronarien deux fois plus élevé pour les présentéistes, comparativement aux fonctionnaires qui se sont absentés. L'un des vecteurs identifiés est la sensibilité au stress, qui est un facteur de risque de l'infarctus du myocarde : les individus accumuleraient un poids psychologique qui finirait par avoir des résurgences physiologiques négatives. (Johns, 2011 ; MacGregor et al., 2008).

Plus récemment, Bergström et al. (2009a) ont confirmé les conclusions de Kivimäki et al. (2005) en analysant les effets du présentéisme sur la santé générale des individus. Leur étude couvre 9 583 salariés suédois dont 72% travaillent dans le secteur public (et dans le domaine de la santé pour la plupart). Les individus sont interrogés à la fois sur leur propension au présentéisme au cours des 12 derniers mois et sur leur état de santé perçu. Bergström et al. (2009a) observent que le fait d'avoir été présentéiste au point de départ (au moins deux fois au cours de l'année passée) est un facteur de détérioration de la santé 18 mois et 36 mois plus tard. Ils vérifient que ce constat est valable quel que soit l'état de santé des individus au point de départ.

D'après Demerouti et al. (2009), le présentéisme implique clairement un risque de burnout. Dans un cadre hospitalier, ils aboutissent à la conclusion que les infirmières les plus présentéistes ont des risques plus élevés de burnout dans les six mois, et cela dans ses deux aspects, à savoir l'épuisement émotionnel et le sentiment de dépersonnalisation (i.e. perte du sens de la réalité), avec le même résultat. Taloyan et al. (2012), sur un échantillon représentatif de la population active suédoise (7 445 salariés) confirment cette conclusion, tout en soulignant également une dégradation de la santé générale sur les deux années qui suivent des épisodes fréquents de présentéisme. L'épuisement psychologique et physique résultant du présentéisme figurent, d'après eux, parmi les éléments les plus nuisibles à la santé. Parallèlement, Virtanen et al. (2012) montrent que les salariés (parmi 2 123 fonctionnaires britanniques) ayant un temps de travail quotidien élevé (11 à 12 heures) ont un risque de dépression plus élevé six années plus tard.

Une situation beaucoup plus spécifique est évoquée dans l'étude qualitative de Gatrell (2011). Elle décrit le comportement de présentéisme vécu par des femmes enceintes souhaitant prouver à leur entourage de travail qu'elles sont toujours capables de tenir leur poste. Cela a donné lieu à des implications de santé sérieuses : pour certaines d'entre elles, le présentéisme aurait même contribué à un accouchement prématuré.

De rares études supposent, certes, que le présentéisme pourrait avoir des implications positives sur la santé de celui qui le pratique, mais aucun lien empirique n'a été démontré. Ainsi, Aronsson et Gustafsson (2005) se demandent si, dans certains cas, le présentéisme ne pourrait conduire à une amélioration de l'état général, grâce au sentiment de satisfaction et au lien social que permettent le travail. C'est aussi ce que déclarent certains employés interviewés par Krohne et Magnussen (2011) : venir travailler étant malade permet d'éviter l'inactivité et l'isolement social suscités par certaines maladies¹⁹, ce qui aurait, d'après eux, des effets bénéfiques sur la santé. Mais ces résurgences positives du présentéisme semblent ne pouvoir se développer que dans des environnements de travail très propices (Aronsson & Gustafsson, 2005) et dans le cas de pathologies bien précises.

4.2 Conséquences sur le niveau d'absence ultérieur

Par le biais d'une dégradation de la santé des salariés, le présentéisme pourrait lui-même être générateur d'absences-maladie. Au travers d'entretiens qualitatifs, Grinyer et Singleton (2000), mais également McKevitt et al. (1997), laissent entendre que certains salariés négligeant de prendre une absence-maladie brève pour soigner une maladie bénigne risquaient de connaître une période d'absence plus longue à cause de la dégradation de leur état de santé. Si le présentéisme génère des absences, il s'agirait donc plutôt d'absences de longue durée.

C'est également l'hypothèse d'Hansen et Andersen (2009). Dans leur étude, ils ont été en mesure d'associer deux éléments : les réponses d'un échantillon aléatoire de la population active danoise (11 838 individus) à un questionnaire comportant un item sur le présentéisme et les données d'absences de ces mêmes individus, issues du ministère du travail danois. Ils distinguent les absences de durée moyenne (2 à 8 semaines) et celles de durée plus longue (plus de 8 semaines). Pour ces deux catégories d'absence,

19. Sans doute les auteurs évoquent-t-ils ici des maladies longues qui laisseraient malgré tout aux individus une capacité suffisante pour travailler.

ils montrent que les individus présentéistes ont un risque plus élevé de connaître une absence-maladie dans les 18 mois. Le présentéisme entraînerait donc aussi bien des absences de durée moyenne que de longue durée. Pour Hansen et Andersen (2009) ce constat est d'autant plus vrai chez les présentéistes « chevronnés », c'est-à-dire ceux qui déclarent avoir été présentéistes plus de 6 fois l'année précédente. Leur risque d'être absent par la suite est de 53% supérieur à celui des salariés n'ayant pas été présentéistes.

Dans un travail d'une envergure comparable, Bergström et al. (2009b) obtiennent le même résultat. Ils avaient déjà établi un lien entre présentéisme et santé perçue sur un panel de salariés suédois (cf. ci-dessus). Pour la même population, ils exploitent les relevés d'absence des employeurs afin d'analyser l'effet du présentéisme sur les absences ultérieures. Ils montrent notamment que les individus qui, en début d'enquête, déclaraient venir travailler fréquemment étant malades ont un niveau d'absence significativement plus élevé au cours de la troisième et quatrième année suivant l'enquête. Comme Hansen et Andersen (2009), ils trouvent qu'une forte fréquence de présentéisme (plus de 5 épisodes par an) entraîne un niveau d'absence plus élevé ultérieurement.

Enfin, l'article plus récent de Taloyan et al. (2012) confirme les conclusions précédentes. La seule différence est que les données d'absence ne proviennent pas des employeurs, mais sont déclaratives. Les interviewés qui, en 2008, déclaraient plus de 7 jours de présentéisme avaient un niveau d'absence significativement plus élevé en 2010.

4.3 Conséquences sur la productivité au travail

D'autres travaux soulignent que le présentéisme implique une baisse directe de la productivité. Ce constat invite à dépasser une approche binaire considérant d'un côté l'absence comme une situation de productivité nulle et de l'autre la présence comme une situation de productivité totale. En effet, un salarié malade n'est pas en mesure de produire l'effort habituel nécessaire à l'accomplissement de ses tâches, comme nous allons le voir.

Signalons d'emblée que la plupart des articles évoqués dans cette section proposent un chiffrage des « coûts de présentéisme ». Notre propos se limite volontairement aux conséquences du phénomène sur la productivité et ne tient pas compte de ces analyses en termes monétaires. À l'instar de plusieurs auteurs (Gosselin & Lauzier, 2011 ; Vingård et al., 2004), nous estimons que les évaluations du coût économique du pré-

sentéisme sont à prendre avec une certaine réserve. On passe trop rapidement d'une estimation par les salariés de leur propre baisse de productivité (en raison d'une maladie) à une évaluation chiffrée des coûts du présentéisme (Johns, 2010). Ce chiffrage se fait le plus souvent en considérant que la perte liée à une heure de travail perdue équivaut au salaire horaire (Lofland et al., 2004). Les auteurs se contentent alors simplement d'utiliser le salaire horaire moyen de l'ensemble de la population active (Goetzel et al., 2004) ou d'une catégorie professionnelle (Letvak et al., 2012) pour imputer un coût direct au présentéisme.

L'étude de Goetzel et al. (2004) est très souvent citée dans la littérature portant sur les conséquences économiques du présentéisme. Leur travail fait la synthèse de plusieurs enquêtes nationales américaines et évalue le coût total de dix maladies chroniques spécifiques. À côté des coûts directs de la maladie, ils évoquent l'impact du présentéisme. En effet, les auteurs considèrent le coût total d'une maladie donnée comme la somme de trois éléments. Premièrement, ils intègrent les coûts médicaux directs, c'est-à-dire l'ensemble des dépenses relatives aux soins hospitaliers, ambulatoires et pharmaceutiques, dont l'estimation provient de données gouvernementales. Deuxièmement, ils évaluent les pertes économiques liées aux jours d'absence des salariés. Troisièmement, ils calculent les pertes liés aux heures de travail perdues chaque jour à cause de la maladie en question, qu'ils désignent comme le coût du présentéisme. Ces deux derniers éléments (pertes liés à l'absence et au présentéisme) s'appuient sur les déclarations des salariés : ceux-ci sont interrogés sur le nombre de leurs journées d'absence et sur le nombre d'heures qu'ils estiment perdre quotidiennement lorsqu'il viennent travailler étant malade.

Sur cette base, les auteurs calculent un coût total de la maladie²⁰. Mais au-delà de ce chiffrage, un autre constat retient notre attention. Les salariés interrogés estiment que, lorsqu'ils travaillent étant malades, leur productivité est limitée. Parmi les affections ayant le plus d'impact, on trouve ainsi la migraine, les maladies respiratoires ou la dépression : les salariés évaluent respectivement à 20,5%, 17,2% et 15,3% la baisse de productivité moyenne liée à ces trois maladies. Autrement dit, le fait de travailler alors que l'on souffre de l'un de ces problèmes de santé a des conséquences importantes sur l'activité professionnelle.

20. Par exemple, le coût total de la migraine s'élèverait à environ 400 000 USD pour 100 salariés touchés et 89% de cette somme serait à imputer au seul présentéisme.

D'autres auteurs utilisent la même approche pour la mesure de la perte de production liée au présentéisme, et se focalisent souvent sur une affection spécifique, ou un secteur d'emploi spécifique (Schultz et al., 2009). Dans le secteur hospitalier, Letvak et al. (2012), procèdent de la même façon que Goetzel et al. (2004) pour estimer la perte de productivité de 1171 infirmières de Caroline du Nord (USA). Encore une fois, l'analyse repose sur les déclarations des salariés concernant leur propre productivité lorsqu'ils sont présentéistes. La question est énoncée de la façon suivante : « À quel point vos problèmes de santé physique ou psychologique ont-ils affecté votre productivité au travail ? » (sur les 14 derniers jours). La réponse est donnée sur une échelle de Likert à 10 niveaux, que les auteurs transforment en pourcentage d'improductivité (de 0 à 100%). Letvak et al. (2012) trouvent une perte de productivité moyenne de 16,3% en raison d'un problème de santé²¹.

Zhang et al. (2012), en s'appuyant sur les travaux antérieurs de Pauly et al. (2008), estiment que la méthode consistant à mesurer le nombre d'heures de travail perdues est un moyen trop simpliste pour évaluer les pertes de productivité liées au présentéisme. Ce faisant, on se concentre sur l'input alors qu'il faudrait s'intéresser à l'output (Zhang et al., 2011). Or, d'après une étude réalisée auprès de 800 managers américains, Pauly et al. (2008) montrent que certains types d'emplois sont plus sensibles au présentéisme des salariés. Ces emplois se caractérisent par des coûts d'opportunité²² (le travail perdu ne peut pas être rattrapé le lendemain) ou par un travail collectif où l'ensemble de la production peut être ralenti par un membre de l'équipe dont la productivité n'est pas optimale (Zhang et al., 2012). Il existerait donc une perte plus large que la production marginale de l'absent : les auteurs expriment cela sous la forme d'un « coefficient multiplicateur des pertes » liées au présentéisme.

S'il paraît délicat de chiffrer l'impact du présentéisme, il n'en demeure pas moins que les études réalisées montrent un lien net entre le fait de travailler avec une santé dégradée et une baisse de la productivité au travail (Schultz et al., 2009). Est-ce à dire pour autant, que le présentéisme a plus d'impact sur la productivité que l'absence ? Gosselin et Lauzier (2011, p. 19) proposent un exemple simple : « quatre jours de présentéisme amenant une baisse de productivité de 25% du travailleur équivaut à une seule journée d'improductivité ». Dans ce cas, on a donc le sentiment qu'un salarié présent étant

21. Puis, en multipliant ce chiffre par le salaire annuel moyen des infirmières de Caroline du Nord (87 880 USD), ils estiment le coût annuel du présentéisme à 14 439 USD par salarié.

22. Une sensibilité temporelle de la production disent Pauly et al. (2008).

malade sera toujours plus productif qu'un salarié absent, sauf si un jour d'absence permet d'éviter plus de quatre jours de maladie. En l'état actuel de la littérature, ce débat semble difficile à trancher.

4.4 Autres conséquences organisationnelles

Des questions se posent donc clairement sur les moyens de mesurer la baisse de rendement induite par le présentéisme, mais la littérature tient largement pour acquis le fait qu'un salarié malade ne peut accomplir parfaitement et intégralement ses activités professionnelles (Johansen, 2012 ; Johns, 2010). Ce constat procède autant de l'intuition que de la déclaration des salariés eux-mêmes.

Cependant, d'autres dimensions de la vie organisationnelle peuvent supporter les conséquences du présentéisme. Il peut par exemple avoir des répercussions sur le climat social de l'équipe de travail, voire de l'entreprise. Johns (2010) a largement développé la notion de « culture de présentéisme » dans la lignée de la culture de l'absence (Nicholson & Johns, 1985). Le statut du présentéisme peut être très différent suivant le contexte. Comme nous l'avons vu, dans certaines équipes, il peut y avoir une certaine pression des collègues tendant à dissuader un salarié malade de venir travailler afin d'éviter la contagion (Grinyer & Singleton, 2000 ; Krohne & Magnussen, 2011), tandis que d'autres équipes vont développer un regard positif sur le présentéisme qui évite les reports de charge de travail (Caverley et al., 2007 ; Dew et al., 2005). Si des salariés viennent à déroger à ces normes locales, des frictions peuvent aisément naître. Grinyer et Singleton (2000) mettent également en avant que le présentéisme, lorsqu'il est la conséquence de mesures destinées à lutter contre l'absence, peut susciter une dégradation plus large du climat social : les salariés qui subissent des injonctions à la présence peuvent développer un vrai ressentiment à l'égard de l'employeur ou de politiques de sanction de l'absence perçues comme injustes.

Par ailleurs, certains articles portant sur le milieu médical ou hospitalier évoquent les conséquences assez alarmantes du présentéisme sur la qualité des soins. L'article de McKevitt et al. (1997), l'un des premiers ayant spécifiquement traité du présentéisme, s'intéressait déjà au cas des médecins, dont la réputation de présentéistes a été confirmée par la suite (Aronsson & Gustafsson, 2005). Les médecins interviewés ont tout à fait conscience que le fait de travailler étant malade nuit à la qualité de la prise en charge des patients. Bien plus encore, Letvak et al. (2012) montrent que les infirmières

(américaines) déclarant être souvent présentéistes sont aussi celles qui déclarent le plus grand nombre d'erreurs médicales. Les résultats de Demerouti et al. (2009), sur un autre panel d'infirmières va clairement dans le sens d'une dégradation de la qualité des soins. L'un des vecteurs identifiés est celui du développement d'attitudes négatives vis-à-vis des patients. Les infirmières présentéistes ont en effet tendance à ne réaliser que les tâches médicales incontournables et à établir une certaine distance relationnelle avec le patient. En dernier lieu, les conclusions de Letvak et al. (2012) soulignent même que le présentéisme est associé à une augmentation du nombre de chutes de patients au cours des soins infirmiers. De leur point de vue, une infirmière absente, pouvant éventuellement être remplacée, vaut toujours mieux qu'une infirmière présentéiste, étant données les conséquences catastrophiques sur la qualité des soins.

Conclusion

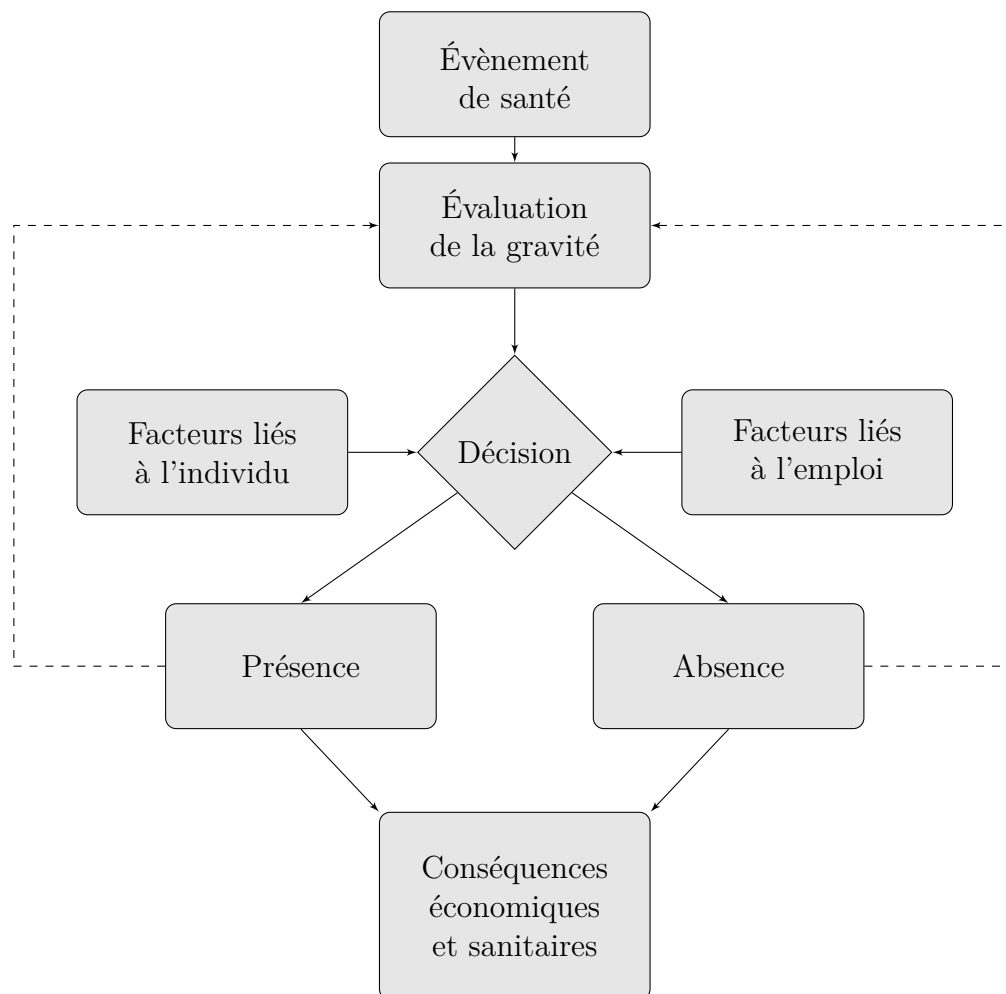
Conjointement aux preuves empiriques apportées dans la littérature ont émergé des tentatives de modélisation du présentéisme (Aronsson & Gustafsson, 2005 ; Gosselin & Lauzier, 2011 ; Gosselin et al., 2013 ; Johansson & Lundberg, 2004 ; Johns, 2010). L'objectif essentiel de la construction d'un modèle est de proposer un regard général sur l'ensemble du processus décisionnel qui conduit au présentéisme, en articulant sa conceptualisation, ses causes et ses conséquences. Ces modèles ont une forme relativement similaire, les auteurs insistant parfois plus ou moins sur certains éléments. En guise de bilan de ce chapitre, nous nous proposons d'en réaliser une synthèse sous la forme d'une représentation graphique (figure 1.3 page 56).

La situation qui donne lieu au présentéisme peut alors être énoncée ainsi : le cours habituel de la vie professionnelle, matérialisé par une présence au travail et un état de santé permettant une productivité optimale, est perturbé par la survenue d'un événement de santé (Johansson & Lundberg, 2004). Les individus prennent la décision de s'absenter ou d'être présentéistes, après avoir évalué la nature et la gravité de cet événement (Gosselin & Lauzier, 2011). Mais comme nous l'avons développé dans ce chapitre, les individus sont aussi influencés dans leur choix par des facteurs individuels (leur statut démographique, leur vie familiale, leur personnalité) et organisationnels (la nature et la quantité de leurs tâches, leur statut dans l'entreprise, les politiques liées à l'absence). La décision d'être présentéiste est en cela le résultat d'une équation à

plusieurs inconnues. Ce choix final a lui-même des conséquences. Celles-ci sont principalement sanitaires (contagion, dégradation de la santé) et économiques (productivité nulle pour l'absence, partielle pour le présentéisme). Dans le graphique 1.3, l'effet sur la santé de l'individu est matérialisé par une flèche pointillée évoquant une boucle de rétroaction, puisque le présentéisme peut produire une dégradation de l'état de santé initial. Une nouvelle décision est alors à prendre.

L'un des enjeux essentiels pour compléter ce modèle explicatif est de parvenir à une mesure adéquate du présentéisme. Dans ce chapitre, nous avons montré que les méthodes de mesure employées présentent certaines limites, notamment en ce qui concerne leur opérationnalisation en entreprise. Dans les développements à venir, nous proposons une autre perspective sur cette problématique de la mesure, qui sans être exempte de limites, permet une évaluation plus aisée de la propension des salariés à s'absenter étant malades.

Graphique 1.3 – Synthèse des modèles théoriques



Chapitre 2

Regard sur les données d'absence

Introduction

Quoique le présentéisme soit devenu ces dernières années un domaine de recherche à part entière (Taloyan et al., 2012), il demeure encore relativement méconnu en dehors de la sphère académique. À ce titre, l'étonnement suscité par la description du phénomène dans les services ressources humaines des entreprises avec lesquelles nous avons travaillé, est suffisamment explicite. Les dirigeants d'entreprise sont sensibles au « taux d'absentéisme » des salariés, mais ne mesurent pas toujours les implications des dispositifs qu'ils mettent en place pour le réduire. Le présentéisme, sans être exclusivement un effet corollaire des mesures de réduction de l'absence, fait partie des conséquences négligées de ces dispositifs.

Dans le chapitre précédent, nous avons émis l'hypothèse que la raison essentielle de la méconnaissance du présentéisme en entreprise est la difficulté de sa mesure. En effet, aucune investigation sur le présentéisme ne semble pouvoir se départir d'une interrogation directe des salariés, ce qui implique des dispositifs de nature à décourager les entreprises

souhaitant procéder à un premier diagnostic. De plus, les données déclaratives obtenues par l'administration de questionnaires ou d'entretiens qualitatifs présentent certaines limites liées notamment à la subjectivité des répondants (cf. chapitre précédent). On comprend donc pourquoi le monde de l'entreprise est resté relativement imperméable à la question du présentéisme.

À titre de comparaison, la littérature autour des comportements d'absence des salariés regorge de travaux s'appuyant sur des micro-données d'entreprise (Johns, 2003). La multiplication des études de cas d'organisations spécifiques a permis de faire émerger les caractéristiques saillantes des comportements d'absence. Ces micro-données proviennent principalement des enregistrements par l'employeur du nombre de jours d'absence pour chaque salarié. Les employeurs ont compris très tôt qu'il était dans leur intérêt d'avoir ce réflexe du « pointage » des absences. Une recherche documentaire, nous a par exemple permis de trouver, en annexe d'une note interne des mines d'Anzin, un tableau statistique datant de 1881, pointant les absences des mineurs et s'interrogeant sur l'augmentation du phénomène¹. Si, à cette époque, les chiffres des absences provenaient des relevés quotidiens des chefs d'équipes, les méthodes d'enregistrement et d'analyse se sont progressivement développées pour aboutir à des systèmes d'information des ressources humaines individualisant et automatisant l'enregistrement des absences. Aujourd'hui, pour la plupart des entreprises, ces données d'absences sont obtenues de façon systématique, sont faciles à rassembler, sont de bonne qualité et surtout ont un coût très faible. C'est donc assez naturellement qu'est apparue dans notre esprit la possibilité d'utiliser ces micro-données d'absence pour fournir un premier aperçu du présentéisme dans les limites d'une entreprise.

Plusieurs types d'absences sont observées dans la vie organisationnelle : congés de maternité, congés sans solde, absences pour événements familiaux, enfants malades, etc. Dans cette thèse nous exploiterons uniquement les *absences-maladie*, c'est-à-dire l'absence en raison d'un problème de santé. D'après nous cette variable d'absence-maladie est composite et recouvre en fait bien plus que l'absence des salariés. Sa richesse émerge lorsqu'on considère l'absence et la présence d'un salarié à l'aune de son état de santé. Comme nous l'avons indiqué en introduction, et dans l'esprit de la typologie proposée

1. Centre Historique Minier de Lewarde, *Notes à la Régie [des Mines d'Anzin] sur les opérations du mois de juin et celles du premier semestre 1881*, dossier n° 6283. On peut également y lire le passage suivant : « nous réagissons de notre mieux contre cette tendance de nos ouvriers à travailler moins qu'auparavant ; les absences du lundi sont sévèrement punies d'une amende de 3 francs ».

par Chatterji et Tilley (2002), nous distinguons quatre situations : la situation normale de présence au travail (se rendre au travail en bonne santé), l'absence (rester chez soi étant malade), le présentéisme (se rendre à son travail étant malade) et l'absentéisme (rester chez soi étant en bonne santé).

Cette typologie est loin d'être purement rhétorique : son implication est réelle pour une thèse dont l'objet est de mesurer ces phénomènes. Qualifier la nature et les relations respectives de ces quatre éléments et comprendre qu'ils participent d'un même processus est un préalable nécessaire à toute tentative de mesure. Ces quatre comportements se trouvent en effet intriqués au sein des données d'absence-maladie où chacun d'entre eux génère des valeurs typiques. La présence normale et le présentéisme produisent des valeurs nulles². L'absence-maladie et l'absentéisme des valeurs positives. Il peut *a priori* sembler impossible de différencier l'absence et l'absentéisme, la présence et le présentéisme. Comment savoir si une valeur positive d'absence est due à une maladie ou à une absence injustifiée ? Comment savoir si un salarié n'ayant jamais été absent (valeur nulle) a été présentéiste ou n'a connu aucune maladie durant l'année ?

Le véritable défi est ici de trouver une méthode statistique capable de dissocier ces phénomènes. La section 3 de ce chapitre retracera le cheminement intellectuel qui nous a conduit à examiner successivement plusieurs outils statistiques, à comprendre leurs caractéristiques et leurs implications et enfin à choisir l'un d'entre eux, qui sera détaillé dans le troisième chapitre de cette thèse. Mais avant cela, nous consacrons une large partie du chapitre à la découverte de cette variable d'absence. D'après nous les entreprises disposent là d'une source d'information dont la richesse n'a pas été entièrement mesurée, et qui, par ailleurs, permet de dépasser les limites traditionnelles des données déclaratives (Ferrie et al., 2005 ; Kivimäki et al., 2003)³. Qu'elles soient obtenues par déclaration des salariés ou par enregistrement dans un système d'information en entreprise, les données d'absence ont des caractéristiques communes qu'il est nécessaire de mettre en évidence afin de modéliser pertinemment ce type de variables. Sur la base de plusieurs études existantes autour des données d'absence et en présentant quelques statistiques, il nous paraît d'abord opportun de montrer comment s'est affiné le choix de la variable dépendante qui constitue le pilier de cette thèse et de « faire connais-

2. Comme on le verra, on peut cependant admettre que le présentéisme peut aussi générer des valeurs faiblement positives.

3. D'après Ferrie et al. (2005, p. 79), « recorded absence from employers' registers should remain the "gold" standard for studies ».

sance » avec elle (section 1). Ce sera l'occasion de décrire sa distribution statistique particulière et ses caractéristiques essentielles. D'autre part, comme un préliminaire à la présentation de notre méthodologie, il nous semble utile de consacrer quelques pages aux liens existant entre l'absence-maladie et la santé des individus (section 2). La santé constitue elle-même un relais entre le niveau d'absence-maladie et l'âge des salariés. Cette dernière variable tient une place importante dans la suite de notre analyse, c'est pourquoi nous nous y arrêterons également dans la deuxième section.

1 Choix de la variable dépendante

En premier lieu, plusieurs aspects des données d'absence sont présentés. La méthodologie de mesure du présentisme proposée dans cette thèse repose largement sur la variable d'absence utilisée, qui deviendra ultérieurement la variable dépendante de notre outil économétrique. Les choix effectués quant à la nature et à la portée de cette variable sont donc centraux et méritent d'être présentés en détail.

1.1 Gravité et fréquence d'absence

Comme nous le verrons, les données d'absence sont des données de comptage. Mais que s'agit-il de dénombrer ? L'analyste a le choix entre le nombre de jours d'absence et le nombre d'épisodes d'absence. Dans la littérature, ces deux approches prédominent lorsqu'il s'agit de mesurer quantitativement l'absence des salariés (Johns, 2003 ; Kristensen et al., 2006 ; Steel, 2003). D'une part, on peut comptabiliser les épisodes d'absence, autrement dit le nombre d'occurrences sur une période donnée : on parle de « fréquence » d'absence. Tous ces épisodes d'absence possèdent un poids équivalent, qu'ils occasionnent un jour d'arrêt ou plusieurs semaines d'indisponibilité. Les valeurs les plus élevées sont attribuées aux individus s'étant absentés le plus fréquemment. D'autre part, on peut dénombrer le nombre de jours d'absence, c'est-à-dire exprimer chaque absence en fonction de sa durée : on parle alors de « gravité » d'absence. Dès lors, le fait de s'absenter pour plusieurs jours n'a plus le même poids dans l'analyse que le fait d'être absent un seul jour : les valeurs les plus élevées sont attribuées aux individus s'étant absenté le plus longuement.

Ces deux approches ont souvent été mises en relation avec deux significations de l'absence. Très tôt, il a été mis en évidence que la mesure de fréquence d'absence est

un indicateur plus pertinent des absences dites « volontaires » (autrement dit l'absentéisme), tandis que la mesure de gravité est plus adéquate pour décrire des absences dites « involontaires » (Chadwick-Jones et al., 1971 ; Shapira-Lishchinsky & Rosenblatt, 2009 ; Steel, 2003). En effet, un comportement d'absence répétitif, donnant lieu à de nombreuses périodes de retrait de très courte durée (un ou deux jours), est souvent envisagé à la lumière du contexte de travail ou de l'attitude du salarié, plutôt qu'en relation directe avec sa santé. Ces absences épisodiques proviennent davantage d'une perception négative du travail (Shapira-Lishchinsky & Rosenblatt, 2009) : leurs causes s'expriment en termes de motivation, de satisfaction liée au travail (Kristensen et al., 2006). Dans le même esprit, la mesure de l'absence en termes de gravité (durée) est un meilleur indicateur des comportements d'absence liés à la santé. Évidemment, les auteurs s'empressent de reconnaître que cette typologie de la mesure n'est pas parfaitement cloisonnée : une même méthode de mesure peut recouvrir des absences volontaires et des absences involontaires (Hammer & Landau, 1981). Des données d'absences idéales, « purifiées » de tout ce que le chercheur ne souhaite pas prendre en compte, n'existent pas (Steel, 2003). Mais ce débat a le mérite d'ouvrir la réflexion sur ce que contiennent réellement les données d'absence : bien plus qu'il n'y paraît *a priori*.

Dans cet esprit, nous optons pour une mesure du nombre de jours d'absence, c'est-à-dire une mesure de gravité. Dans cette thèse, nous nous intéressons aux absences liées à un événement de santé, à une maladie. Or, la durée de l'absence donne une indication supplémentaire sur la nature de la maladie et permet de tenir compte de la diversité des incidents de santé. Telle maladie laissera le choix à l'individu de venir travailler ou pas, tandis que telle autre l'obligera à entamer une démarche de soin, voire à être hospitalisé. L'information de la durée de l'absence, appréhendée dans la littérature comme une « pondération » des absences, revêt donc dans notre cas une importance particulière. Par ailleurs, nous verrons que les données en fréquence se caractérisent également par une très faible variance : dans nos bases de données 96,8% des individus ont une valeur inférieure ou égale à 3. Des régressions dont la variable dépendante est la fréquence d'absence-maladie sont toutefois présentées en annexe 3 : elles demeurent congruentes avec les régressions basées sur la gravité d'absence, mais impliquent sans doute d'autres choix de modélisation.

1.2 Période d'analyse

Les variables d'absence utilisées dans les travaux empiriques peuvent couvrir des périodes de profondeurs diverses, variant de quelques mois à quelques années (Johns, 2003). Les observations effectuées plus haut au sujet du présentéisme sont également vraies pour les données déclaratives d'absence : une période de rappel d'un an implique un effort important pour l'interviewé et nuit à la qualité des données. Cette limite ne s'applique pas aux données provenant de l'employeur. L'année reste clairement la période d'analyse la plus pratiquée (Harrisson & Martocchio, 1998), et c'est aussi celle que nous choisissons dans cette thèse.

Steel et Rentsch (1995) ont montré que le choix d'une période d'analyse trop courte peut conduire à d'importants écarts de mesure : pour une même population, les résultats diffèrent par exemple considérablement si l'on se fonde sur une période d'analyse de deux mois comparativement à une période d'un an ou plus. Nous voyons deux autres limites au choix d'une période d'analyse infra-annuelle. D'une part, comme nous le montrerons plus bas (section 2), les données d'absence sont cycliques : leur niveau varie considérablement selon la période de l'année pendant laquelle se déroule l'étude, notamment en raison des épidémies de maladies saisonnières. D'autre part, de très courtes périodes d'analyse produisent des données comportant presque exclusivement des valeurs nulles. Pour illustrer cette assertion on peut prendre l'exemple de l'enquête Emploi réalisée en France entre 2003 et 2011. La période de référence y est très courte : une seule semaine. Une étude de la DARES (2013) montre qu'en moyenne seuls 3,6% de l'ensemble des salariés ont déclaré avoir été absents au moins une journée pendant la semaine de référence. Autrement dit, il y a en moyenne 96,4% de valeurs nulles sur une telle période d'analyse.

Certains travaux affirment que, lorsque l'absence est mesurée en termes de durée (gravité), il y aurait un intérêt à choisir des périodes d'analyse plus longues (Steel, 2003). À l'instar de Steel et Rentsch (1995), ils suggèrent qu'une période d'analyse de trois ans permettrait de mettre à jour davantage d'effets (les corrélations étant plus marquées) mais également d'obtenir une distribution statistique moins asymétrique (cf. 1.4) et donc plus facile à exploiter. Pour les raisons évoquées plus haut, nous conserverons toutefois une période d'analyse d'une année. D'autres éléments confortent ce choix. D'une part, notre travail se positionne sur le terrain opérationnel de l'entreprise. Or, les données d'absences n'y sont pas toujours enregistrées de manière extensive :

celles de la dernière année ou de l'année en cours sont plus facilement accessibles et nécessitent moins de vérifications et autres corrections. D'autre part, dans la mesure où notre travail comprend une dimension diachronique, une période d'analyse d'une année facilitera également l'étude de l'évolution inter-annuelle du phénomène. Toutefois, si tel est le choix posé dans cette thèse, le cumul sur de plus longues périodes peut s'avérer fructueux et n'est pas exclu dans des travaux ultérieurs.

1.3 Absences individuelles et absences agrégées

Autre choix possible dans la mesure de l'absence, le niveau d'agrégation des données. La littérature s'appuie traditionnellement sur des données d'absences individuelles (Rentsch & Steel, 2003). Mais, récemment, quelques travaux ont choisi d'aller au-delà des comportements individuels pour s'intéresser à la dimension collective de l'absence (Hausknecht et al., 2008 ; Kristensen et al., 2006 ; Nielsen, 2008). Pour cela, les données d'absences sont agrégées sur des unités de travail. On ne mesure plus alors le niveau d'absence d'un individu, mais celui d'un groupe d'individus : d'une équipe de travail, d'un département de l'entreprise, etc. (Rentsch & Steel, 2003). Selon Hausknecht et al. (2008), une telle agrégation aurait plusieurs effets bénéfiques. D'abord, elle permettrait de mesurer des effets inobservables lorsque l'on se limite à l'angle individuel. Un groupe de travail développe ses propres normes et ses règles informelles : cette culture d'absence (Rentsch & Steel, 2003) et ses implications ne peuvent être prises en compte que par l'agrégation des données. Ensuite, d'un point de vue méthodologique, « agréger les absences individuelles au niveau de l'unité génère des données normalement distribuées, ce qui ouvre au chercheur de nouvelles options analytiques » (Hausknecht et al., 2008, p. 1224). Enfin, ce type de mesure aurait l'avantage de mieux orienter les décisions d'intervention des dirigeants d'entreprise sur les unités où se concentrent les niveaux d'absence les plus élevés.

Mais Rentsch et Steel (2003) reconnaissent que la littérature a toujours prioritairement considéré l'absence comme une décision individuelle. Dans cet esprit, notre thèse optera généralement pour une approche individuelle, considérant que le présentisme et l'absence-maladie sont deux modalités possibles du choix d'un individu confronté à un problème de santé. Nous ne négligeons pas pour autant l'importance du collectif et de sa culture d'absence : dans la partie empirique de ce travail, nous serons par exemple amenés à étudier l'effet de la relation manager-subordonné sur le présentisme. Bien

que les données d'absence ne seront pas agrégées, les enjeux de la dimension interpersonnelle seront évoqués à cette occasion.

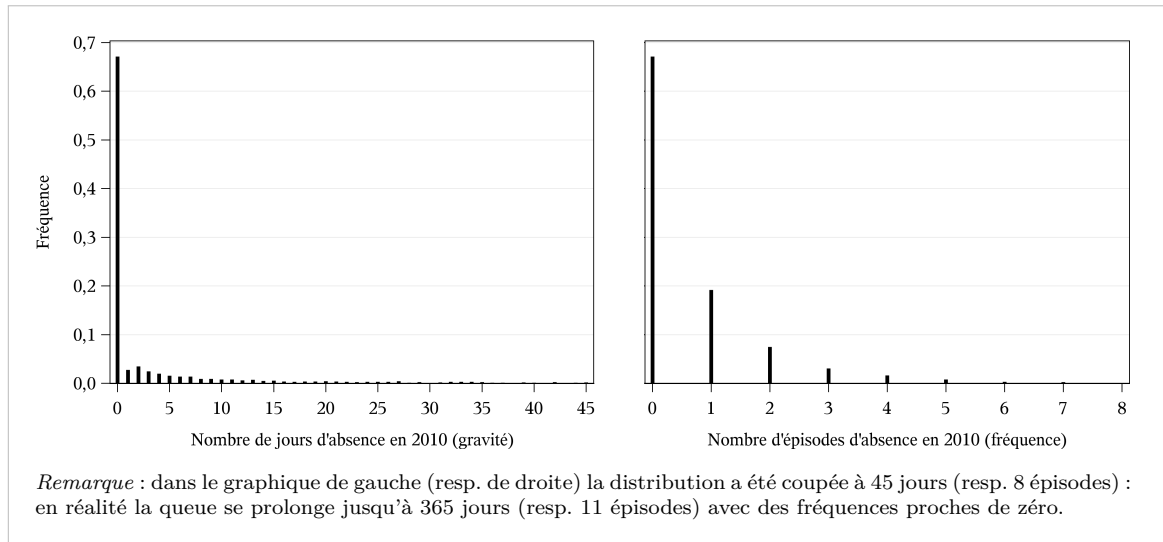
1.4 Distribution statistique

Il est remarquable que les travaux sur les données d'absence ont souvent cherché des moyens pour contourner les difficultés méthodologiques liées aux spécificités de la distribution statistique des absences. L'allongement de la période d'analyse ou l'agrégation des données à un niveau collectif ne sont que deux exemples parmi d'autres. Le fait est que les absences, qu'elles soient mesurées selon leur fréquence ou selon leur gravité, ont une distribution statistique atypique. En anticipant sur des données d'entreprise qui seront utilisées dans les deux derniers chapitres de cette thèse (un centre hospitalier régional, noté CHR ci-après), nous présentons dans cette section quelques éléments caractéristiques. Le graphique 2.1 (page 65) illustre la distribution des absences suivant leur modalité de mesure. De plus, en comparant ces distributions de données enregistrées par l'employeur à celle de données déclaratives (graphique 2.2 page 65), on peut faire le constat d'une proximité. Les données déclaratives présentées sont issues de l'enquête *European Working Conditions Survey* (EWCS)⁴, dont nous nous sommes procuré les données brutes à des fins de comparaison. Pour obtenir ces données d'absence, les interviewés sont invités à indiquer le nombre de jours pendant lesquels ils ont été absents pour des raisons de maladie lors de l'année écoulée.

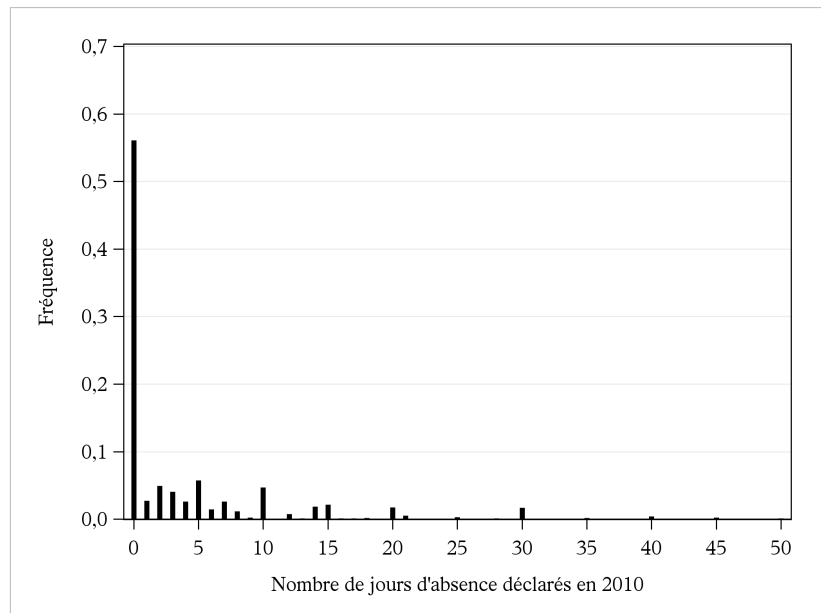
Avant d'entrer dans les détails, on peut noter l'existence de valeurs typiques dans les distributions présentées. Les données déclaratives de l'EWCS sont caractérisées par des pics aux valeurs 5, 10, 15, 20 ou encore 30. Lorsque les individus évaluent la quantité de leurs jours d'absence, ils ont tendance à répondre par des valeurs arrondies. S'il en était besoin, il s'agit là encore d'une indication que les données déclaratives peuvent être biaisées. Cependant, les données d'entreprise ont elles aussi leurs valeurs typiques. Dans la mesure du nombre de jours d'absence au CHR en 2010, on voit par exemple que la valeur 1 est sous-représentée tandis que la valeur 7 est légèrement surreprésentée. Cela provient du fait qu'un certificat d'absence-maladie est assez rarement établi pour une seule journée : lorsqu'un événement de santé nécessite une consultation médicale, au moins deux jours sont généralement prescrits. De même la prescription d'un repos

4. Réalisée en 2010, sur 38 000 personnes issues de 34 pays européens. Nous conservons ici l'ensemble des données, mais les résultats sur ce périmètre sont similaires à ceux obtenus en se concentrant sur la France.

Graphique 2.1 – Nombre de jours et d'épisodes d'absence (CHR en 2010)



Graphique 2.2 – Nombre de jours d'absence déclarés (EWCS en 2010)



d'une semaine (7 jours) est sans doute plus courant qu'un repos de 6 jours. Rappelons que dans le contexte hospitalier, le travail n'est pas organisé sur les jours ouvrés, mais sur les jours calendaires, l'hôpital fonctionnant également le samedi et le dimanche⁵.

On peut déterminer plusieurs autres particularités statistiques des données d'absence, également relevées par certains auteurs (Hammer & Landau, 1981 ; Lu et al., 2010 ; Steel, 2003). Ces constats sont résumés dans le tableau 2.1. La spécificité la plus visible est la surreprésentation des valeurs nulles. Dans les données du CHR, 67,12% des individus n'ont pas été absents en 2010 et n'ont donc aucun jour d'absence (valeur nulle). Ce grand nombre de zéros est l'une des sources de surdispersion des données, ce qui signifie que la variance est beaucoup plus élevée que la moyenne. Ensuite, la forme de la distribution est asymétrique : le coefficient d'asymétrie (noté *skewness* dans le tableau 2.1) est fortement positif lorsque l'absence est mesurée d'après sa gravité. Le coefficient d'aplatissement élevé (noté *kurtosis*) renvoie à la longueur de la queue de la distribution⁶. À côté d'un grand nombre de valeurs nulles, la variable contient en effet des valeurs positives rares, mais élevées : certains individus ont été malades pendant la majeure partie de l'année, voire l'intégralité.

Tableau 2.1 – Description statistique des données d'absence

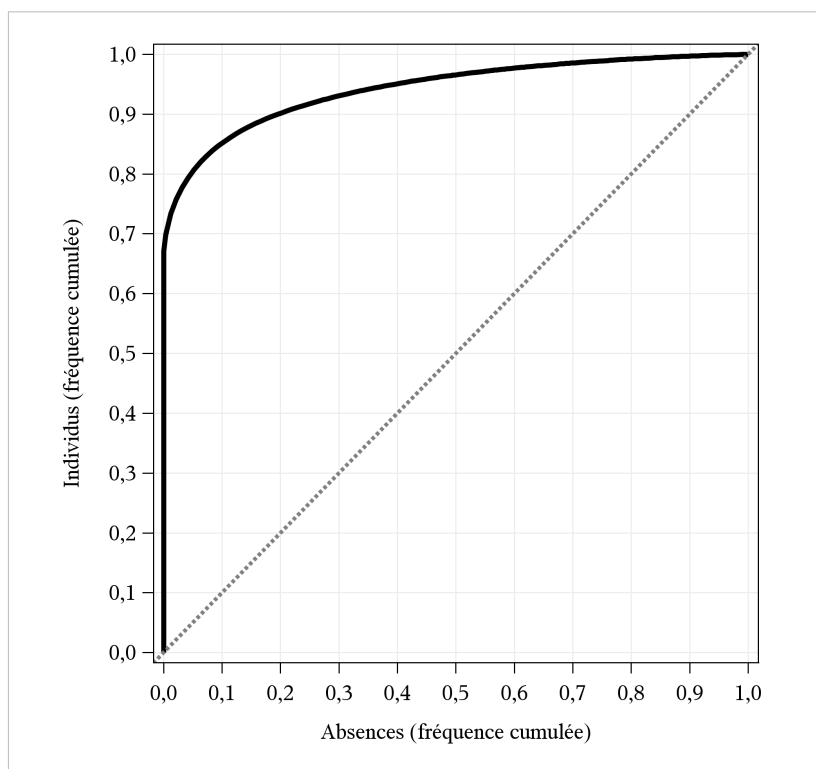
Statistique	CHR (2010)	CHR (2010)	EWCS (2010)
	Gravité	Fréquence	Gravité (déclarée)
Moyenne	7,93	0,59	6,45
Écart-type	25,15	1,12	19,74
Variance	632,66	1,26	389,69
<i>Skewness</i>	6,24	2,91	8,02
<i>Kurtosis</i>	54,07	11,81	90,74
Minimum	0	0	0
Maximum	365	11	365
Part de zéros	67,12%	67,12 %	56,10%

5. Dans les entreprises fonctionnant sur les jours ouvrés, comme dans la banque régionale que nous serons amenés à étudier, c'est la valeur 5 qui est surreprésentée.

6. Dans le cas d'une distribution normale les coefficients d'asymétrie et d'aplatissement prennent une valeur nulle.

Cela conduit également à faire le constat d'une grande concentration des données d'absence, comme le montre la courbe de Lorenz pour le CHR en 2010 (graphique 2.3). Cette année là, 3,41% des individus étaient à l'origine d'environ 50% des jours d'absence enregistrés dans l'établissement⁷. Les 38 individus les plus malades (environ 1% de la population) cumulaient à eux seuls 23,3% de l'ensemble des jours d'absences. Le coefficient de Gini est de 0,88 (la valeur unitaire correspondant à une inégalité totale).

Graphique 2.3 – Concentration du nombre de jours d'absence (CHR en 2010)



Pour contourner ce problème de forte concentration des absences, un choix méthodologique crucial s'est imposé à nous : l'exclusion des absences supérieures à 45 jours. Nous aurons l'occasion de justifier ce seuil lors de l'application empirique, mais ceci revient à exclure environ 5% des individus les plus malades. Ce choix s'avère nécessaire dans la perspective de l'étude du présentisme : il nous faut exclure les absences sur lesquelles l'individu n'a aucun pouvoir de décision, les absences qui s'imposent à lui

7. La médiane s'établit à 80 jours d'absence.

par la gravité de la maladie. Pour mesurer la propension de certains individus à venir travailler étant malades, il n'est pas souhaitable de déséquilibrer l'analyse en intégrant ces longues maladies.

En résumé, nous faisons différents choix méthodologiques sur la variable dépendante que l'on peut synthétiser par quatre propositions. Nous optons pour :

- une mesure de la gravité (nombre total de jours) plutôt que de la fréquence ;
- une période d'analyse annuelle plutôt qu'infra-annuelle ;
- une étude des données individuelles plutôt qu'une agrégation ;
- une variable limitée aux absences de moins de 45 jours.

2 Relation entre santé et absence au travail

Les données d'absence sont très souvent utilisées dans la littérature comme un critère exprimant la motivation ou l'engagement au travail des salariés (Hausknecht et al., 2008). De nombreuses études ont souligné que les individus satisfaits de leur travail et épanouis dans les tâches qu'ils accomplissent ont un niveau d'absence moins élevé (Kristensen et al., 2006 ; Punnett et al., 2007). Dans la pratique de l'entreprise, le niveau d'absence est donc peu à peu devenu un indicateur de satisfaction au travail. Plus largement, le « taux d'absentéisme », à l'instar du taux de turnover, est entré dans la gamme des instruments permettant de mesurer le climat social d'une organisation. À l'extrême, on en vient à considérer qu'un salarié à l'aise dans son travail, dont la rémunération et les conditions de travail sont satisfaisantes (Ose, 2005), ne devrait guère s'absenter.

Ce faisant, on a parfois omis qu'une absence-maladie pouvait également être liée... à la santé de l'individu. Comme l'indiquent fort justement Afsa et Givord (2009, p. 83), dans la littérature « les comportements d'absence sont [traditionnellement] considérés comme un révélateur de l'effort du salarié. Curieusement, l'état de santé intervient rarement dans ces modèles ». Nous l'avons souligné dans le premier chapitre, la question de la santé n'a finalement pris sa place que tardivement dans la littérature managériale sur l'absence. Or, des travaux du champ de l'économie de la santé mettent en évidence les relations existantes entre l'absence et la santé des salariés. À cette relation entre absence et santé, nous ajoutons une troisième variable nous paraissant essentielle, à savoir l'âge des salariés. Cette variable, trop souvent considérée comme une simple variable

de contrôle, devrait toujours être prise en considération dans l'analyse des données d'absence (Donders et al., 2012; Kivimäki et al., 1997). Cette intercorrélation entre l'absence au travail, la santé et l'âge des salariés viendra par la suite appuyer l'un des arguments montrant la pertinence de notre méthodologie de mesure du présentisme.

2.1 Absence et indicateurs de santé

2.1.1 *Résultats de l'enquête Whitehall II*

Avant de donner quelques éléments statistiques sur les relations entre absence, santé et âge, il est intéressant de s'arrêter sur les résultats d'une enquête de grande envergure : l'enquête Whitehall II. Elle s'intéresse à une cohorte de 10 308 fonctionnaires britanniques basés à Londres et âgés de 35 à 55 ans. En première vague (entre 1985 et 1988), ces individus ont subi des examens médicaux divers et ont été interrogés sur leurs conditions de travail, leur santé perçue, leurs incidents de santé, etc. Le suivi de santé s'est ensuite poursuivi sur les années suivantes. En particulier, les absences-maladie enregistrées par l'employeur pour chaque salarié de la cohorte sont disponibles sur 13 années entre 1985 et 1998.

Cette enquête a produit un résultat particulièrement intéressant pour notre propos : de manière générale, le niveau d'absence des individus est corrélé à leur état de santé (Ferrie et al., 2005; Kivimäki et al., 2003; Marmot et al., 1995). Ferrie et al. (2005) soulignent que cette relation est confirmée quelle que soit la méthode de mesure de la santé (état de santé perçu ou nombre d'incidents de santé avérés) et quelle que soit la méthode de recueil des données d'absence (absences déclarées ou enregistrées). Les absences au travail fournissent également une bonne prédiction du niveau de mortalité (Head et al., 2008; Kivimäki et al., 2003). La mortalité augmente lorsque le niveau d'absence-maladie a précédemment augmenté : selon Kivimäki et al. (2003), les salariés présentant en moyenne plus d'un épisode d'absence (de plus de 7 jours) sur deux ans, ont un niveau de mortalité deux à cinq fois plus élevé que les autres, quel que soit leur sexe. Head et al. (2008) ajoutent que le fait d'avoir des informations sur la cause médicale de l'absence améliore cette capacité de prédiction.

Enfin, on peut noter que certaines études exploitant l'enquête Whitehall II évoquent une relation en forme de U entre l'absence et la santé. Ainsi, les personnes prenant quelques jours d'absence seraient globalement en meilleure santé (et leur mortalité se-

rait plus faible) que les individus n'ayant jamais d'absence (Kivimäki et al., 2003). D'autres articles s'appuient également sur ce constat (Hansen & Andersen, 2009 ; McKevitt et al., 1997) pour pointer l'existence du présentéisme et ses répercussions sur la santé. Dans le premier chapitre, nous nous étions d'ailleurs déjà intéressé à une étude, exploitant également les données de cette enquête, qui montrait que les personnes s'absentant peu étant malades, voyaient augmenter leur risque d'accident coronarien les années suivantes (Kivimäki et al., 2005). Mesurer le présentéisme conjointement à l'absence-maladie est alors un moyen de tenir compte de l'aspect multiforme de la santé au travail : présentéisme et absence sont deux réponses possibles à un incident de santé (Caverley et al., 2007) et les salariés peuvent être amenés à substituer l'un à l'autre (MacGregor et al., 2008).

2.1.2 European Working Conditions Survey

Dans la section précédente, nous nous sommes concentrés sur les travaux exploitant l'enquête Whitehall II reconnue pour sa qualité. Mais ces résultats concernant le lien entre état de santé et absence au travail sont largement partagés dans le reste de la littérature, notamment française (Boot et al., 2011 ; Chaupain-Guillot & Guillot, 2007 ; Missègue, 2007). Chaupain-Guillot et Guillot (2011) montrent que l'état de santé fait partie des déterminants significatifs de l'absence dans l'ensemble des pays européens.

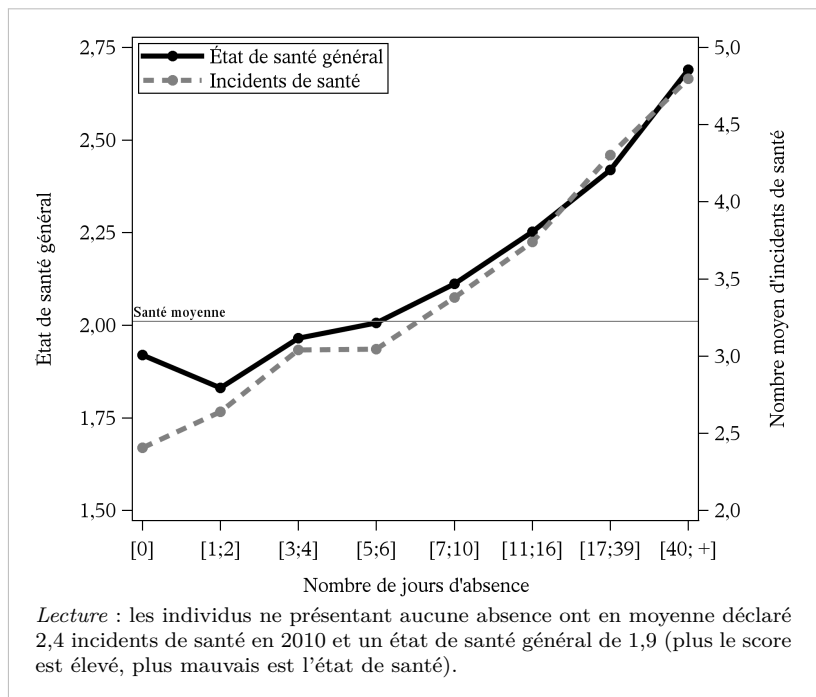
L'enquête EWCS confirme l'ensemble de ces constats. En plus d'interroger les individus sur leurs jours d'absence⁸, l'enquête EWCS recueille des informations sur leur santé. Une première question porte sur la perception générale de la santé⁹ : la réponse est donnée sur une échelle de Likert à cinq niveaux de très bon (1) à très mauvais (5). La seconde question porte sur le nombre d'incidents de santé vécus par l'individu au cours des 12 derniers mois. Une typologie de 14 symptômes est présentée¹⁰.

8. Sur les 12 derniers mois, combien de jours au total avez-vous été absent du travail pour des raisons de santé ?

9. Quel est votre état de santé en général ? Diriez-vous qu'il est très bon (1), bon (2), ni bon ni mauvais (3), mauvais (4) ou très mauvais (5) ?

10. Durant les 12 derniers mois, avez-vous souffert de l'un ou plusieurs des problèmes de santé suivants ? Problèmes d'audition, problèmes de peau, maux de dos, douleurs musculaires dans les membres supérieurs, douleurs musculaires dans les membres inférieurs, maux de tête et fatigue oculaire, douleurs d'estomac, difficultés respiratoires, maladies cardio-vasculaires, blessure(s), dépression ou anxiété, fatigue générale, problèmes d'insomnie, autres problèmes.

Graphique 2.4 – Absence, santé perçue et incidents de santé (EWCS en 2010)



Globalement, comme on le voit dans le graphique 2.4, plus l'état de santé perçu par l'individu est mauvais (valeur élevée), plus le nombre d'incidents de santé recensés est élevé, et plus le nombre de jours d'absences déclaré est élevé. Les données EWCS semblent également confirmer *a minima* l'idée que les salariés sans absence ont une moins bonne santé : le présentisme peut en être la cause. Certes des nuances s'imposent, puisque la santé moyenne de ces individus ne présente qu'une faible différence avec les catégories suivantes et parce qu'il s'agit également de la catégorie la plus représentée (56,10% de la population). Néanmoins, n'être jamais absent ne signifie pas être en parfaite santé : les individus n'ayant pas été absents déclarent également 2,4 incidents de santé en moyenne. Dans cette catégorie d'individus, seuls 28,1% des individus déclarent n'avoir eu aucun incident de santé au cours des douze derniers mois. Il est donc probable que les 71,9% restants aient travaillé au moins une fois en étant malade. Autrement dit, ils ont été présentéistes au moins une fois.

De plus, le tableau 2.2 montre une distribution typique de la santé perçue : très peu d'individus déclarent une mauvaise santé, tandis que la grande majorité s'estime en bonne santé (Bergström et al., 2009a ; Böckerman & Laukkanen, 2010 ; Montaut, 2010). Un test du χ^2 montre sans surprise que les variables d'absence et de santé ne sont pas indépendantes : les principales contributions se concentrent sur les niveaux d'absence les plus élevés : les salariés en bonne santé y sont sous-représentés tandis que les salariés en mauvaise santé y sont surreprésentés.

Tableau 2.2 – Absences et santé perçue (EWCS en 2010)

Jours d'absence	Bonne santé (1)	Ni bonne ni mauvaise	Mauvaise santé (2)	Fréquence marginale
Aucun	59,0%	48,8%	29,4%	56,1%
De 1 à 3	12,7%	8,8%	4,9%	11,8%
De 4 à 6	10,1%	9,2%	7,1%	9,8%
De 7 à 13	9,1%	12,0%	9,3%	9,7%
De 14 à 58	7,9%	16,9%	30,0%	10,3%
De 59 à 365	1,3%	4,3%	19,3%	2,3%
Total	100,0%	100,0%	100,0%	100,0%
Fréquence Marginale	76,6%	20,9%	2,5%	

(1) Salariés ayant déclaré une bonne ou très bonne santé

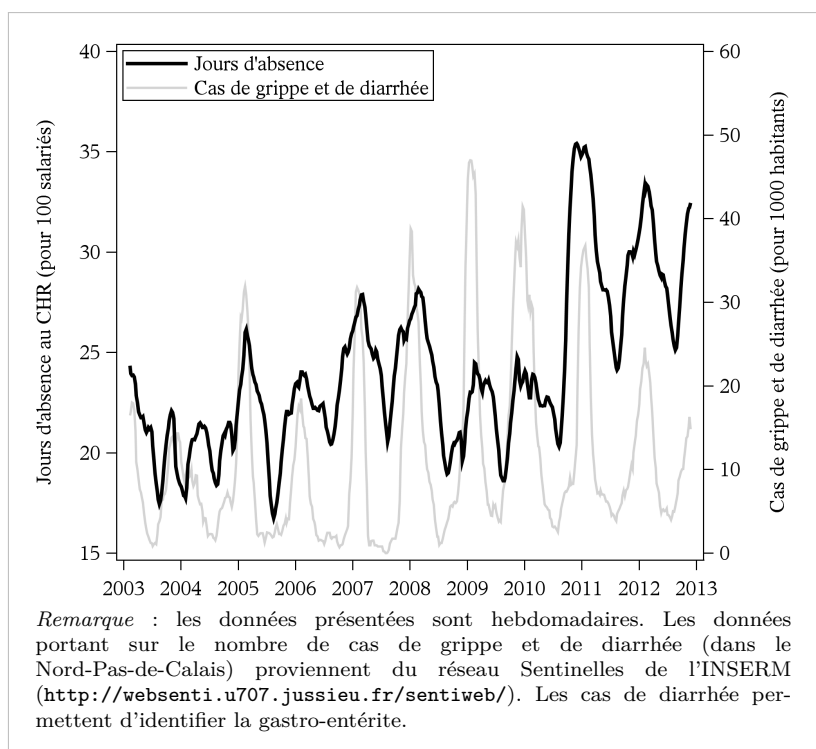
(2) Salariés ayant déclaré une mauvaise ou très mauvaise santé

2.1.3 Sensibilité aux maladies saisonnières

Bien entendu, les micro-données d'entreprise ne fournissent pas d'information directe sur la santé perçue ou le nombre d'incidents de santé des salariés. Ce qui précède laisse cependant penser que les données d'absences sont elles-mêmes porteuses d'une partie de cette information. Même si certains outils statistiques permettent de produire une analyse pertinente de l'absence-maladie sans cette information (cf. section suivante), il est intéressant de constater que les absences-maladie enregistrées par l'employeur présentent une sensibilité au contexte sanitaire et épidémiologique. D'un point de vue national, une enquête de la DARES (2013) montre par exemple que les maladies saisonnières ont un impact sur les absences-maladies : aux pics épidémiques (grippe et gastro-entérite) correspondent des pics d'absence-maladie. Nous avons transféré cette

analyse au monde de l'entreprise. Le graphique 2.5 présente le niveau hebdomadaire des absences au CHR sur une dizaine d'années et les confronte aux pics épidémiologiques enregistrés dans la région Nord-Pas-de-Calais chaque semaine. On s'aperçoit clairement que le niveau d'absence est sensible aux épidémies de grippe et de gastro-entérite.

Graphique 2.5 – Absences au CHR et épidémies saisonnières (2003-2012)



2.2 Absence, santé et âge

2.2.1 Résultats de la littérature

La littérature économique et gestionnaire regorge d'études s'attardant sur les liens entre l'âge et la durée des absences. Hélas, cette abondance de résultats n'est pas synonyme de consensus. Comme l'indique Nielsen (2008) les preuves d'une relation positive entre âge et absence côtoient celles d'une relation négative. Les tenants d'une relation positive (Frick & Malo, 2008, par exemple) estiment que celle-ci est due à la dégradation naturelle de la santé avec l'âge. D'autres auteurs, trouvant une relation

négative, avancent plusieurs explications : un plus grand bien-être dans l'emploi et dans l'entreprise lié à l'ancienneté, la difficulté de retrouver un travail en cas de licenciement (Dionne & Dostie, 2007) ou la diminution des tâches familiales pour les femmes.

Une récente étude néerlandaise semble réconcilier ces deux points de vue *a priori* divergents. Donders et al. (2012) montrent que les plus de 55 ans auraient une fréquence d'absence moins élevée, mais des périodes d'absences beaucoup plus longues. En France, Missègue (2007) met également en évidence que lorsque l'absence-maladie est étudiée en termes de durée (gravité), les seniors semblent plus touchés. Mais lorsqu'on s'intéresse au nombre de périodes d'absence (fréquence), la différence n'est plus significative. En soi, les seniors ne seraient donc pas plus fréquemment absents que les autres, mais « toutes choses égales par ailleurs, les actifs [seniors ont] un nombre de jours d'arrêt maladie plus important que leurs cadets » (Missègue, 2007, p. 12). Lorsqu'ils sont malades, les seniors seraient donc plus sévèrement touchés. Ce résultat est cohérent avec la dégradation de la santé perçue dans les dernières années de la vie au travail (DARES, 2013 ; Missègue, 2007 ; Montaut, 2010), mais aussi avec l'augmentation des dépenses de santé avec l'âge (Montaut, 2010).

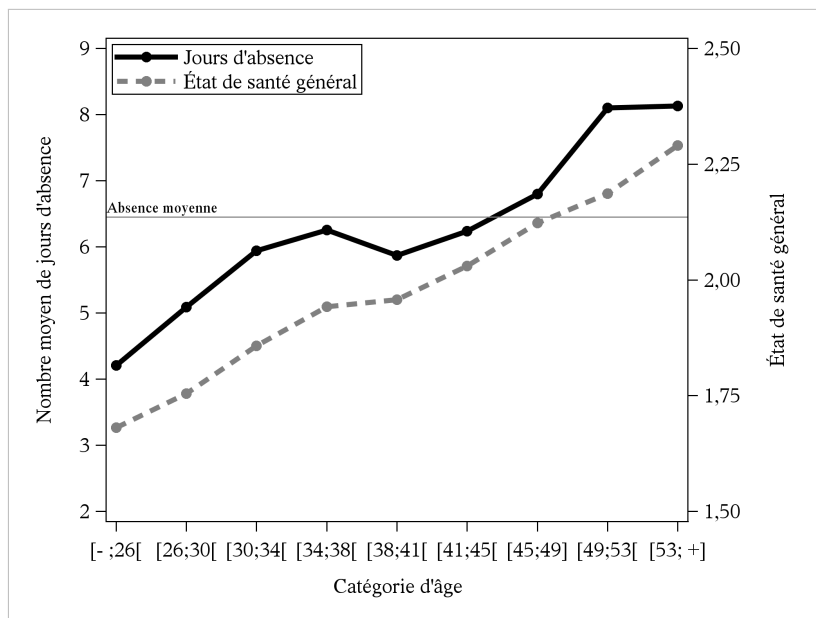
2.2.2 *European Working Conditions Survey*

Si l'on prend l'ensemble des jours d'absence et que l'on répartit les salariés en plusieurs tranches d'âge, qu'obtient-on ? Le graphique 2.6 (page 75) présente cette statistique pour l'enquête EWCS. Le nombre moyen de jours d'absence est significativement plus faible pour les premières catégories d'âge (4,2 pour les moins de 25 ans et 5,3 pour les 25-35 ans) comparativement aux seniors (8,1 pour les plus 53 ans). Entre ces deux extrêmes, la progression est relativement linéaire, hormis une légère recrudescence des absences pour les individus âgés de 30 à 38 ans. Ceci peut être lié aux pathologies relatives à la grossesse pour les femmes, voire plus largement aux maladies d'enfants à charge¹¹. La santé perçue évolue quant à elle presque linéairement avec l'âge.

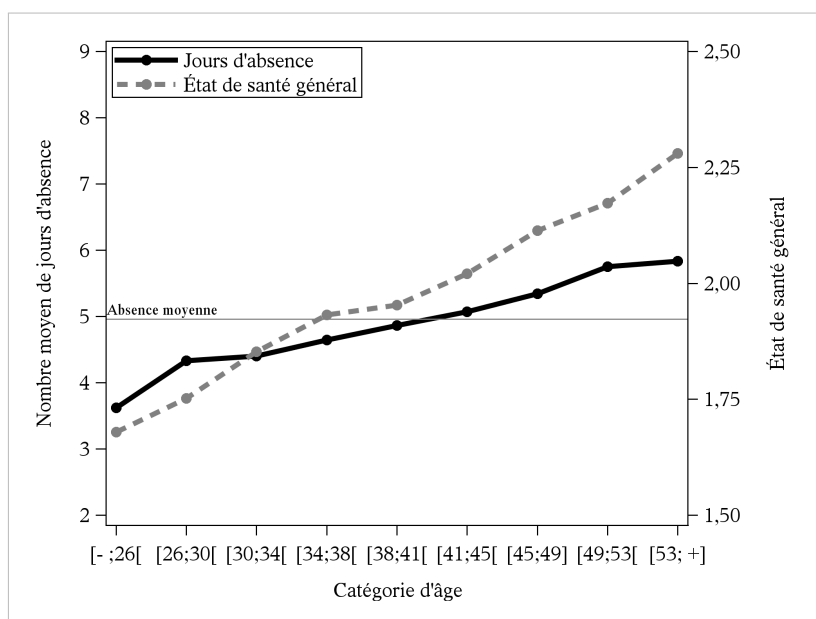
Mais la prise en compte de l'ensemble des absences a des implications qu'il est important de décoder à ce stade. En effet, deux types d'absences-maladie sont ici placés sur un pied d'égalité : les longues maladies et les absences de courte durée. Pourtant, ces deux catégories d'absence ne relèvent pas du même comportement. Les absences de longue durée sont liées à des maladies plus graves supposant de longues périodes de

11. La question posée dans l'enquête n'est pas suffisamment précise pour trancher ce point.

Graphique 2.6 – Âge, santé perçue et absences (EWCS en 2010)



Graphique 2.7 – Âge, santé perçue et absences de moins de 90 jours (EWCS en 2010)



convalescence. Si l'on renouvelle l'exercice en modifiant le périmètre d'analyse, c'est-à-dire en excluant les absences de longue durée, le constat n'est plus si tranché. Dans le graphique 2.7 (page 75), les 25% des individus ayant les plus longues absences sont exclus. La pente de la courbe d'absence est bien moins prononcée. En particulier, la moyenne d'absence des catégories d'âge les plus élevées baisse de façon nette (de 8,1 à 5,8 pour les plus de 53 ans) tandis qu'elle reste relativement stable pour les plus jeunes (de 4,2 à 3,6 pour les moins de 26 ans).

Ceci vient confirmer notre choix d'exclure de l'analyse les individus présentant un trop grand nombre d'absence (au-delà de 45 jours). Ce faisant, nous excluons les individus dont la santé est très dégradée. Or, la catégorie des seniors est davantage touchée par les longues maladies. La limitation du périmètre d'analyse permet donc également de comparer les catégories d'âge sur un pied d'égalité.

3 Modéliser les données d'absence

Jusqu'ici nous avons décrit les données d'absences très globalement. Quelques lots de données ont été présentés au moyen d'indicateurs statistiques rudimentaires, comme la moyenne ou la variance. Pour entrer plus avant dans la compréhension de cette variable d'absence et, par la suite, mettre en évidence des facteurs expliquant les phénomènes qui y sont emboîtés, il est nécessaire de recourir à une modélisation. L'idée est alors de parvenir à une distribution théorique qui s'ajusterait à la distribution empirique des données d'absence et qui serait adaptée à leurs spécificités. En partant d'une loi de Poisson, incontournable pour qui étudie des données de nature discrète, nous verrons que, lorsqu'elles sont analysées avec des instruments idoines, les données d'absence peuvent livrer des informations encore insoupçonnées. C'est là l'objet de la présente section et du chapitre à venir.

3.1 Outils classiques

Le premier outil économétrique qui se présente à l'esprit lorsqu'on cherche à modéliser une variable et à évaluer ses relations avec des variables explicatives est le modèle linéaire classique estimé par les moindres carrés ordinaires (MCO). Plusieurs arguments théoriques laissent cependant penser que l'utilisation de la méthode des MCO sur des données de comptage n'est pas optimale. De nombreux articles signalent en effet son

inadéquation (Atkins & Gallop, 2007 ; Dionne & Dostie, 2007 ; Kivimäki et al., 1997 ; Lu et al., 2010). L'un des arguments récurrents est que les données d'absences ne sont pas normalement distribuées et que l'hypothèse de normalité des résidus posée par les MCO est probablement violée (Kivimäki et al., 1997 ; Lu et al., 2010). Delgado et Kniesner (1997, p. 41) indiquent qu'une « régression par les moindres carrés ordinaires produit des estimateurs inefficaces lorsque la variable dépendante est un comptage des absences et [qu'ignorer] son hétéroscédasticité est une source d'erreurs standards inconsistantes ». Hammer et Landau (1981) conviennent également du fait que les données d'absences violent l'hypothèse d'homoscédasticité inhérente aux MCO et que certaines inférences peuvent donc s'avérer erronées à cause d'erreurs standard surestimées ou sous-estimées. Par ailleurs, l'utilisation des MCO peut nécessiter une transformation logarithmique des données, ce qui s'avère parfois problématique (O'Hara & Kotze, 2010), notamment à cause du grand nombre de valeurs nulles. La littérature ne plaide donc pas en faveur des moindres carrés ordinaires, loin s'en faut. Nous aurons l'occasion de confirmer cela dans le chapitre suivant, en nous appuyant sur une simulation.

Une variable d'absence pourrait aussi être traitée comme une variable dichotomique ou polytomique. Il serait alors envisageable d'utiliser une régression logistique pour l'analyse. Même s'il n'y a pas d'objection technique majeure, nous nous rangeons à l'avis de Lu et al. (2010) : une catégorisation des données conduirait à une perte d'information importante sur la durée des absences. L'intérêt d'exploiter la mesure de gravité des absences qui renvoie directement à la santé des individus serait fortement restreint par un tel choix.

3.2 Loi de Poisson

Les inconvénients inhérents à l'usage de méthodologies d'analyse classiques nous pousse à envisager d'autres pistes. Pour des données composées de valeurs entières positives ou nulles, les lois classiques discrètes s'imposent à l'évidence. Parmi celles-ci, la loi de Poisson semble la plus pertinente. En effet, les données d'absences sont typiquement des données de comptage et la distribution de Poisson s'applique précisément à un dénombrement (Greene, 2005). Dans notre cas, il s'agit de dénombrer pour chaque salarié le nombre de jours d'absence pendant une année donnée. Comme d'autres (Dionne & Dostie, 2007 ; Kivimäki et al., 1997 ; Missègue, 2007 ; Winkelmann, 1999), nous sommes donc amenés à considérer les avantages et les inconvénients d'une loi de Poisson.

3.2.1 *Présupposés*

La loi de Poisson comptabilise un nombre d'événements rares se produisant sur une période de temps défini. Elle peut être envisagée « comme la loi du nombre de succès dans des épreuves de Bernoulli indépendantes et de même paramètre » (Delsart et al., 2009, p. 104). Or, le nombre annuel de jours d'absences peut aussi être appréhendé en ce sens (Winkelmann, 1999). Au cours d'une année de 365 jours, il y a 365 tirages de Bernoulli (ou 260 si l'on raisonne en jours ouvrés) : chaque jour, un salarié est absent ou présent, avec une probabilité p ou $1 - p$, respectivement. Si la probabilité p est constante et que les tirages sont indépendants, alors le nombre annuel de jours d'absence suit une loi de Poisson. On peut l'exprimer formellement de la façon suivante : si Y suit une loi de Poisson dont l'espérance est λ (avec $\lambda > 0$), alors

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad (2.1)$$

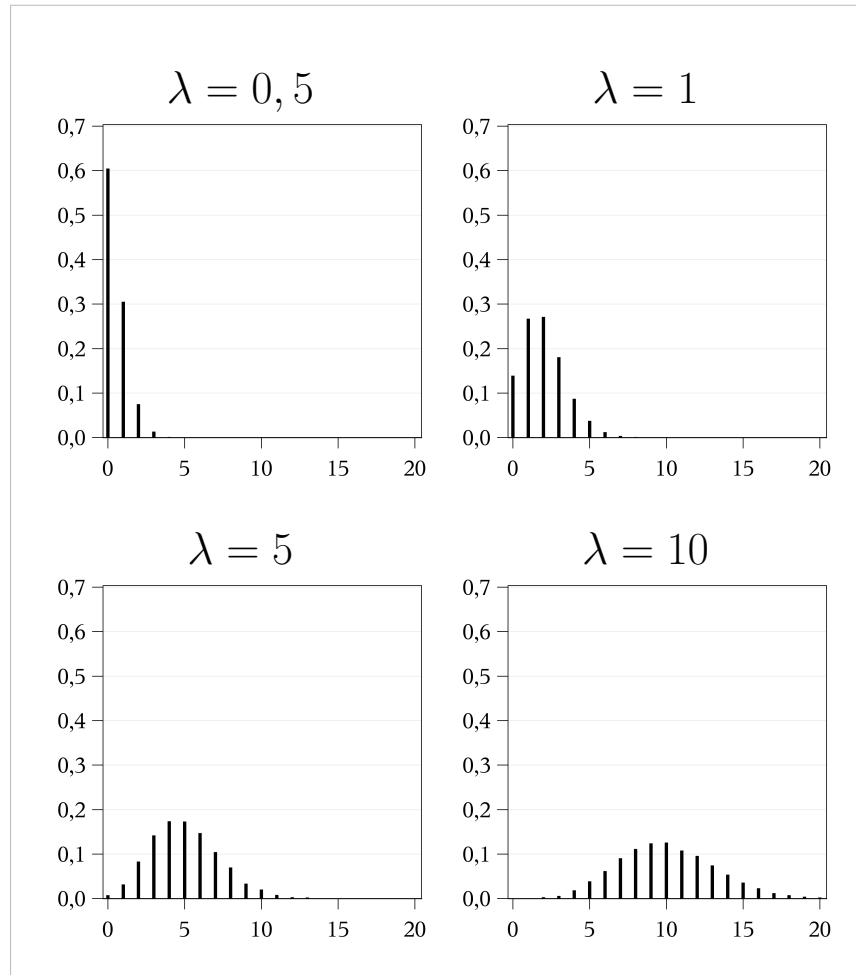
où la variable observée y est un entier positif ou nul.

L'espérance λ est ici égale à la variance : une variable strictement poissonnienne est donc *equidispersée*. Si cela n'est pas vérifié, on parle de *surdispersion* lorsque la variance est supérieure à l'espérance ou de *sous-dispersion* dans le cas inverse (Winkelmann, 2008).

La loi de Poisson prend des formes diverses suivant la valeur de son espérance λ . Comme on le voit sur le graphique 2.8 (page 79), pour des valeurs faibles de λ , la distribution est fortement asymétrique, les valeurs nulles étant nombreuses. Mais pour des valeurs plus élevées de λ , la distribution se symétrise (Delsart et al., 2009).

3.2.2 *Loi de Poisson et individus hétérogènes*

La loi de Poisson ne semble pas inadaptée pour décrire les absences d'un individu donné. Dans l'ensemble de ce travail, les distributions individuelles sont toujours considérées comme poissonniennes. Un argument pourrait être opposé à ce choix. La loi de Poisson postule que les tirages successifs, présentés plus haut comme des épreuves de Bernoulli, sont indépendants. Or, les données d'absences incluent une sorte d'effet de mémoire ou de contagion : le fait d'être absent un jour dépend en partie du fait que le salarié était déjà malade la veille (Barmby, 2002 ; Barmby & Sibly, 2004). Néanmoins, comme le montre parfaitement Winkelmann (2008), ce problème de dépendance des événements

Graphique 2.8 – Distributions de Poisson selon la moyenne λ 

peut être pris en compte du point de vue de la distribution totale des absences par l'utilisation d'une loi binomiale négative. Si l'on s'en tient aux distributions individuelles, la loi de Poisson reste donc de mise.

Mais dès lors que l'on s'intéresse à plusieurs individus ayant des moyennes d'absence différentes, l'égalité entre moyenne et variance supposée par la loi de Poisson ne tient plus. En effet, les données d'absences sont caractérisées par une surdispersion. Cet argument est parfois avancé pour rejeter la loi de Poisson dans la modélisation des absences. En fait, il serait plus juste de dire qu'une loi de Poisson unique d'espérance donnée ne régit pas les absences de l'ensemble des individus. Ceux-ci sont en effet hétérogènes : leurs distributions de probabilités individuelles sont certes poisson-

niennes, mais de paramètres différents. De ces différences inter-individuelles découle la surdispersion caractérisant la distribution de l'ensemble des individus. Elle pourrait en partie être résorbée par l'inclusion de variables indépendantes dans le modèle (démographiques, organisationnelles, etc.), mais peut aussi être due à des éléments inobservés : les différences d'état de santé ou de robustesse physique, par exemple. Il est donc nécessaire d'envisager les traitements possibles de cette surdispersion.

3.3 Traitement de l'hétérogénéité inter-individuelle

Pour résumer ce qui vient d'être dit, la loi de Poisson convient pour modéliser l'absence d'un individu isolé. Mais lorsqu'on étudie un grand nombre d'individus dont les distributions individuelles sont hétéroclites, émerge un problème de surdispersion. Comment, sans se départir des forces d'une loi de Poisson, peut-on tenir compte de l'hétérogénéité interindividuelle qui est à l'origine de la surdispersion des données d'absence ?

La réponse la plus simple consiste à repérer directement les causes de surdispersion dans les données. Hilbe (2011) propose plusieurs pistes pour y remédier :

- Y a-t-il un grand nombre de valeurs nulles ? C'est le cas des données d'absences, mais il paraît inopportun d'exclure ces valeurs nulles.
- Existe-t-il des données aberrantes ? À l'origine oui, mais lorsque le périmètre d'analyse est limité, les absences très longues sont exclues.
- Une variable explicative importante a-t-elle été omise ? Nous soupçonnons effectivement qu'une source inobservable produit de l'hétérogénéité.
- Des sous-populations ayant un comportement spécifique cohabitent-elles ? Il s'agit là d'une forme d'hétérogénéité particulière qui sera analysée dans le chapitre suivant.

Lorsqu'on ne peut résoudre le problème en corrigeant *ex ante* ces sources d'hétérogénéité, par exemple lorsqu'une partie de l'hétérogénéité demeure inobservable, il est possible d'inclure un paramètre captant cette hétérogénéité. Ceci nous conduit à la loi binomiale négative qui demeure une distribution de forme poissonnienne. Il s'agit en fait d'un mélange Poisson-Gamma. Dans le même esprit, l'hétérogénéité caractéristique des données d'absence pourrait aussi être décelée en faisant l'hypothèse d'une coexistence de plusieurs sous-populations dont les comportements diffèrent. Il pourrait donc être envisageable de modéliser les données de chaque sous-population par une loi de Poisson, autrement dit par un mélange de distributions poissonniennes.

3.3.1 Mélange de lois de Poisson

Une première piste pour tenir compte de la surdispersion des données d'absence est de distinguer des sous-populations dont les absences suivraient chacune une loi de Poisson d'espérance donnée. En prenant l'exemple de deux sous-populations dont les données respectives suivent une loi de Poisson, i.e. $X_1 \sim P(\lambda_1)$ et $X_2 \sim P(\lambda_2)$ avec une probabilité d'occurrence de p (resp. $q = 1 - p$), on peut en effet montrer que la variance est globalement plus élevée que l'espérance :

$$\begin{aligned}
 E(X) &= \sum p_i \lambda_i = p\lambda_1 + q\lambda_2 \\
 V(X) &= \sum p_i \lambda_i + \sum p_i (\lambda_i - E(X))^2 \\
 &= p\lambda_1 + q\lambda_2 + p(\lambda_1 - (p\lambda_1 + q\lambda_2))^2 + q(\lambda_2 - (p\lambda_1 + q\lambda_2))^2 \\
 &= p\lambda_1 + q\lambda_2 + pq^2(\lambda_1 + \lambda_2)^2 + qp^2(\lambda_1 + \lambda_2)^2 \\
 &= p\lambda_1 + q\lambda_2 + pq(p + q)(\lambda_1 + \lambda_2)^2 \\
 &= p\lambda_1 + q\lambda_2 + pq(\lambda_1 + \lambda_2)^2 \\
 &= E(X) + pq(\lambda_1 + \lambda_2)^2
 \end{aligned}$$

On trouve bien que $V(X) \geq E(X)$. Bien entendu, cela peut se généraliser à un nombre j de sous-populations. Avec $X_j \sim P(\lambda_j)$ et une probabilité d'occurrence p_j , le résultat obtenu est le même :

$$\begin{aligned}
 E(X) &= \sum p_j \lambda_j \\
 V(X) &= \sum p_j \lambda_j + \sum p_j (\lambda_j - E(X))^2 \\
 &= E(X) + \sum p_j (\lambda_j - E(X))^2
 \end{aligned}$$

En soi, l'idée d'un mélange de loi de Poisson serait donc envisageable. L'une des raisons nous ayant progressivement conduit à privilégier une distribution binomiale négative est que les données dont nous disposons n'offrent aucun moyen de diviser *ex ante* la po-

pulation en plusieurs groupes dont les distributions individuelles seraient comparables. Étant donnée la nature des données d'absence, on pourrait en venir à diviser la population en autant de distributions poissonniennes qu'il y a d'individus, mais l'estimation s'avèrerait alors impossible. La distribution binomiale négative, qui est elle-même un mélange de lois, est à la fois d'utilisation plus souple et d'implémentation plus fréquente dans les logiciels d'analyse statistique. La base de l'estimation demeure poissonnienne, mais le mélange de distributions se joue différemment.

3.3.2 Loi binomiale négative

Le moyen traditionnel de contourner les problèmes inhérents à une loi de Poisson standard pour la modélisation des absences d'un grand nombre d'individus, est donc d'utiliser une loi binomiale négative (Barmby et al., 2001 ; Frick & Malo, 2008 ; Jensen & MacIntosh, 2007 ; Winkelmann, 1999). Cette distribution est une généralisation ou une amélioration technique de la loi de Poisson : il s'agit d'y introduire un paramètre prenant en compte l'hétérogénéité entre les espérances poissonniennes des individus de la population. On introduit donc un paramètre u multiplicatif de λ :

$$P(Y = y) = \frac{e^{-\lambda u} (\lambda u)^y}{y!} \quad (2.2)$$

Pour modéliser ce paramètre u , on utilise habituellement une loi Gamma, qui est bien adaptée à la nature particulière du paramètre d'hétérogénéité. Deux raisons essentielles conduisent à ce choix. D'une part, le paramètre u , de par sa nature, doit être positif, puisqu'il est entré multiplicativement de λ dans la loi de Poisson. Or, λ est toujours strictement positif¹². D'autre part, en l'absence d'informations sur la distribution de ce paramètre d'hétérogénéité, il est nécessaire d'opter pour une loi statistique suffisamment souple pour s'ajuster à des distributions de formes diverses. C'est l'une des grandes forces de la loi Gamma.

12. D'après Winkelmann (2008), trois distributions générant des valeurs réelles positives conviendraient *a priori* : la distribution log-normale, l'inverse-gaussienne et la loi Gamma. L'inconvénient des deux premières est que, dans un mélange avec une loi de Poisson, il n'existe pas de solution analytique pour l'estimation du maximum de vraisemblance. Même si des techniques d'estimation adéquates ont peu à peu vu le jour par la suite (Winkelmann, 2008), le mélange Poisson-Gamma s'est imposé comme un standard à cause de sa commodité mathématique.

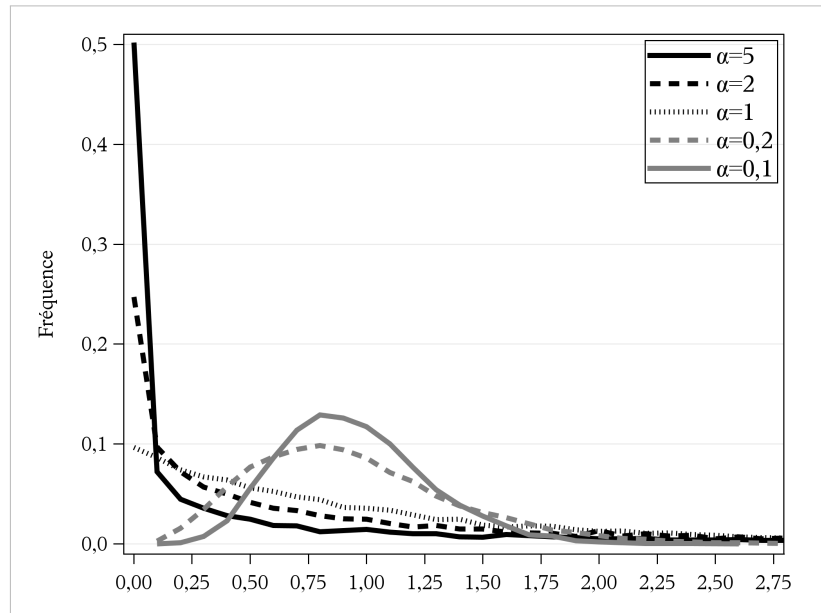
Lorsque la loi Gamma prend un paramètre de forme β et d'échelle δ , elle peut s'écrire :

$$g(u) = \frac{\delta^\beta}{\Gamma(\beta)} u^{\beta-1} e^{-\delta u} \quad (2.3)$$

Son espérance est alors $E(u) = \beta/\delta$ et sa variance $V(u) = \beta/\delta^2$. Dans la distribution binomiale négative, les deux paramètres de la loi Gamma incluse sont égaux, tel que $u \sim \Gamma(\theta, \theta)$. On obtient donc $E(u) = 1$ et $V(u) = 1/\theta$. L'espérance de la loi binomiale négative sera donc finalement égale à l'espérance λ de la loi de Poisson, mais la variance, dont la valeur dépend du paramètre θ , est différente.

Comme le montre le graphique 2.9, les situations d'hétérogénéité interindividuelle dont est capable de tenir compte la loi Gamma sont très diverses. La forme de la loi Gamma est souvent exprimée à partir de l'inverse de θ , c'est-à-dire $\alpha = \frac{1}{\theta}$. Le paramètre α est en effet d'interprétation plus commode : lorsqu'il augmente, la variance augmente également, et avec elle la surdispersion¹³. Pour des valeurs d' α tendant vers zéro, l'hétérogénéité entre individus est très faible, puis, à mesure que la valeur d' α augmente, les écarts entre les individus augmentent également.

Graphique 2.9 – Distributions de la loi Gamma selon son paramètre α



13. La plupart des logiciels (comme SAS, celui que nous utilisons) estiment d'ailleurs la valeur d' α .

Lorsqu'on poursuit l'idée et que l'on intègre la distribution Gamma du paramètre d'hétérogénéité à la loi de Poisson, on obtient la distribution binomiale négative (la démonstration est proposée en annexe 1) :

$$P(Y = y) = \frac{\Gamma(\theta + y)}{\Gamma(y + 1)\Gamma(\theta)} \left(\frac{\theta}{\theta + \lambda} \right)^\theta \left(\frac{\lambda}{\theta + \lambda} \right)^y \quad (2.4)$$

L'espérance de la distribution binomiale négative est donc identique à celle de la distribution purement poissonnienne, $E(Y) = \lambda$, mais la variance est désormais :

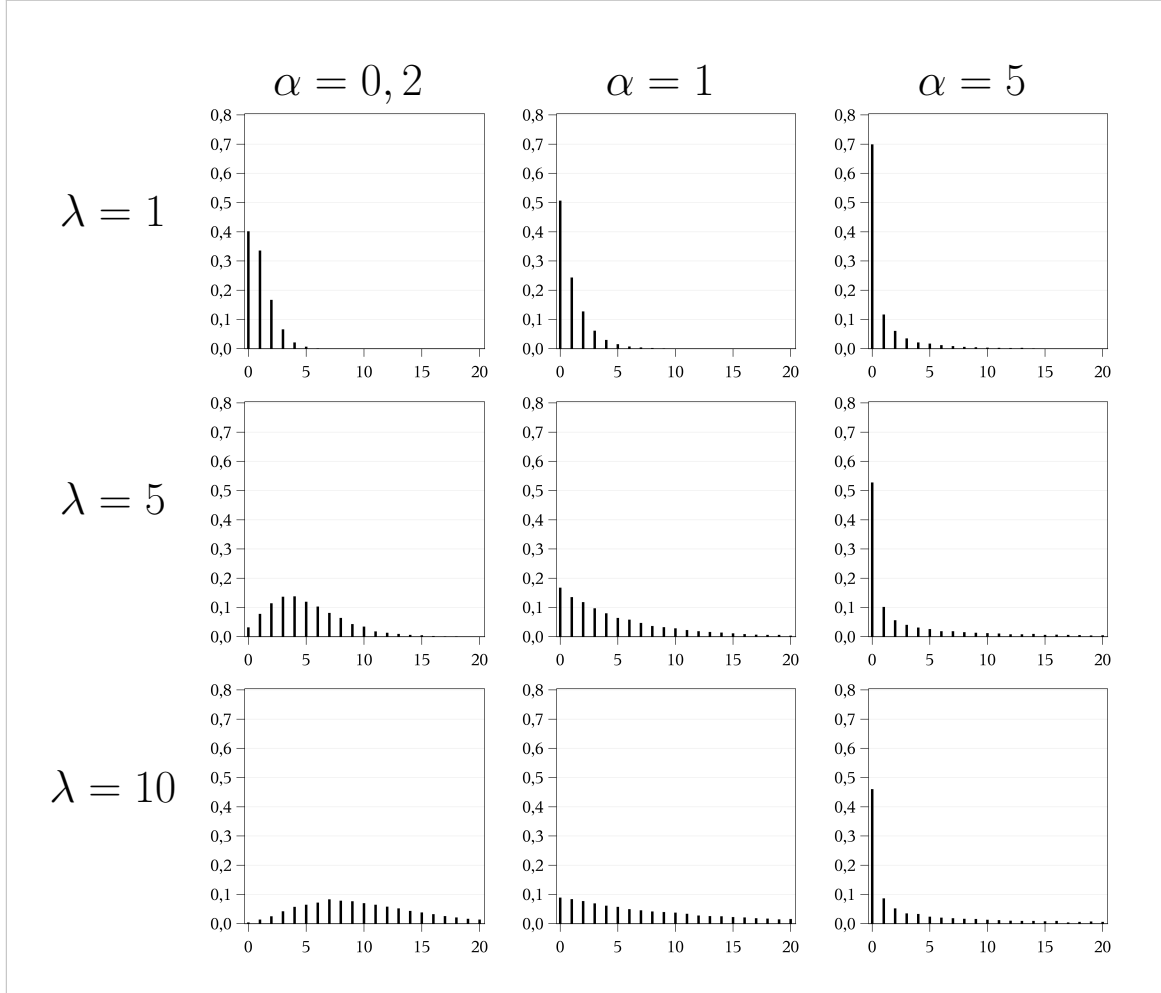
$$V(Y) = \lambda \left(1 + \left(\frac{1}{\theta} \right) \lambda \right)$$

$$V(Y) = \lambda(1 + \alpha\lambda) = \lambda + \alpha\lambda^2 \quad \text{avec } \alpha = \frac{1}{\theta} \quad (2.5)$$

Il existe donc désormais une relation quadratique entre l'espérance et la variance de la loi binomiale négative qui dépend du paramètre α . Ce paramètre α sera estimé comme une variable explicative du modèle et apparaîtra dans les résultats finaux. Nous reviendrons largement sur ce paramètre dans les lignes à venir et dans le chapitre suivant, parce qu'il est d'une importance majeure.

La loi binomiale négative, comme celle de Poisson, permet donc de modéliser l'espérance d'occurrence d'un événement. Elle intègre un paramètre d'hétérogénéité α autorisant une surdispersion de la variable dépendante. Le graphique 2.10 (page 85) présente plusieurs distributions selon différentes valeurs de λ et de α . Lorsque α est très proche de zéro, la distribution binomiale négative est comparable à une distribution de Poisson simple. Pour $\alpha = 0,2$ et $\lambda = 1$, le nombre de valeurs nulles générées dans la loi binomiale négative est quasiment équivalent. Mais lorsque α augmente, un très grand nombre de valeurs nulles apparaissent et la queue de la distribution s'allonge, même pour des espérances élevées. La loi binomiale négative s'avère donc en mesure d'appréhender à la fois un grand nombre de valeurs nulles et des valeurs positives très élevées, ce qui est impossible avec une simple loi de Poisson.

Il est important de bien distinguer le paramètre α (qui donne une indication de la forme de la loi Gamma) du facteur d'hétérogénéité u (qui est lui distribué selon la loi Gamma de paramètre α). Le paramètre α est intégré à la loi binomiale négative et une

Graphique 2.10 – Distributions binomiales négatives selon λ et α 

estimation en est fournie dans les logiciels implémentant cette distribution. En tant que lien entre l'espérance et la variance, le paramètre α indique donc directement le degré de surdispersion de la distribution. On peut réaliser un test de nullité du paramètre α pour confirmer qu'un modèle binomial négatif est préférable à un modèle de Poisson. En effet, lorsque α tend vers zéro, alors la variance du modèle tend vers λ : le modèle de Poisson est dit « emboîté » dans le modèle binomial négatif. Si le paramètre α est significativement différent de zéro, alors un modèle binomial négatif est à privilégier.

3.3.3 Deux aspects de l'hétérogénéité

L'hétérogénéité inter-individuelle des données d'absence a un double visage. L'utilisation d'une loi binomiale négative permet de tenir compte des différences inter-individuelles sans réellement les qualifier. En effet, jusqu'ici, il n'a pas encore été question de variables explicatives. Schématiquement, si l'on devait modéliser à ce stade, on aurait un modèle de Poisson incluant simplement un effet aléatoire dont on pense qu'il suit une loi Gamma. Cependant, parallèlement aux données d'absence, nous disposons d'un ensemble d'informations individuelles permettant de caractériser plus finement les différences entre les individus et de proposer un modèle incluant un vecteur de variables explicatives. Il s'agit en quelque sorte de la dimension observable de l'hétérogénéité inter-individuelle.

L'espérance λ est une fonction de ce vecteur de variables explicatives. Les variables explicatives sont introduites dans le modèle sous une forme exponentielle¹⁴. En effet, si on considère que $X_i = [x_{1_i} \ x_{2_i} \ \dots \ x_{k_i}]$ est un vecteur de k variables explicatives :

$$\lambda_i = \exp(X_i^T \beta) \quad (2.6)$$

ou sous une forme logarithmique :

$$\log(\lambda_i) = X_i^T \beta \quad (2.7)$$

À partir du moment où une partie de l'hétérogénéité inter-individuelle est expliquée par une batterie de variables, le paramètre u inclus dans le modèle binomial négatif capte l'hétérogénéité résiduelle qui demeure inobservée. C'est le deuxième aspect de l'hétérogénéité inter-individuelle. L'enjeu véritable de l'inclusion d'un paramètre d'hétérogénéité est alors de permettre au chercheur de modéliser efficacement un phénomène sur la base des variables disponibles. Une variable (ou un vecteur de variables) peut demeurer inobservée sans que la pertinence de l'analyse soit fondamentalement remise en cause. La connaissance de ces éléments inobservés captés par le paramètre d'hétérogénéité découle bien entendu de l'expérience de l'analyste. Dans le cas des données d'absence, l'interprétation la plus vraisemblable est que les niveaux de santé des individus soient variables : n'ayant pas les mêmes risques de contracter une maladie,

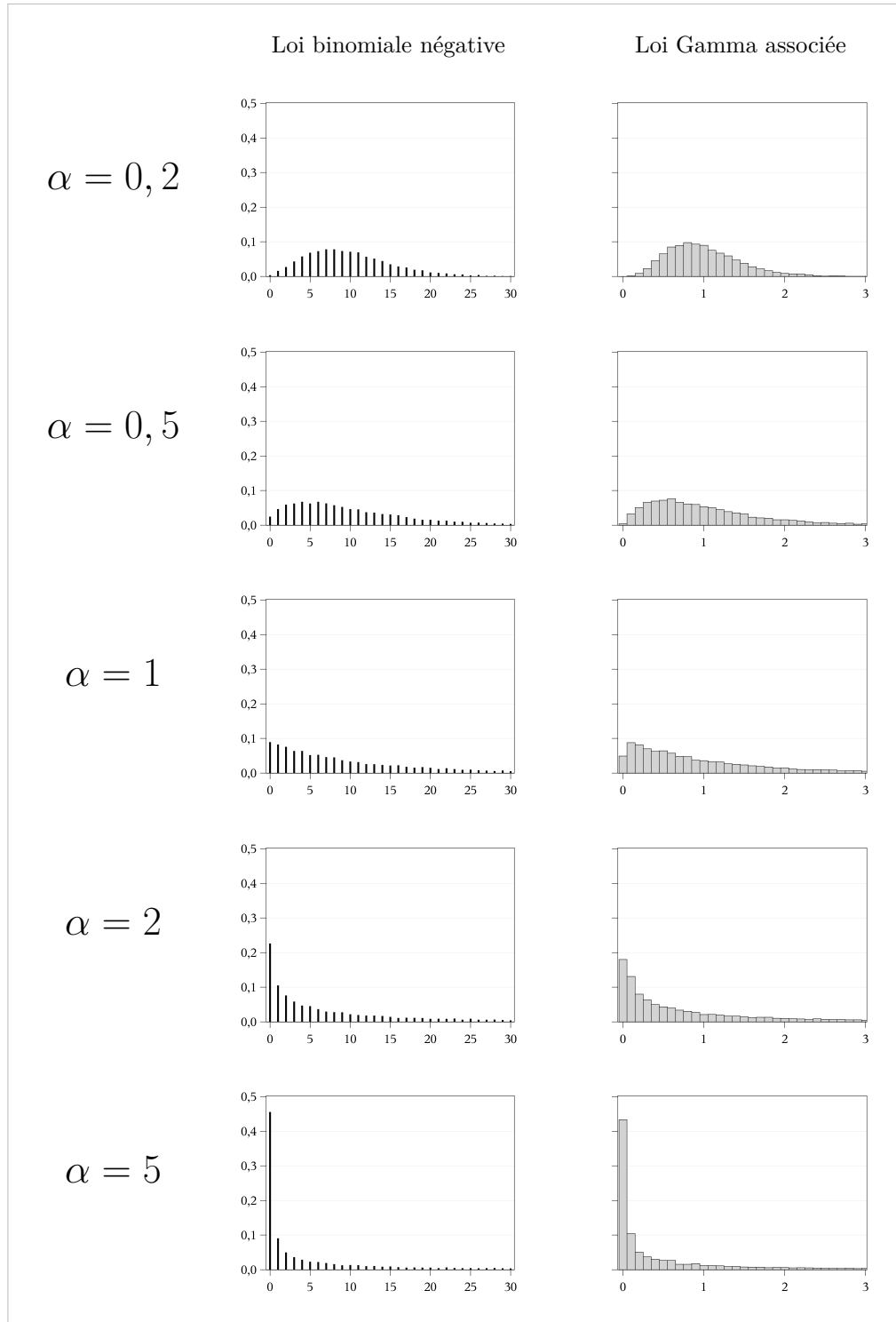
14. Ce qui fait de la loi de Poisson un modèle de forme log-linéaire (Greene, 2005).

il n'ont pas la même propension à l'absence ni au présentéisme. Il peut aussi s'agir de la robustesse physique des individus, de traits psychologiques, de dispositions morales, etc. Étudier des micro-données d'entreprise implique souvent de renoncer à la mesure de ces phénomènes.

La souplesse du paramètre d'hétérogénéité est une réelle force pour modéliser les données d'absences. Sans information complémentaire sur la distribution du paramètre, il est possible de faire face à de nombreux scénarios. Si l'on assimile par exemple le paramètre d'hétérogénéité à l'état de santé des individus, l'estimation reste pertinente aussi bien lorsque l'ensemble des individus ont un état de santé homogène (concentration autour de la moyenne) que lorsque les niveaux de santé sont très divergents (dispersion autour de la moyenne). En cas de concentration autour de la moyenne, α est proche de zéro et tous les individus ont une santé proche de 1. En cas de forte hétérogénéité, par exemple avec $\alpha = 5$, de nombreux individus ont une santé comprise entre 0 et 1, et quelques-uns une valeur très élevée. L'espérance d'absence des premiers (λ multiplié par une valeur inférieure à 1) baisse alors considérablement, tandis que l'espérance des seconds atteint une valeur très élevée (λ est démultiplié). Comme nous l'avons vu au début de ce chapitre, le scénario le plus probable est d'ailleurs qu'il y ait beaucoup d'individus en bonne santé et quelques individus en très mauvaise santé. Dans l'enquête EWCS, par exemple, 25,3% des individus se disaient en très bonne santé en 2010 tandis que seuls 2,48% se disaient en mauvaise ou en très mauvaise santé. Ce type de distribution est d'ailleurs observé dans d'autres travaux (Bergström et al., 2009a ; Böckerman & Laukkanen, 2010 ; Montaut, 2010).

Sous un autre angle, on peut aussi interpréter directement les valeurs individuelles du paramètre u . Si pour un individu i , $u_i = 0,2$, cela signifie que son espérance d'absence sera 5 fois moins élevée que celle d'un individu de santé standard ($u_i = 1$). Inversement, si $u_i = 3$, son espérance d'absence sera trois fois plus élevée.

Le graphique 2.11 (page 88) présente les formes du paramètre d'hétérogénéité u suivant différentes valeurs du paramètre α . Dans cet exemple, l'espérance du modèle binomial négatif est toujours $\lambda = 10$. On voit bien comment, lorsque α augmente, la distribution du terme d'hétérogénéité et la loi binomiale négative se transforment peu à peu.

Graphique 2.11 – Distributions de la loi binomiale négative (avec $\lambda = 10$)

Conclusion

Ce second chapitre était consacré à la découverte de la variable dépendante qui sera exploitée dans la suite de notre thèse. Ses spécificités méthodologiques et sa nature composite ont souvent alimenté les débats de la recherche autour de l'absence-maladie. Il ne nous semblait donc pas superflu de nous étendre sur la description de cette variable et sur les moyens à envisager pour sa modélisation. La méthodologie de mesure présentée dans ce travail repose largement sur le fait que les données d'absence captent, en parallèle à l'absence-maladie, un ensemble de comportements de la main d'œuvre, dont le présentéisme fait partie. Selon nous cette propriété n'a pas suffisamment attiré l'attention de la recherche et il est urgent de combler une telle lacune. En ce sens, notre thèse peut également constituer une contribution à l'analyse des données d'absence.

En guise de synthèse, nous reprenons ici les principaux enseignements de ce chapitre sur les données d'absence. La première section se résume en une simple phrase : la mesure de l'absence pour laquelle nous optons repose sur le dénombrement du nombre de jours d'absence-maladie par salarié et par an, à l'exclusion des longues maladies. Chaque élément de cette phrase relève d'un choix précis.

« **Le nombre de jours d'absence-maladie** ». L'absence est exprimée en termes de durée plutôt que de fréquence (i.e. le dénombrement du nombre d'épisodes d'absence). Appréhendée sous l'angle de la gravité (durée), l'absence est davantage liée à la question de la santé et de la maladie. Ceci est indispensable dans la perspective de mesurer le présentéisme, puisque ce comportement est une réponse possible à un incident de santé.

« **Par salarié** ». Le point de vue individuel est privilégié à la perspective collective. Puisque nous cherchons à analyser le choix d'un individu lors d'un épisode de maladie, agréger les absences au niveau d'une équipe de travail ou d'un département de l'entreprise serait moins cohérent, dans notre cas.

« **Par an** ». Une période d'analyse d'une année permet de tenir compte de l'aspect cyclique de l'absence et constitue la norme en entreprise (e.g. bilan social).

« **À l'exclusion des longues maladies** ». Les données d'absences sont caractérisées par une forte concentration sur quelques individus. Autrement dit, un très petit nombre d'individus génère la majorité des absences. Les cas de maladies lourdes (et donc longues) sont exclues du périmètre d'analyse. En effet, un individu se trouvant

dans cette situation n'a de toute façon pas d'autre choix que de se retirer et de se soigner : le présentéisme n'entre pas en ligne de compte.

Ensuite, dans la deuxième partie de ce chapitre, nous sommes revenu sur un aspect de l'absence-maladie qui, tout en paraissant presque trivial, a pourtant été négligé dans la recherche en économie (hormis en économie de la santé) et en gestion. L'absence de la main d'œuvre est souvent considérée comme un indicateur d'engagement et de motivation de celle-ci, ou comme la conséquence d'une pression de l'organisation sur l'individu. En mettant l'accent sur ces facteurs relatifs au travail, on a presque oublié que l'absence est d'abord en relation avec l'état de santé de l'individu et avec l'incidence de diverses maladies. Cette deuxième partie a donc donné la parole à des travaux réhabilitant cette relation entre absence et santé et a proposé quelques statistiques descriptives en ce sens. La corrélation de l'absence et de l'état de santé est très importante pour les chapitres à venir, notamment lorsqu'il s'agira d'interpréter nos résultats.

Enfin, pour remplir notre objectif de faire la part du présentéisme dans la variable dépendante, il était nécessaire de trouver une loi statistique capable de modéliser au mieux le nombre de jours d'absence. La capacité d'une loi de Poisson à modéliser des événements rares se produisant sur une période de temps donnée aurait pu être exploitée pour les absences d'un individu donné. Mais les absences d'un grand nombre d'individus se caractérisent par une forte hétérogénéité inter-individuelle produisant une surdispersion statistique, c'est-à-dire une variance plus élevée que la moyenne. En raison de cette surdispersion, la loi de Poisson doit être aménagée pour modéliser les absences de plusieurs centaines d'individus ayant chacun leur propre espérance d'absence. La loi binomiale négative, qui peut être appréhendée comme une généralisation de la loi de Poisson, inclut un paramètre de dispersion captant ces différences inter-individuelles. Ce paramètre suit lui-même une loi statistique suffisamment souple (loi Gamma) pour s'adapter à différents scénarios. Deux formes d'hétérogénéité entre les individus ont été identifiées dans cette section. Soit une forme observable, que l'on peut mesurer par les variables explicatives introduites dans le modèle, soit une forme inobservée, résultant de l'absence de certaines variables dans le modèle. Dans notre cas, il s'agit par exemple de l'état de santé des individus ou de leur profil psychologique. En utilisant une loi binomiale négative, il est possible d'estimer correctement le poids respectif des variables explicatives observées, tout en isolant l'hétérogénéité non-observée.

Chapitre 3

Modèles à inflation de zéros

Introduction

Dans le chapitre précédent, nous avons donné un aperçu de la nature des données d'absence et avons trouvé un moyen de les modéliser correctement, en tenant compte de leurs spécificités. Parmi celles-ci, la présence d'un grand nombre de valeurs nulles a été citée en premier lieu. On peut trouver de nombreux autres exemples de données dans lesquelles les valeurs nulles sont représentées en abondance. En économie, les travaux s'intéressant au nombre de rebuts dans un processus de production (Lambert, 1992), au nombre de brevets déposés par un chercheur (Carayol, 2007) ou encore à la fréquence des accidents du travail (Campolieti, 2002) sont également confrontés à cette problématique de la surreprésentation des valeurs nulles. L'analyste peut prendre différents partis vis-à-vis de ce problème. D'abord, il peut considérer que les zéros constituent simplement une valeur parmi d'autres et qu'il n'y a aucune raison de s'en soucier. Il peut également ignorer ces valeurs et choisir de tronquer la variable pour ne

s'intéresser qu'aux observations positives. Enfin, il peut adopter un regard binaire et étudier la variable selon la réalisation ou la non-réalisation de l'événement.

Les mêmes questions se posent vis-à-vis des données d'absences. On peut étudier les seuls individus qui s'absentent, dichotomiser la variable d'absence ou considérer les valeurs nulles comme équivalentes aux autres. Dans cette thèse nous allons faire un autre choix. Non seulement nous prenons en considération l'ensemble des valeurs prises par la variable d'absence, valeurs nulles comprises, mais nous posons l'hypothèse (que nous testons par la suite) que ces valeurs nulles peuvent désigner dans certains cas des individus qui n'ont pas été malades et dans d'autres des individus malades ayant travaillé malgré tout.

Ce problème de « double-origine » des valeurs nulles peut-être envisagé comme une autre forme d'hétérogénéité, que l'on pourrait nommer « comportementale ». En effet, dans le chapitre précédent, deux formes d'hétérogénéité ont été mises en évidence. La première est inhérente à tout travail économétrique. Les individus se comportent différemment par rapport à un standard (en termes statistiques on dirait qu'il y a des écarts par rapport à la moyenne) : l'analyste cherche à expliquer cette variabilité en recherchant des relations avec des variables explicatives. On pourrait parler d'hétérogénéité « observée ». Le second type d'hétérogénéité demeure quant à lui inobservé. Dans les données d'absence, il s'agit typiquement de l'état de santé ou de la robustesse physique, dont le modèle binomial négatif est capable de tenir compte. Ici apparaît un troisième type d'hétérogénéité. Il concerne les différences comportementales observées entre deux sous-populations. Dans l'une, les individus s'absentent lorsqu'ils sont malades, dans l'autre, celle des présentéistes, les individus travaillent même malades.

Dans ce chapitre, nous montrerons comment, en enrichissant le modèle binomial négatif standard, il est possible de mesurer cette troisième forme d'hétérogénéité en « triant » les valeurs nulles. La boîte à outils des modèles de comptage contient en effet des modèles dits à inflation de zéros (*zero-inflated*) qui le permettent. Les deux premières sections de ce chapitre sont consacrées à une description des modèles *zero-inflated* et à quelques illustrations de leur exploitation dans la littérature. Même si nous entrons dans les détails théoriques des modèles, nous souhaiterions faire œuvre de pédagogie afin de permettre au lecteur de saisir notre stratégie et la richesse de l'instrument que nous présentons. C'est l'esprit de la seconde section dans laquelle est présentée une batterie de simulations de Monte-Carlo. Le but de celles-ci est à la fois de faciliter

la compréhension des modèles à inflation de zéros par une application à des données simulées et d'observer leur comportement lorsque leurs hypothèses sont mal respectées.

1 Traitement de l'hétérogénéité comportementale

Dans cette section, nous présentons les fondements théoriques des modèles à inflation de zéros, de leur estimation et de leur interprétation. Mais auparavant, nous voudrions encore une fois insister sur l'idée de double-origine des valeurs nulles d'absence. C'est l'intuition centrale de ce travail, sur laquelle reposent les résultats à venir.

1.1 Idée maîtresse

Le modèle binomial négatif, comme nous l'avons observé, est capable de tenir compte d'un nombre conséquent de valeurs nulles, grâce à la souplesse du paramètre d'hétérogénéité qu'il intègre. Cependant, ce modèle considère que l'ensemble des valeurs de la distribution émanent d'un seul et même processus de génération. Schématiquement, l'absence au travail est alors appréhendée comme un choix binaire que chaque salarié pose quotidiennement : absence en cas de maladie, sinon présence au travail. La valeur obtenue n'est alors rien d'autre que la somme annuelle des jours de maladies, les valeurs nulles exprimant seulement l'absence de maladie.

Or, la réalité est bien plus complexe et les options possibles plus nombreuses. Nous l'avons déjà évoqué, une variable d'absence est composite : elle reflète d'autres décisions possibles, d'autres comportements. Les valeurs positives qu'elle comporte peuvent certes renvoyer à l'absence-maladie, mais également à l'absentéisme, c'est-à-dire à des absences non liées à des raisons de santé. Les valeurs nulles quant à elles, ne sont pas seulement l'expression d'une présence au travail permise par une bonne santé, mais également du présentéisme.

En d'autres termes, ces valeurs nulles peuvent être produites par deux processus majeurs, un processus normal et un processus structurel. D'une part, certains individus, dont l'état de santé ou la robustesse physique sont meilleures que la moyenne, ne subissent aucun épisode de maladie et sont présents tout au long de l'année. Ces salariés, s'ils avaient été malades, se seraient toutefois absentés. C'est le processus normal qui est ici décrit, celui qu'un modèle binomial négatif standard est tout à fait capable

de mesurer. Au vu de la distribution de la santé et des événements de maladie, on peut d'ailleurs supposer qu'un nombre non négligeable de valeurs nulles provient de ce processus normal (cf. chapitre précédent).

Une seconde source de valeurs nulles vient cependant s'ajouter à la première. Si certains individus, dans le processus normal, décident de prendre des jours d'absence pour se soigner (ou plutôt suivent la recommandation de leur médecin), d'autres se rendent à leur travail malgré un état de santé dégradé. Dans les données d'absence, ce comportement de présentéisme se traduit majoritairement par des valeurs nulles, qui sont toutefois d'une autre forme que les précédentes. Elles proviennent, non pas de la bonne santé de l'individu, mais de sa décision de travailler nonobstant la maladie. Ces zéros relèvent d'un processus « structurel ».

L'outil qui va être présenté dans les lignes à venir tient compte de ces deux situations.

1.2 Fonctionnement des modèles à inflation de zéros

C'est Lambert qui, en 1992, a proposé pour la première fois un modèle *zero-inflated*. Si ces modèles ont donc déjà 20 ans, ils ont été relativement peu exploités en comparaison des autres modèles de comptage. Les connaissances produites grâce à l'utilisation des modèles *zero-inflated* peuvent pourtant être importantes, comme on pourra en juger. L'objectif est à présent d'en considérer les mécanismes généraux. L'interprétation des éléments constitutifs de ces modèles sera élargie à l'absence et au présentéisme dans un second temps seulement.

1.2.1 Structure générale des modèles à inflation de zéros

Les modèles à inflation de zéros s'appuient largement sur les modèles de comptage standard explorés dans le chapitre précédent. Le modèle de Poisson et le modèle binomial négatif reposent sur l'estimation d'une espérance λ à partir d'une « équation de comptage » intégrant des variables explicatives. Dans les modèles à inflation de zéros une « équation d'inflation » estimée par un modèle logistique est adjointe à l'équation de comptage. Les modèles *zero-inflated* sont donc en fait des mélanges de modèles.

Qu'apporte l'introduction de cette équation d'inflation ? Elle va permettre d'évaluer la probabilité qu'un individu ayant une valeur nulle fasse partie de l'une ou de l'autre des deux catégories de zéros. En effet, on fait l'hypothèse que les valeurs nulles peuvent être

expliquée par deux processus (Winkelmann, 2008). Dans le premier processus, les zéros font partie d'un régime « normal » mesuré par un modèle de comptage classique (qui peut produire des valeurs nulles ou positives). Comme indiqué plus haut, les modèles standard, en particulier le modèle binomial négatif, permettent déjà de tenir compte d'un certain nombre de valeurs nulles. Il s'agit de valeurs nulles « circonstancielles » (Staub & Winkelmann, 2012). Dans le second processus, les valeurs produites sont toujours nulles et possèdent donc un statut différent, provenant d'un contexte spécifique (de nombreux exemples seront donnés dans la deuxième partie de ce chapitre). Il s'agit alors de zéros dits « structurels » ou « stratégiques » (Staub & Winkelmann, 2012) qui apparaissent en surplus des zéros habituellement observés dans la distribution.

La distribution des probabilités des modèles *zero-inflated* schématise ce double mécanisme :

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)f(0) & \text{si } y_i = 0 \\ (1 - p_i)f(y_i) & \text{si } y_i > 0 \end{cases} \quad (3.1)$$

La fonction $f(\cdot)$ désigne l'un des modèles de comptage standard présentés plus haut. La seule différence est que les valeurs nulles $f(0)$ et les valeurs positives $f(y_i > 0)$ sont distinguées. Mais l'apparente séparation des valeurs nulles et des valeurs positives dans l'équation 3.1 ne doit pas laisser penser que ces deux types de valeurs sont estimées séparément. Il existe des modèles fonctionnant sur ce principe, dits modèles « *hurdle* », qui estiment deux processus séparés, l'un produisant seulement des valeurs nulles, l'autre seulement des valeurs positives. Tel n'est pas le cas ici. En effet, dans le cas où $y_i = 0$ on s'aperçoit que deux processus entrent en ligne de compte, chacun étant pondéré par une probabilité p_i . Elle correspond à la probabilité d'être dans un régime structurel, aussi appelé « état parfait » (Lambert, 1992) où seules des valeurs nulles sont générées. La probabilité $1 - p_i$ recouvre alors les autres situations où la valeur de y_i peut être nulle ou positive, relevant du modèle de comptage standard. La probabilité p_i est calculée sur la base d'une régression logistique dont le vecteur de variables explicatives (noté Z_i ci-après) peut être identique, partiellement identique ou totalement différent du vecteur de variables explicatives (noté X_i) utilisé pour le modèle de comptage (Winkelmann, 2008). Ainsi :

$$p_i = \frac{\exp(Z_i^T \delta)}{1 + \exp(Z_i^T \delta)} \quad (3.2)$$

Statistiquement, la probabilité p_i informe sur la part de zéros (structurels) s'ajoutant aux zéros naturels du modèle de comptage (Staub & Winkelmann, 2012).

Si l'on note $E_f(Y_i)$ l'espérance du modèle de comptage standard utilisé, l'espérance du modèle *zero-inflated* est :

$$E(Y_i) = (1 - p_i)E_f(Y_i) \quad (3.3)$$

Une question légitime se pose également au moment d'implémenter un modèle *zero-inflated*. À partir de quel seuil doit-on considérer que le nombre de zéros est excessif? Autrement dit, comment savoir si un modèle à inflation de zéros se justifie? La première solution repose sur l'observation de la variable dépendante. Cependant, une forte fréquence de valeurs nulles ne signifie pas toujours qu'il existe une surabondance : une distribution de Poisson d'espérance faible et *a fortiori* une distribution binomiale négative à forte hétérogénéité génèrent naturellement une part importante de zéros. La seconde aide à la décision est fournie par le test de Vuong. Classiquement utilisé par les auteurs exploitant les modèles à inflations de zéros il teste l'hypothèse qu'un modèle *zero-inflated* est plus adapté aux données qu'un modèle standard en se basant sur les contributions individuelles aux log-vraisemblances respectives des modèles. Nous aurons l'occasion de l'exploiter par la suite (chapitre 4).

1.2.2 Modèle de Poisson à inflation de zéros (ZIP)

On peut à présent intégrer le modèle de Poisson dans la structure générale présentée plus haut pour obtenir la distribution des probabilités du modèle *zero-inflated Poisson* (noté ZIP). Ce modèle s'écrit de la façon suivante :

$$P(Y_i = y_i | p_i, \lambda_i) = \begin{cases} p_i + (1 - p_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} & \text{si } y_i = 0 \\ (1 - p_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} & \text{si } y_i > 0 \end{cases} \quad (3.4)$$

En sachant que y_i est toujours nul dans le cas « structurel », ceci se réécrit :

$$P(Y_i = y_i | p_i, \lambda_i) = \begin{cases} p_i + (1 - p_i) e^{-\lambda_i} & \text{si } y_i = 0 \\ (1 - p_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} & \text{si } y_i > 0 \end{cases} \quad (3.5)$$

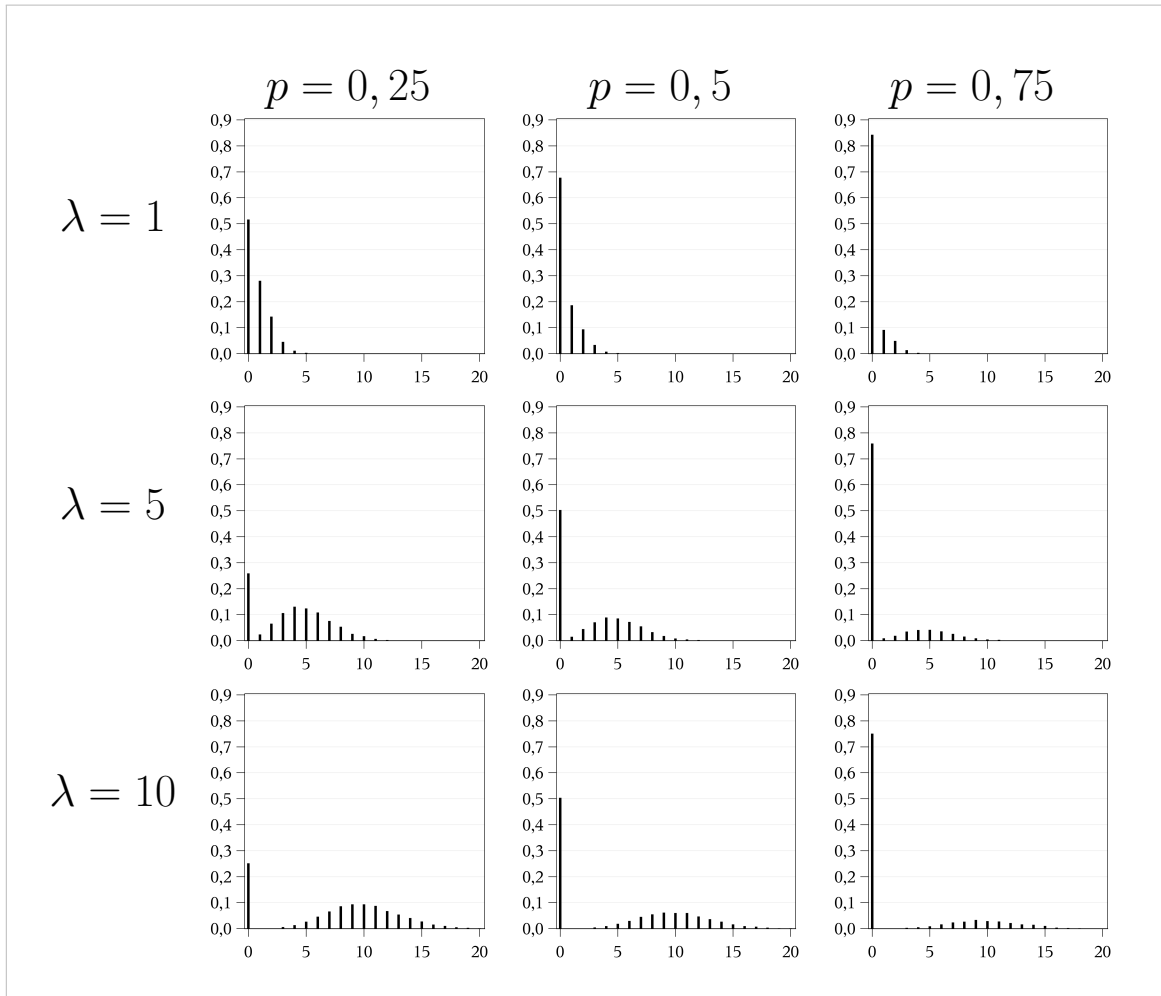
avec $p_i = \frac{\exp(Z_i^T \delta)}{1 + \exp(Z_i^T \delta)}$ et $\lambda_i = \exp(X_i^T \beta)$.

L'espérance et la variance du modèle ZIP sont respectivement ¹ :

$$E(Y_i|p_i, \lambda_i) = \lambda_i(1 - p_i) \quad (3.6)$$

$$V(Y_i|p_i, \lambda_i) = E(Y_i)(1 + p_i\lambda_i) = \lambda_i(1 - p_i)(1 + p_i\lambda_i) \quad (3.7)$$

Graphique 3.1 – Distributions ZIP selon λ et la probabilité d'inflation de zéro p



Le graphique 3.1 compare plusieurs distributions du modèle ZIP selon l'espérance λ du modèle de Poisson et la probabilité p d'inflation de zéros (la part de zéros dits « structurels »). La cohabitation d'un régime producteur de valeurs nulles et d'un

1. Le rapport entre l'espérance et la variance montre que la surdispersion dépend seulement de p_i .

régime de comptage est facilement observable. Pour $\lambda = 1$ et $p = 0,25$, environ la moitié des valeurs nulles proviennent de la loi de Poisson, tandis que l'autre moitié est générée par l'inflation de zéros. Puis, lorsque λ augmente (à partir de $\lambda = 5$), les deux processus sont de plus en plus distincts : pour des valeurs fortes de λ les valeurs nulles proviennent uniquement du processus d'inflation, puisque le modèle de Poisson n'en génère plus à ce niveau d'espérance.

La spécificité du modèle ZIP est de considérer que l'abondance de valeurs nulles est la source unique de surdispersion dans la variable dépendante. Il tient donc compte de l'existence d'un processus structurel générant uniquement des valeurs nulles, mais n'admet pas que des différences inter-individuelles puissent provenir d'une autre source. Autrement dit, il tient compte de la troisième forme d'hétérogénéité que nous avons décelée, sans tenir compte de la seconde.

1.2.3 Modèle binomial négatif à inflation de zéros (ZINB)

En plus de l'hétérogénéité observable (variables explicatives), les modèles binomial négatif et ZIP ne considèrent qu'un seul autre type d'hétérogénéité : l'hétérogénéité interindividuelle non-observée pour le premier et l'hétérogénéité comportementale pour le second. Il est donc nécessaire d'aboutir à un outil qui permettra de traiter ensemble les trois types d'hétérogénéité. En intégrant le modèle binomial négatif dans la structure *zero-inflated*, on obtient :

$$P(Y_i = y_i | p_i, \lambda_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{\theta}{\theta + \lambda_i} \right)^\theta & \text{si } y_i = 0 \\ (1 - p_i) \left[\frac{\Gamma(\theta + y_i)}{\Gamma(y_i + 1)\Gamma(\theta)} \left(\frac{\theta}{\theta + \lambda_i} \right)^\theta \left(\frac{\lambda_i}{\theta + \lambda_i} \right)^{y_i} \right] & \text{si } y_i > 0 \end{cases} \quad (3.8)$$

avec $p_i = \frac{\exp(Z_i^T \delta)}{1 + \exp(Z_i^T \delta)}$ et $\lambda_i = \exp(X_i^T \beta)$.

L'espérance et la variance du modèle ZINB sont respectivement² :

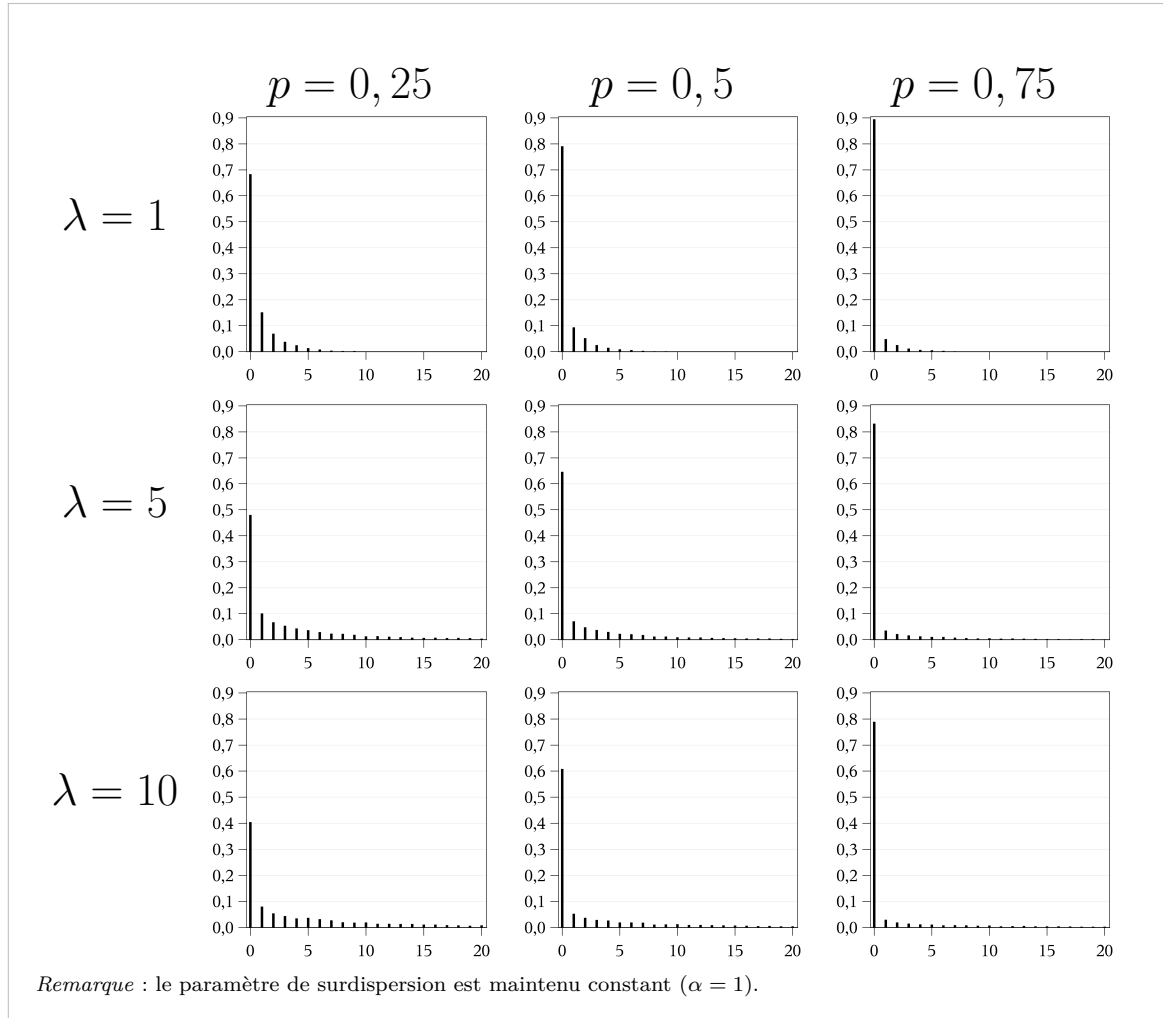
$$E(Y_i | p_i, \lambda_i) = \lambda_i(1 - p_i) \quad (3.9)$$

$$V(Y_i | p_i, \lambda_i) = \lambda_i(1 - p_i)(1 + p_i\lambda_i + \alpha\lambda_i) \quad (3.10)$$

2. Le rapport entre $E(Y_i)$ et $V(Y_i)$ indique que la surdispersion dépend à la fois de p_i et d' α .

Rappelons aussi que l'espérance λ_i du modèle de comptage, autrement dit l'espérance conditionnelle au fait d'être dans le régime normal, se déduit en posant $\lambda_i = \frac{E(Y_i)}{(1-p_i)}$.

Graphique 3.2 – Distributions ZINB selon λ et la probabilité d'inflation de zéro p



Comme on l'observe sur le graphique 3.2 (distribution du modèle selon l'espérance de l'équation de comptage et la part de zéros structurels), la proportion de valeurs nulles dans le modèle ZINB peut atteindre des valeurs extrêmement fortes (ici, 90% des observations pour $\lambda = 1$ et $p = 0,75$). Les zéros liés à l'inflation viennent alors s'ajouter à de nombreux zéros normalement présents dans le modèle, sans modifier l'espérance du modèle binomial négatif sous-jacent. Contrairement au modèle ZIP, le modèle ZINB peut en effet prendre en compte une quantité importante de valeurs nulles naturelles

issues du processus de comptage. Cette capacité, propre au modèle binomial négatif sous-jacent, provient de l'inclusion du paramètre d'hétérogénéité, comme nous l'avons vu précédemment. Dans la même distribution, on peut donc trouver conjointement des valeurs d'absence très élevées et de nombreuses valeurs nulles, ce qui produit une distribution très asymétrique.

1.2.4 *Modèle zero-inflated pour données longitudinales*

Jusqu'ici, nous avons seulement évoqué des modèles en coupe transversale (dits *cross-section*). Ceux-ci s'appliquent aux observations d'une année donnée et représentent la situation de façon statique. Bien que, dans cette thèse, l'analyse en coupe transversale constitue le cœur de l'analyse, les absences des individus peuvent naturellement être mesurées sur plusieurs périodes et justifier l'usage de modèles plus complexes.

Ouvrons ici une parenthèse sur les données d'absence longitudinales. Dans cette configuration, les données prennent la forme d'un panel dans lequel plusieurs observations décrivent un même individu de façon répétée. Deux éléments nouveaux apparaissent alors. D'abord, il existe une relation entre les données d'absence d'un même individu. Comme cela a très bien été décrit dans la littérature (Barmby, 2002 ; Barmby & Sibly, 2004 ; Winkelmann, 1999), l'absence d'un individu en t est largement déterminée par son absence ou sa présence en $t - 1$. Il existe une corrélation entre les observations d'un même individu, relayée par un état de santé spécifique au salarié. L'approche longitudinale a donc une conséquence : celle de fournir davantage d'information sur l'hétérogénéité inter-individuelle non observée. Détenir plusieurs valeurs d'absence pour un même individu permet de caractériser plus précisément son comportement et les facteurs qui l'expliquent, y compris les facteurs qui demeurent inobservés, comme sa santé. En parallèle à ces effets individuels, il peut y avoir des effets temporels, liés par exemple à des épidémies saisonnières plus prononcées certaines années. Mais il ne s'agit pas là de l'information que nous privilégions, puisque nous nous intéressons aux comportements des individus³. De plus ces effets temporels s'imposent de la même manière à l'ensemble des individus.

Une piste possible aurait été de cumuler les absences en élargissant la période d'analyse (Steel, 2003 ; Steel & Rentsch, 1995) et d'utiliser un modèle en coupe transversale.

3. D'ailleurs, parmi les modèles de comptage en panels qui vont être proposés, l'effet auquel on s'intéresse est, la plupart du temps, l'effet individuel (Boucher et al., 2009).

Malgré un certain gain dans la robustesse de l'analyse, cela induirait néanmoins une perte d'informations sur l'évolution des variables, sur les spécificités annuelles ainsi que sur l'hétérogénéité non-observée. La stratégie choisie est la mise en œuvre d'un modèle *zero-inflated* en panels. Mais l'élaboration de ce type de modèles est encore un chantier largement ouvert (Winkelmann, 2008) impliquant des défis très particuliers (Min & Agresti, 2005). Néanmoins, différentes variantes du modèle ZIP incluant des effets aléatoires⁴ ont été proposés dans la littérature (Boucher et al., 2009 ; Hur et al., 2002 ; Yau et al., 2004). Ces effets aléatoires sont en particulier destinés à tenir compte de la corrélation entre les observations d'un même individu. La vraie différence avec un modèle ZINB en coupe transversale est qu'un modèle en panels tient compte de l'aspect répétitif des données et ne traite pas les observations d'un même individu comme celles de deux individus différents. Certains auteurs estiment néanmoins qu'un modèle en coupe transversale basé sur la loi binomiale négative est capable d'offrir des performances raisonnables sur des données en panels (Allison, 2009). En effet, le modèle binomial négatif standard n'est rien d'autre qu'un modèle de Poisson incluant un effet individuel aléatoire suivant une distribution Gamma (Winkelmann, 2008).

Dans les modèles ZIP en panels proposés (Hur et al., 2002 ; Yau et al., 2004), l'effet aléatoire introduit suit une loi normale dont la variance est estimée au cours de la régression. Nous avons montré plus haut que, dans notre cas, une loi Gamma convenait sans doute mieux à la mesure de l'hétérogénéité. D'après Boucher et al. (2009), une transformation est possible pour que l'effet aléatoire suive une loi Gamma. Nelson et al. (2006) proposent un moyen de transformer ce terme aléatoire afin qu'il suive une loi Gamma⁵. Dans la section consacrée aux simulations de Monte-Carlo, nous aurons l'occasion d'appliquer trois modèles sur des données longitudinales simulées : le modèle ZINB standard et deux modèles ZIP en panels dont les effets aléatoires suivent respectivement une distribution normale et une distribution Gamma.

4. Rappelons, qu'en économétrie des panels, les effets individuels peuvent être traités soit comme des effets fixes, soit comme des effets aléatoires. Dans le premier cas, les effets individuels sont appréhendés comme des constantes, c'est-à-dire des paramètres à estimer (ou du moins à isoler). Une manière de spécifier un modèle à effets fixes serait de spécifier une dummy pour chaque individu. Mais pour une population de taille conséquente, les limites des logiciels statistiques sont vite atteintes (Winkelmann, 2008). D'autres méthodes permettraient l'estimation. Néanmoins, à notre connaissance, aucun modèle *zero-inflated* à effets fixes n'a été proposé. Dans le second cas, les effets individuels sont considérés comme un terme aléatoire suivant une distribution donnée pour l'ensemble de la population. On parle parfois de modèles à erreur composée, une partie des résidus étant liée aux effets individuels.

5. Il proposent une méthode avec le logiciel SAS.

1.3 Interprétation pour les données d'absence

1.3.1 Éléments principaux

Nous venons de détailler les aspects techniques des modèles *zero-inflated*. À présent, il serait souhaitable de revenir au processus qui nous a conduit à les exploiter sur des données d'absence et à leur interprétation dans ce contexte. La variable que nous explorons, à savoir le nombre annuel de jours d'absence par salarié a été abordée comme une variable composite dans laquelle on peut déceler d'autres phénomènes que le binôme absence pour cause de maladie / présence en bonne santé. Le présentéisme s'y trouve emboîté et se manifeste spécialement par une surabondance de valeurs nulles. Si l'on considère que tous les zéros caractérisent une présence en bonne santé, nul n'est besoin de chercher un modèle plus sophistiqué que le binomial négatif. Mais si l'on soupçonne qu'il existe deux sources de valeurs nulles, l'utilisation d'un modèle à inflation de zéros s'impose.

Sur la base des mêmes arguments qui nous ont poussé à privilégier le modèle binomial négatif standard au modèle de Poisson standard, nous privilégions le modèle ZINB plutôt que le ZIP. Rappelons les deux arguments majeurs qui nous conduisent à ce choix. Premièrement, le modèle ZINB permet de prendre en compte la surdispersion des données. Celle-ci n'est pas seulement liée à l'inflation de zéros comme le postule le modèle ZIP : il y a d'autres sources d'hétérogénéité interindividuelle, qui peuvent être captées dans un modèle ZINB. Deuxièmement, ce modèle reste pertinent même lorsque les occurrences ne sont pas parfaitement indépendantes, ce qui est très certainement le cas pour les données d'absence. Dès lors, à quels types de comportements renvoient les différents éléments constitutifs du modèle ZINB ?

Repartons de la structure générale du modèle (f étant le modèle binomial négatif) :

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)f(0) & \text{si } y_i = 0 \\ (1 - p_i)f(y_i) & \text{si } y_i > 0 \end{cases} \quad (3.11)$$

Aux trois éléments constitutifs du modèle correspondent trois comportements distincts.

Le présentéisme. Il est mesuré par la probabilité p_i . Cette probabilité renvoie en effet au régime générant toujours des valeurs nulles que les utilisateurs des modèles

zero-inflated appellent régime structurel (Staub & Winkelmann, 2012) ou état parfait (Lambert, 1992). Loeys et al. (2012) insistent particulièrement sur le fait que cette partie du modèle est bien souvent négligée et/ou mal interprétée dans la littérature. L'idée est en effet essentielle. Il s'agit bien ici d'une probabilité d'appartenir à un groupe dont les individus présentent des valeurs nulles « inattendues » provoquées par une règle qu'ils se sont imposée. Il ne s'agit en aucun cas d'une probabilité de ne pas être absent. D'ailleurs, en contraste, on s'aperçoit que les zéros naturels ou circonstanciels (Staub & Winkelmann, 2012) n'entrent pas dans cette probabilité. Ils sont pris en compte avec les valeurs positives dans le régime normal de probabilité $1 - p_i$.

La probabilité p_i est donc bien une probabilité de présentéisme et non une probabilité d'être présent (ou de zéro occurrence). Le présentéisme se définit alors comme un comportement générant un surplus de valeurs nulles ne faisant pas partie du modèle de comptage standard. Il est le fait d'individus touchés par une maladie dont la règle est de se rendre au travail, alors que le comportement normal serait de prendre un arrêt de travail.

La présence en bonne santé. Elle est mesurée par $f(0)$ et apparaît avec une probabilité $(1 - p_i)f(0)$. Même si la valeur qui la représente est identique à celle du présentéisme (zéro), ces deux comportements ne relèvent pas du même processus. Ce qui est capté ici est l'ensemble des zéros dits circonstanciels qui entrent dans le processus de comptage binomial négatif. En d'autres termes, il peut arriver qu'un individu ne soit pas malade pendant une année donnée. S'il avait eu un incident de santé justifiant un arrêt de travail, il se serait tabsenté.

L'absence-maladie. Elle est mesurée par $f(y_i)$ et entre dans le même régime que la présence en bonne santé avec une probabilité $(1 - p_i)f(y_i)$. Ici, on revient à la mesure traditionnelle de la gravité d'absence.

1.3.2 Autres contextes d'exploitation

Certains articles appliquent également le modèle ZINB (ou un modèle proche) à des données d'absence, mais négligent l'interprétation de l'équation d'inflation (Lu et al., 2010; Taimela et al., 2007). À notre connaissance, seuls quelques travaux exploitant un modèle *zero-inflated* ont effleuré la question du présentéisme, sans le nommer. Un

article produit dans le cadre de la DREES⁶, indique que la surabondance de zéros dans les données d'absence peut provenir de deux sources. D'un côté, « les individus, face à un problème de santé, décident de ne pas s'arrêter alors qu'ils pourraient le demander » et, de l'autre, ils « ne font face à aucun événement susceptible d'entraîner un arrêt de travail » (Missègue, 2007, p.11). Ce constat est partagé par Frick et Malo (2008, p. 518) dont l'hypothèse est qu'il existe « une règle comportementale de ne pas être absent volontairement »⁷ pour les individus s'inscrivant dans le régime structurel.

Cette interprétation des modèles *zero-inflated* lorsqu'ils sont appliqués aux absences-maladie peut être complétée par la description qu'en font les auteurs issus d'autres champs ou utilisant d'autres types de données. L'article de Staub et Winkelmann (2012, p. 1) propose par exemple une formulation qui, par analogie, peut aider à la compréhension de notre démarche : « considérons le recours aux soins mesuré par le nombre de visites chez le médecin. Une personne peut ne pas se rendre chez son médecin pendant une période donnée, premièrement parce qu'il est adepte des médecines alternatives et ne consulte jamais un médecin, deuxièmement parce qu'il n'a pas eu l'occasion de le faire pendant la période d'observation », c'est-à-dire qu'il n'a pas été malade. Les deux régimes de production de valeurs nulles sont bien distingués ici. D'autres formulations de ce type peuvent être proposées.

Rebuts. Dans l'article séminal présentant pour la première fois un modèle de Poisson à inflation de zéros, Lambert (1992) étudie le cas d'une entreprise produisant des cartes électroniques. Dans le processus de production, des machines soudent les composants sur des circuits imprimés et l'entreprise dispose de plusieurs critères permettant d'évaluer la qualité de la soudure. Dans un très grand nombre de cas, le produit est acceptable. À cause de l'abondance de ces valeurs nulles, un modèle de Poisson standard s'avère incapable de modéliser le nombre de rebuts résultant d'un défaut de soudure. En recherchant une solution, l'auteur s'est aperçu que les valeurs nulles proviennent en fait de deux causes. L'absence de défaut peut signifier soit qu'une machine est parfaitement calibrée, soit qu'une machine mal paramétrée parvient tout de même à confectionner un produit acceptable. De là est née l'idée de créer un modèle qui parviendrait conjoin-

6. Direction de la Recherche, des Études, de l'Évaluation et des Statistiques.

7. « *They follow an absolute rule of no voluntary absenteeism* ». Signalons au passage que nous sommes entrés en contact avec les auteurs de ce papier. Ils ont confirmé la pertinence de notre approche.

tement à rendre compte du processus de comptage des rebuts et de la probabilité pour une machine d'être dans un état parfait d'absence de rebuts.

Dépôts de brevets. Carayol (2007) utilise un modèle ZINB pour décrire le processus de production de brevets dans le contexte d'une université française. Lorsque sur une période donnée un chercheur n'a pas déposé de brevet, il peut y avoir deux explications. D'une part, le chercheur peut avoir mené des recherches qui, tout en étant brevetables dans l'absolu, n'ont pas débouché pendant la période d'observation. D'autre part, il est également possible que le chercheur mène des recherches qui ne peuvent en aucun cas donner lieu à un dépôt de brevet. Ce second type de valeurs nulles entre dans le régime structurel tel qu'il a été décrit plus haut.

Sécurité routière. Allain et Brennac (2001) étudient le nombre d'accidents sur différents sites routiers. Leurs deux groupes de valeurs nulles sont décrits comme suit : « il existe deux états, deux groupes d'appartenance des sites étudiés. Certains sites relèvent de l'état à zéro accident (*zero-accident state*) : le site est fondamentalement sûr, à risque très faible, c'est-à-dire qu'aucun accident ne peut être observé. D'autres sites relèvent de l'état accidentel (*accident state*) : sur ces sites, les observations suivent une distribution connue (loi de Poisson ou loi binomiale négative) et on retrouve parmi eux des sites à moyenne faible où aucun accident n'est observé sur la période d'étude ».

1.3.3 Deux objections

En restant dans la perspective de l'interprétation du modèle ZINB, on pourrait légitimement soulever deux objections à notre analyse. Premièrement, peut-on être entièrement sûr qu'il n'existe pas d'autres processus générateurs de valeurs nulles ? Deuxièmement, le présentéisme ne peut-il pas se traduire également par de faibles valeurs positives ? Cette deuxième question est particulièrement importante et sera spécifiquement examinée dans les dernières sections de ce chapitre à l'occasion des simulations de Monte-Carlo. Il s'agira alors de tester la stabilité du modèle lorsque les hypothèses de départ sont relâchées. Mais il est possible de préciser ici les éléments permettant de penser que ces deux objections ne remettent pas en cause la validité de la mesure du présentéisme. D'abord, peut-il y avoir d'autres sources de valeurs nulles ? Dans cette thèse, nous supposons que le phénomène d'inflation de zéro est très majoritairement dû au com-

portement de présentéisme. Plusieurs arguments plaident en faveur de cette affirmation. Comme indiqué plus haut, seules deux raisons expliquent convenablement la grande quantité de valeurs nulles : un état de santé excellent pour une grande partie de la population ou une règle comportementale de présence invétérée. Le modèle ZINB est précisément conçu de manière à capter des niveaux de santé hétérogènes par le modèle de comptage qu'il intègre. Les salariés présents parce qu'ils sont en bonne santé entrent dans ce processus. Par conséquent, le présentéisme demeure la seule interprétation raisonnable des valeurs nulles produites dans le régime d'inflation. Nos résultats, sur données simulées (section 2) aussi bien que sur données réelles (chapitres suivants) apporteront d'autres confirmations. On pourrait malgré tout imaginer que des valeurs nulles puissent provenir de dysfonctionnements dans le système d'enregistrement des absences. Certaines catégories de personnel ayant des contrats de travail ou des fonctions spécifiques, pourraient aussi négliger de déclarer certaines absences. Ces possibilités ont été envisagées. Mais étant donné la détermination des entreprises avec lesquelles nous avons travaillé pour réduire les absences à tous les niveaux, nous estimons que ces scénarios restent tout au plus marginaux et ne peuvent expliquer une surabondance de valeurs nulles. Par ailleurs, les résultats des régressions viennent également invalider cette hypothèse.

D'autre part, le présentéisme ne pourrait-il se manifester par de faibles valeurs positives ? Empiriquement, on pourrait admettre qu'au cours d'une année, un individu habituellement présentéiste puisse prendre quelques jours d'absence en raison d'une maladie plus grave. Il peut aussi arriver que sa propension au présentéisme le pousse à réduire de quelques jours la période de convalescence adaptée à son affection. Néanmoins, l'hypothèse la plus vraisemblable reste qu'un salarié présentant un profil authentique de présentéiste n'enregistre aucune absence sur la période (valeur nulle). À l'extrême, on pourrait donc considérer qu'il existe deux types de présentéistes : la grande majorité qui n'a pas été absente et un groupe d'individu ayant des valeurs faibles d'absence. De là découle l'analyse suivante. D'abord, il n'y a aucune raison de penser que les facteurs de présentéisme diffèrent dans ces deux « catégories ». Leur profil statistique, que nous cherchons à déterminer, est sans doute similaire ou fortement comparable. Pour déterminer statistiquement un profil de présentéisme fiable, on peut donc se contenter d'utiliser les valeurs nulles. Ce profil statistique est matérialisé par les coefficients des différentes variables explicatives estimés par le modèle. À partir de ces coefficients, on calcule une probabilité de présentéisme pour tous les individus de la population :

les présentéistes obtiendront alors une probabilité élevée même s'ils ont une valeur d'absence non-nulle, puisque leur profil est similaire à celui des présentéistes ayant une valeur nulle. La simulation de Monte-Carlo devra valider cette assertion et, dans le chapitre 4 (application empirique), il sera intéressant de confronter les valeurs d'absences réelles aux probabilités de présentéisme estimées des individus.

2 Simulations de Monte-Carlo

Jusqu'ici nous avons décrit le fonctionnement et les hypothèses du modèle ZINB, qui sera majoritairement utilisé par la suite. L'application à des données réelles et l'analyse qui s'en suivra fourniront des éléments complémentaires sur la pertinence de notre démarche. Mais c'est d'abord à des données simulées que le modèle va être confronté, dans le cadre de simulations de Monte-Carlo. Deux objectifs sont ici poursuivis : premièrement favoriser la compréhension du modèle et deuxièmement vérifier sa robustesse lorsque les hypothèses initiales ne sont pas parfaitement respectées. Pour ce faire, nous procéderons en deux temps. En premier lieu, les données simulées respecteront parfaitement les présupposés du modèle. L'idée est simplement de savoir si la technique d'estimation est bien capable de faire ce que l'on attend d'elle, par exemple mesurer l'hétérogénéité non-observée (section 2.2). En second lieu, des modifications seront introduites selon plusieurs scénarios pour observer la stabilité des estimations (section 2.3). Ces changements seront soit de nature technique (corrélation de variables, distribution du paramètre d'hétérogénéité, etc.), soit de nature plus empirique (pas de présentéisme pour les salariés les plus malades, présentéisme générant également des valeurs positives). Enfin, nous évoquerons une simulation spécifique réalisée sur des données en panels (section 2.4). Mais avant d'entrer dans la simulation à proprement parler, nous décrivons le processus de production des variables explicatives et de la variable expliquée.

2.1 Préliminaires

2.1.1 *Variables explicatives*

La première étape est la création d'une base de données comprenant 10 000 individus. Ceux-ci sont décrits par trois variables explicatives reflétant plus ou moins les différents

types de variables qui se retrouvent classiquement dans les micro-données d'entreprise. Ces trois variables demeurent fixes pour l'ensemble des simulations :

- la variable HOMME est une variable binaire (1 si l'individu est un homme, 0 sinon) : 60% des individus sont des hommes ;
- la variable AGE est une variable quantitative uniformément distribuée et comprise entre 20 et 65 ;
- la variable ENFANTS est une variable quantitative suivant une distribution de Poisson ($\lambda = 1,75$) : 17,2% des individus n'ont pas d'enfants, 30,3% en ont un seul, 3,4% en ont plus de 3.

Si nécessaire, d'autres variables explicatives suivant des hypothèses spécifiques sont ajoutées à ce lot initial.

2.1.2 Variable expliquée

À partir de ces trois variables explicatives sont respectivement générées l'espérance d'absence et la probabilité de présentéisme « théoriques » de chaque individu. Les coefficients utilisés sont inspirés de ceux réellement observés dans la suite de ce travail. L'espérance λ est définie par l'équation suivante :

$$\lambda_i = \exp(\mu_i) \quad (3.12)$$

avec

$$\begin{aligned} \mu_i = & - 1,3 \times (1 + \epsilon_1) - 0,6 \times (1 + \epsilon_2) \times \text{HOMME} \\ & + 0,04 \times (1 + \epsilon_3) \times \text{AGE} + 0,045 \times (1 + \epsilon_4) \times \text{ENFANTS} \end{aligned} \quad (3.13)$$

La probabilité de présentéisme p_i est quant à elle calculée comme suit :

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (3.14)$$

avec

$$\begin{aligned} \eta_i = & - 1,5 \times (1 + \epsilon_5) + 0,3 \times (1 + \epsilon_6) \times \text{HOMME} \\ & + 0,04 \times (1 + \epsilon_7) \times \text{AGE} - 0,04 \times (1 + \epsilon_8) \times \text{ENFANTS} \end{aligned} \quad (3.15)$$

Les coefficients intervenant dans ces équations sont soumis à un aléa uniforme, afin que leurs valeurs ne soient pas parfaitement identiques dans toutes les simulations. Ils peuvent varier de plus ou moins 5% ($-0,05 \leq \epsilon_n \leq 0,05$).

Ensuite, les individus sont répartis en deux catégories : les présentéistes et les autres. Cette catégorisation se fait par tirage aléatoire, suivant la probabilité de présentéisme théorique de l'individu⁸.

Enfin, on génère la nombre de jours d'absence pour chaque individu. Pour les présentéistes, celui-ci est nul (cette hypothèse sera assouplie dans une simulation). Pour les autres, la valeur d'absence suit une loi de Poisson en fonction de leur espérance individuelle λ_i . La distribution des absences simulées possède toutes les caractéristiques classiques des données d'absences. À partir de la deuxième simulation, les valeurs d'absence ne seront plus purement poissonniennes, puisqu'un paramètre d'hétérogénéité sera introduit.

En résumé, la génération des données d'absence se déroule comme suit :

```

    espérance d'absence  $\lambda_i$ 
    probabilité de présentéisme  $p_i$ 

    variable aléatoire uniforme  $x$  comprise entre 0 et 1
    si  $p_i > x$  alors absence = 0
    si  $p_i < x$  alors absence  $\sim$  Poisson( $\lambda_i$ )
  
```

2.1.3 Stratégie globale

Dans les pages à venir, 11 simulations correspondant chacune à un scénario bien défini sont proposées. Pour chaque simulation 100 tirages sont effectués. Chaque tirage consiste à régénérer une variable dépendante à partir des variables explicatives (dont la valeur reste fixe sur l'ensemble des simulations) et à effectuer quatre régressions : une régression de Poisson, une binomiale négative, une ZIP et une ZINB. Pour l'ensemble des modèles, on dispose donc des coefficients estimés, de l'espérance d'absence estimée et, dans le cas des modèles *zero-inflated*, de la probabilité de présentéisme estimée pour

8. Dans la pratique, on compare la probabilité p_i à une valeur aléatoire uniforme (i. e. comprise entre 0 et 1) : si la probabilité est supérieure à la valeur aléatoire, l'individu est considéré comme présentéiste.

chaque individu. Pour chaque simulation, après avoir décrit le scénario, nous présentons les moyennes des coefficients et des probabilités de présentéisme sur 100 tirages.

2.2 Modèles basiques

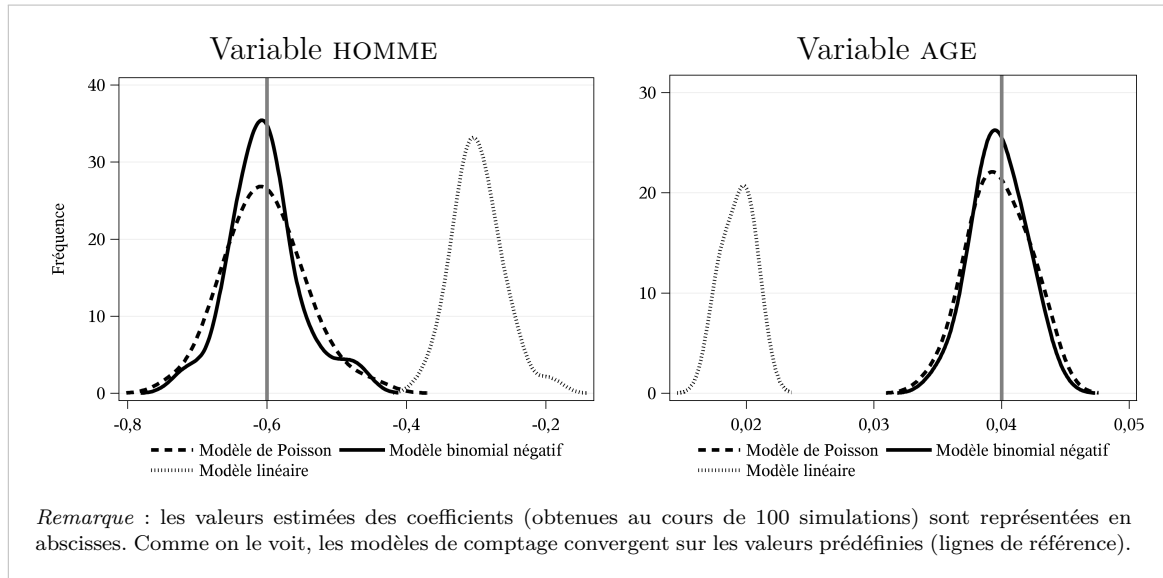
Les trois premières simulations peuvent paraître triviales. Elles montrent simplement que les modèles fonctionnent parfaitement lorsque leurs hypothèses sont respectées. L'intérêt de ces simulations réside dans la confrontation des modèles. Deux assertions énoncées précédemment doivent être vérifiées. D'une part, les modèles purement poissonniens sont-ils effectivement moins performants en présence d'une surdispersion ? D'autre part, les modèles *zero-inflated* produisent-ils des estimations cohérentes malgré la présence « contrainte » d'un grand nombre de valeurs nulles ? Mais avant d'entamer les simulations spécifiquement dédiées aux modèles à inflation de zéros, nous ouvrons une parenthèse concernant les modèles standards.

2.2.1 Modèles linéaires

Dans le chapitre précédent, nous évoquions l'opposition de la littérature à l'usage d'une régression linéaire estimée par les moindres carrés ordinaires pour des données de comptage. Nous avons voulu vérifier cet avis général et avons procédé à quelques essais sur des données purement poissonniennes, sans inflation de zéros. Le graphique 3.3 (page 111) présente les estimations des coefficients HOMME et AGE sur 100 tirages. Comme on le voit, les modèles de comptage convergent sur la valeur prévue (représentée par une ligne de référence), tandis que les estimations du modèle linéaire (ou plus justement log-linéaire après transformation logarithmique) sont biaisées.

Par ailleurs, d'autres simulations montrent que lorsque l'espérance λ est élevée (en moyenne sur 100 tirages, $\lambda = 12, 15$), les valeurs des coefficients estimées par les MCO sont consistantes et comparables à celle de la régression de Poisson. La forme de la variable dépendante est alors quasiment normale. Mais les résultats se dégradent dans deux situations : lorsque λ diminue (e.g. $\lambda = 5, 37$) ou lorsqu'on introduit de l'hétérogénéité ($\lambda = 12, 06$ mais $\alpha = 5$). Dans les deux cas, la distribution devient fortement asymétrique. Dans le dernier cas, le plus réaliste lorsqu'il s'agit de données d'absence, seul le modèle binomial négatif demeure parfaitement fiable.

Graphique 3.3 – Estimation de deux variables : modèles sans inflation de zéro

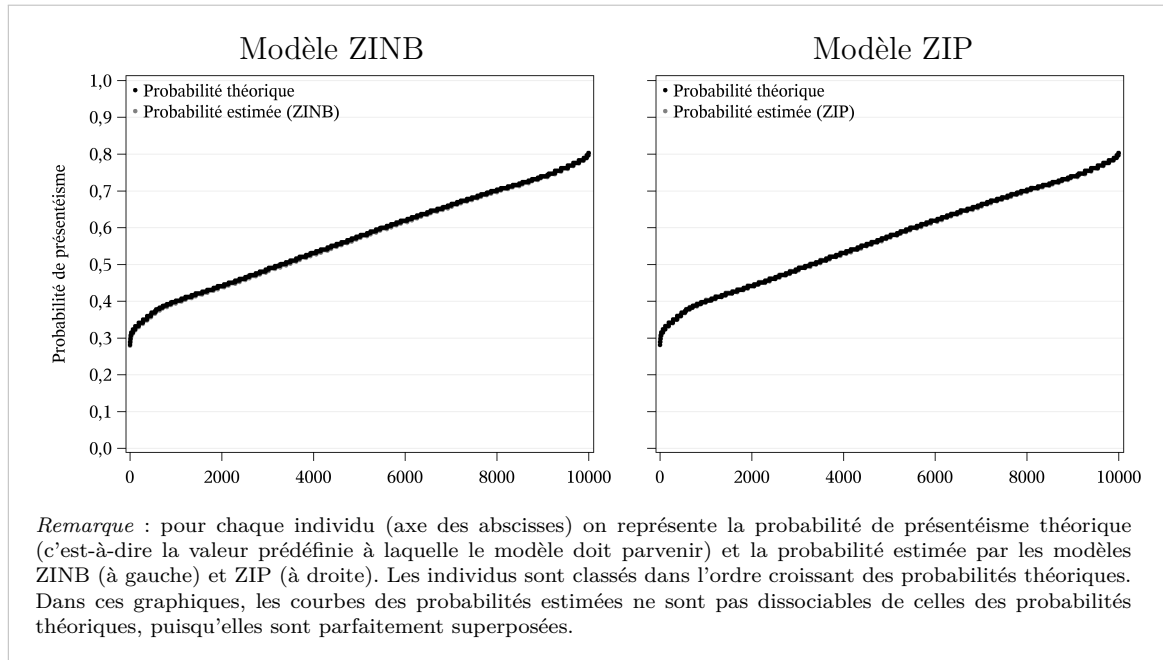


2.2.2 Simulation n° 1 : modèle sans hétérogénéité

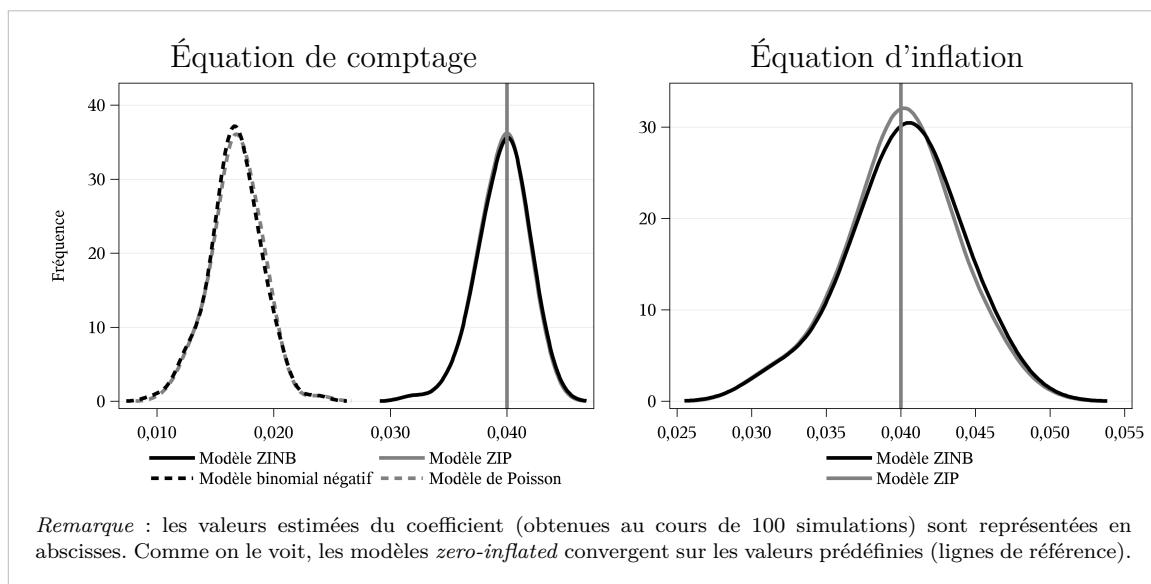
Le cas le plus simple est celui où il n'existe aucune hétérogénéité inobservée. Les différences entre les valeurs d'absences sont alors parfaitement expliquées par les trois variables indépendantes. D'après les éléments théoriques évoqués plus haut, les modèles *zero-inflated* devraient parvenir à une meilleure estimation que les modèles standards à cause de l'inflation de zéros, mais il ne devrait pas y avoir de différences majeures entre le modèle ZIP et le modèle ZINB. En effet, lorsqu'il n'y a pas d'hétérogénéité inobservée ($\alpha \rightarrow 0$), estimer un modèle binomial négatif revient à estimer un modèle de Poisson. Ces hypothèses sont entièrement vérifiées dans les résultats.

Le graphique 3.4 (page 112) montre clairement que pour le modèle ZINB, aussi bien que pour le ZIP, les probabilités de présentisme sont parfaitement estimées : le nuage de points des probabilités estimées se confond avec celui des probabilités théoriques. Rappelons que les probabilités de présentisme sont calculées sur la base des coefficients de l'équation d'inflation, qui sont donc également parfaitement estimés. Concernant les coefficients de l'équation de comptage, l'exemple de la variable AGE est proposé dans le graphique 3.5 (page 112) : on s'aperçoit que les modèles standards sont déstabilisés par la surreprésentation des valeurs nulles, tandis que les modèles *zero-inflated* convergent

Graphique 3.4 – Probabilités de présentéisme théoriques et estimées (simulation n° 1)



Graphique 3.5 – Estimation du coefficient de la variable AGE (simulation n° 1)



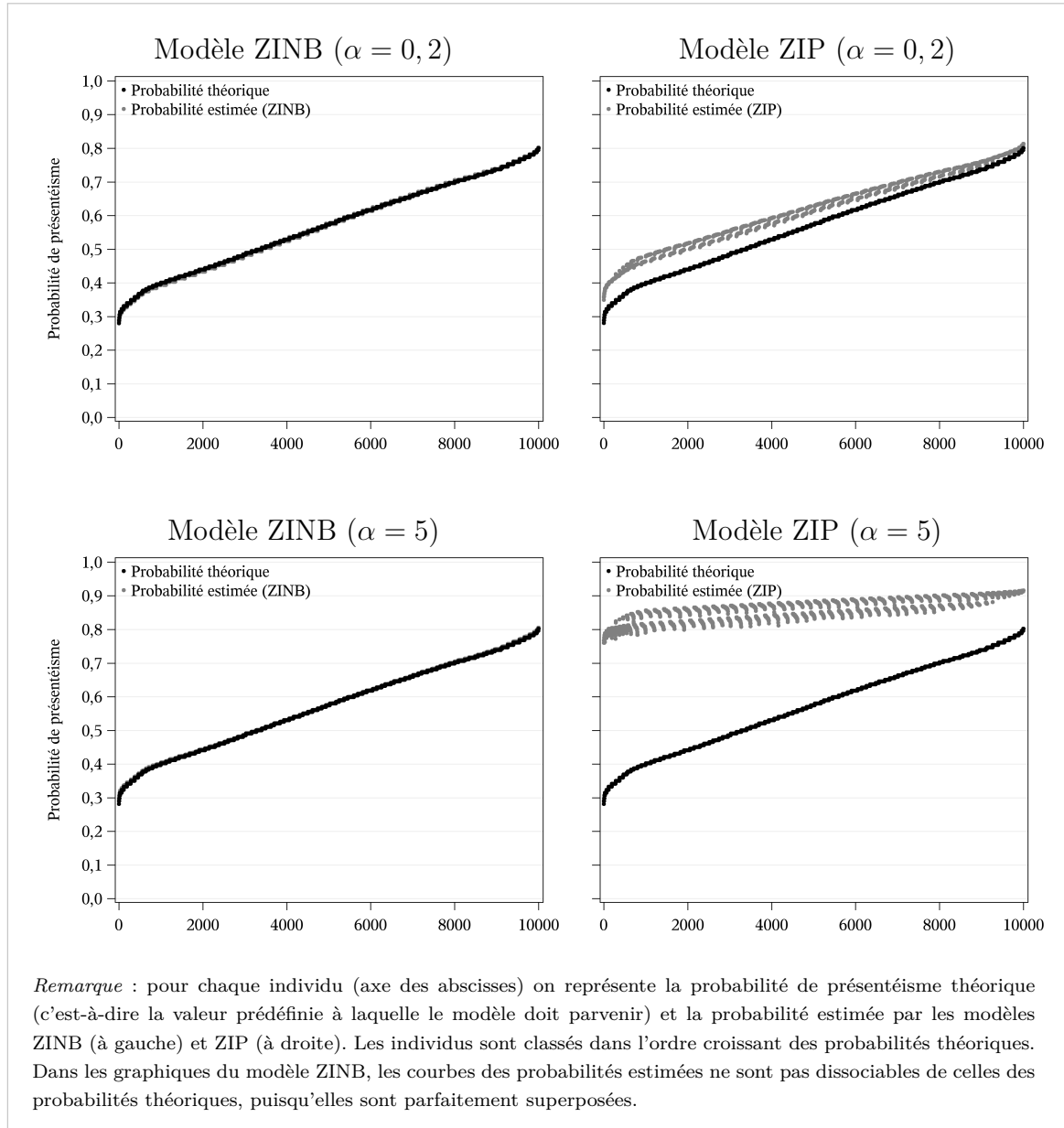
sur la bonne valeur. Le modèle binomial négatif standard cherche à isoler la surdispersion que produisent les nombreux zéros (son paramètre d'hétérogénéité est de 2,32), mais n'aboutit pas à une estimation correcte.

2.2.3 Simulation n° 2 & n° 3 : modèles avec hétérogénéité

Au vu des résultats de la première simulation, le modèle ZIP est parfaitement suffisant en l'absence d'hétérogénéité inobservée. Il convient parfaitement lorsque les variables explicatives essentielles sont à disposition et qu'aucun effet externe ne vient perturber les valeurs d'absence. Pour les données d'absence, ces conditions ne sont habituellement pas réunies et des variables non observées peuvent avoir une influence sur les valeurs. L'état de santé individuel a particulièrement été mis en évidence comme un facteur important demeurant indisponible pour la modélisation. Il est nécessaire d'en tenir compte à présent en intégrant un paramètre d'hétérogénéité dans la simulation, puis d'évaluer l'efficacité des régressions respectives. Par simplification, nous admettrons que les effets inobservés se résument en une variable SANTÉ mesurant l'état de santé général et que celle-ci suit une loi Gamma. Plus bas, nous aurons l'occasion de tester d'autres scénarios. Pour le moment, nous considérons qu'une loi Gamma est suffisamment souple pour s'adapter à diverses distribution de la santé.

Comme nous l'avons expliqué plus haut, les deux paramètres de la loi Gamma sont présumés égaux pour que l'espérance du paramètre d'hétérogénéité soit égal à 1. On a donc $\text{SANTÉ} \sim \Gamma(\theta, \theta)$, c'est-à-dire $\Gamma(\alpha^{-1}, \alpha^{-1})$, puisque les modèles estiment le paramètre α . Un paramètre α faible caractérise une situation dans laquelle les individus ont un état de santé comparable (homogénéité), tandis qu'un paramètre α élevé caractérise une situation de forte hétérogénéité inter-individuelle (diverses distributions de la loi Gamma en fonction de ce paramètre sont présentées dans le graphique 2.9 page 83). Pour illustrer ces deux scénarios, le paramètre SANTÉ est introduit dans la simulation n° 2 avec $\alpha = 0,2$ et dans la simulation n° 3 avec $\alpha = 5$ multiplicativement de λ_i ($\lambda_i = \exp(\mu_i) \times \text{SANTÉ}$). La moyenne de la variable aléatoire est égale à 1. De ce fait, lorsque la variable prend une valeur faible, l'individu est en bonne santé et son absence est également plus faible (multipliée par une valeur inférieure à 1). Inversement lorsque la valeur est élevée, l'état de santé est dégradé et l'absence plus forte (multipliée par une valeur supérieure à 1).

Graphique 3.6 – Probabilités de présentéisme théoriques et estimées (sim. n° 2 & n° 3)



Comme on s'en rend compte sur le graphique 3.6, même pour un α faible (0,2), le modèle ZIP présente déjà des écarts d'estimation assez nets. Ensuite, plus le paramètre α est élevé, moins le modèle ZIP est performant. Dans un régime d'hétérogénéité forte, le modèle ZIP surestime considérablement les probabilités : en moyenne, le coefficient

multiplicateur entre les probabilités théoriques et les probabilités estimées est de 1,59. Sans aucune information sur la santé des individus, le modèle ZINB continue à estimer parfaitement les probabilités et donne une mesure fiable de la dispersion. Concernant l'équation de comptage (tableau 3.1), en situation de faible hétérogénéité, les coefficients du modèle ZIP, tout en restant dans l'intervalle de confiance, sont légèrement sous-estimés, tandis que dans le modèle ZINB, la qualité d'estimation reste inchangée.

Tableau 3.1 – Coefficients de l'équation de comptage (sim. n° 2 et n° 3)

	Base	Simulation n° 2				Simulation n° 3			
		ZIP		ZINB		ZIP		ZINB	
		$(\bar{x})$	(σ)	$(\bar{x})$	(σ)	$(\bar{x})$	(σ)	$(\bar{x})$	(σ)
CONSTANTE	-1,300	-1,121	0,101	-1,308	0,109	0,217	0,151	-1,289	0,246
HOMME	-0,600	-0,572	0,049	-0,609	0,052	-0,437	0,082	-0,606	0,107
AGE	0,040	0,038	0,002	0,040	0,002	0,029	0,003	0,040	0,004
ENFANTS	0,045	0,042	0,013	0,045	0,014	0,033	0,030	0,049	0,035
α (sim. n° 2)	0,200			0,198	0,037				
α (sim. n° 3)	5,000							4,940	1,126

$(\bar{x})$ moyenne des coefficients sur 100 tirages

(σ) écart-type des coefficients sur 100 tirages

2.3 Relâchement des hypothèses préliminaires

Jusqu'ici, les hypothèses posées par les modèles *zero-inflated* étaient entièrement respectées. L'une des forces des simulations de Monte-Carlo est d'évaluer le biais engendré par l'introduction de perturbations dans le modèle. Par exemple, le choix d'une distribution Gamma est, certes, une hypothèse raisonnable, mais demeure un présupposé : il est utile d'évaluer le risque d'erreur au cas où ce choix s'avérerait erroné. C'est l'une des situations testées dans cette partie.

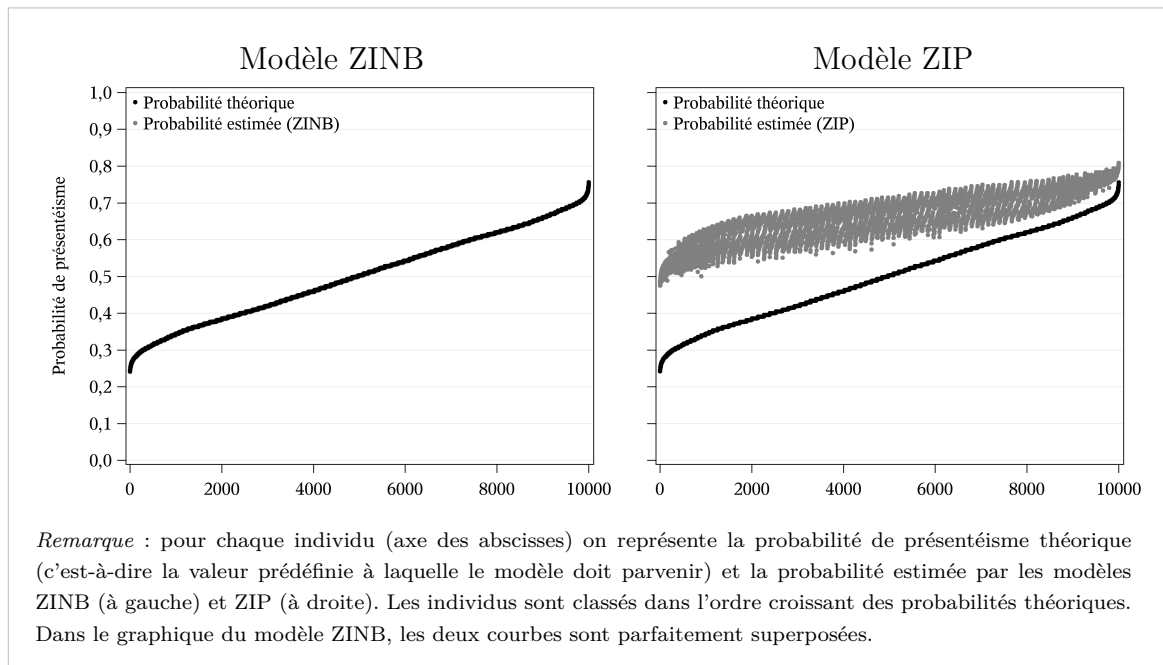
Notons que, sauf indication contraire, le paramètre α est maintenu à 1, ce qui reflète un niveau intermédiaire d'hétérogénéité inter-individuelle.

2.3.1 Simulation n° 4 : variables explicatives corrélées

Une première question se pose en matière de spécification du modèle. Jusqu'ici, les variables explicatives introduites étaient indépendantes. Si deux variables explicatives

présentant un degré de corrélation élevé étaient spécifiées dans le modèle, qu'advierait-il des probabilités de présentéisme? Pour répondre à cette question, deux variables uniformément distribuées sont créées, reflétant respectivement l'AGE (compris entre 20 et 65) et l'ANCIENNETÉ (comprise entre 5 et 25), et sont toutes deux introduites, à la fois dans l'équation de comptage et dans l'équation d'inflation. Les coefficients affectés à la variable AGE demeurent inchangés et ceux de la variable ANCIENNETÉ sont respectivement de 0,02 et -0,02 dans l'équation de comptage et dans l'équation d'inflation. Le coefficient de corrélation entre ces deux variables est calculé à 0,59⁹.

Graphique 3.7 – Probabilités de présentéisme théoriques et estimées (simulation n° 4)



À partir du tableau des résultats 3.2 (page 117), on se rend compte que les coefficients estimés par le modèle ZINB semblent insensibles à la corrélation entre les deux variables. Les coefficients produits reflètent parfaitement les coefficients prédéfinis. Le paramètre de dispersion est évalué à sa juste valeur et n'est pas affecté. Les probabilités de présentéisme sont parfaitement estimées par le modèle ZINB (graphique 3.7),

9. La corrélation est obtenue grâce à une méthode classique dans les simulations de Monte-Carlo. On simule d'abord deux variables normales corrélées en s'appuyant sur la décomposition de Cholesky. Puis on transforme les variables normales en prenant leurs fonctions de répartition pour obtenir deux variables corrélées de distribution uniforme.

tandis que pour le modèle ZIP, la surestimation des probabilités peut sans doute être imputée en grande partie à la présence d'hétérogénéité dans le modèle. La spécification de deux variables corrélées, par exemple l'âge et l'ancienneté, ne semble donc pas être un obstacle à une estimation consistante.

Tableau 3.2 – Coefficients estimés (simulation n° 4)

	Base	ZIP		ZINB	
		$(\bar{x})$	(σ)	$(\bar{x})$	(σ)
CONSTANTE	-1,300	-0,610	0,098	-1,284	0,111
HOMME	-0,600	-0,491	0,039	-0,592	0,043
AGE	0,040	0,033	0,002	0,040	0,003
ENFANTS	0,045	0,037	0,015	0,045	0,016
ANCIENNETÉ	0,020	0,016	0,005	0,019	0,005
α	1,000			0,989	0,093
CONSTANTE	-1,500	0,010	0,102	-1,502	0,207
HOMME	0,300	0,386	0,048	0,308	0,064
AGE	0,040	0,020	0,003	0,040	0,004
ENFANTS	-0,045	-0,043	0,020	-0,042	0,028
ANCIENNETÉ	-0,020	-0,020	0,005	-0,021	0,007

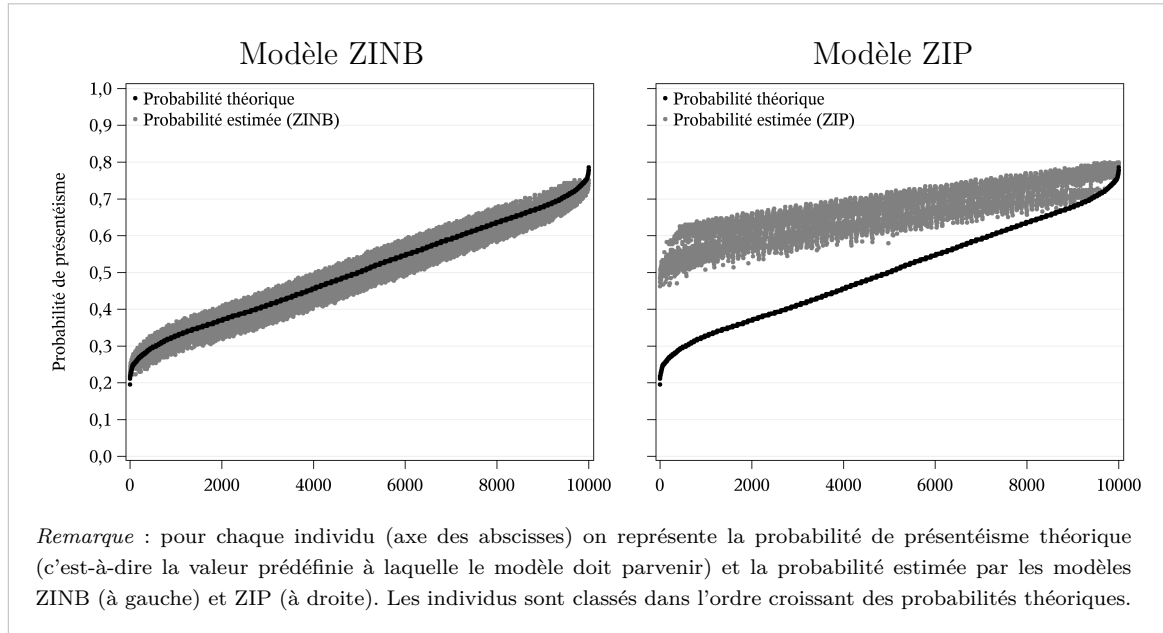
$(\bar{x})$ moyenne des coefficients sur 100 tirages

(σ) écart-type des coefficients sur 100 tirages

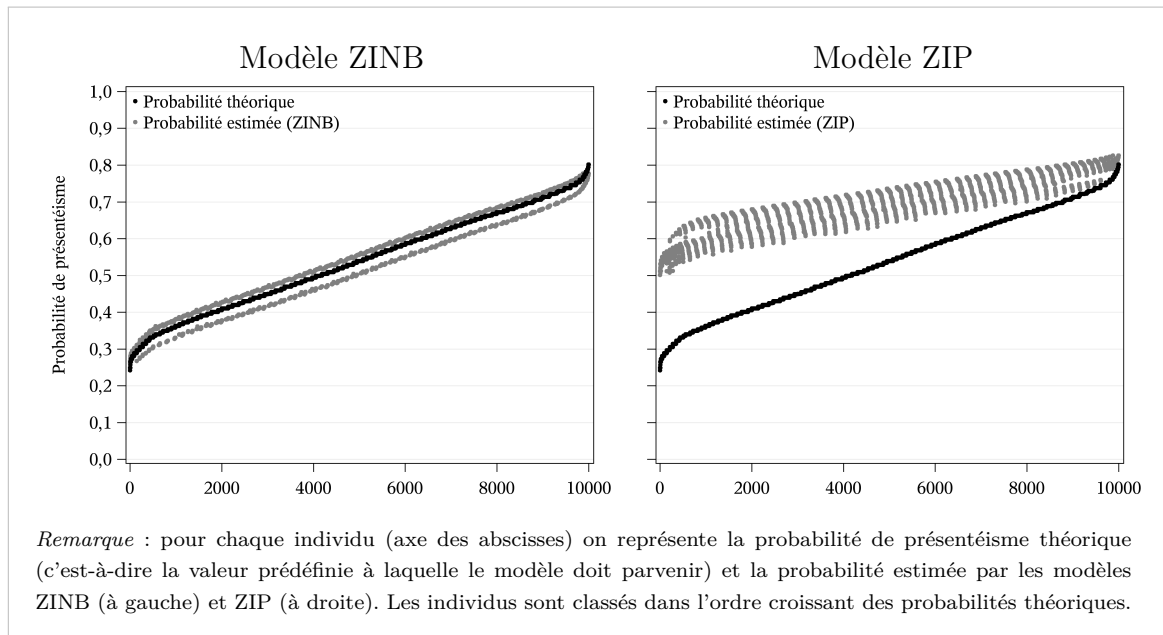
2.3.2 Simulation n° 5 & n° 6 : variables explicatives omises

Une autre manière de savoir si un modèle reste pertinent en présence d'hétérogénéité non-observée est de le confronter à l'omission d'une variable d'importance dans la spécification. Les simulations n° 5 et n° 6 testent respectivement l'omission d'une variable quantitative et l'omission d'une variable binaire. Dans la simulation n° 5, une variable quantitative est créée : ses valeurs, uniformément distribuées, varient entre 5 et 25. La variable entre à la fois dans l'équation de λ avec un coefficient de 0,02 et dans l'équation des probabilités de présentisme avec un coefficient de -0,02. Pour la simulation n° 6, c'est une variable binaire qui est générée, de telle sorte que 30% des individus obtiennent une valeur unitaire. Ses coefficients sont respectivement de 0,2 et de -0,2 dans l'équation de comptage et l'équation d'inflation.

Graphique 3.8 – Probabilités de présentéisme théoriques et estimées (simulation n° 5)



Graphique 3.9 – Probabilités de présentéisme théoriques et estimées (simulation n° 6)



Dans ces deux simulations avec variable omise, le modèle ZINB surestime les constantes de comptage et sous-estime les constantes d'inflation (cf. tableau 3.3). Les coefficients des variables explicatives et le paramètre α restent cependant relativement bien estimés (en dehors de la variable ENFANTS dans la simulation n° 6). Concernant les probabilités de présentisme (graphique 3.8 et 3.9 page 118), ceci a pour effet d'augmenter les écarts entre les valeurs estimées et les valeurs théoriques. Lorsqu'une variable quantitative est omise, les probabilités estimées diffèrent plus ou moins suivant la valeur prise par la variable omise. De même, lorsque la variable omise est binaire, les probabilités sont relativement bien déterminées lorsqu'un individu obtient la modalité nulle (aucun effet) mais sont légèrement sous-estimées lorsqu'il obtient la valeur 1. Dans une variante de la simulation n° 5 (non présentée ici), la variable quantitative n'a été omise que dans l'équation de comptage. Dans ce cas, il n'y a quasiment aucune incidence sur l'estimation des probabilités pour le modèle ZINB.

En conclusion, l'omission d'une variable importante dans la spécification peut avoir des implications importantes, dont il faut être conscient. Mais les simulations montrent également que l'ordonnancement général des probabilités reste correct. Lorsqu'il s'agit d'étudier le comportement de deux sous-populations dans une même organisation, la détection des individus les plus présentistes reste excellente, et pour les autres, les écarts restent minimes.

Tableau 3.3 – Coefficients estimés (simulation n° 5 et n° 6)

	Base	Simulation n° 5				Simulation n° 6			
		ZIP		ZINB		ZIP		ZINB	
		$(\bar{x})$	(σ)	$(\bar{x})$	(σ)	$(\bar{x})$	(σ)	$(\bar{x})$	(σ)
CONSTANTE	-1,300	-0,352	0,105	-1,001	0,119	-0,499	0,100	-1,151	0,119
HOMME	-0,600	-0,492	0,043	-0,607	0,049	-0,490	0,054	-0,596	0,061
AGE	0,040	0,033	0,002	0,040	0,003	0,033	0,002	0,040	0,003
ENFANTS	0,045	0,036	0,016	0,044	0,018	0,036	0,018	0,046	0,020
α	1,000			1,033	0,095			0,995	0,113
CONSTANTE	-1,500	-0,274	0,111	-1,857	0,234	-0,172	0,128	-1,639	0,256
HOMME	0,300	0,399	0,047	0,309	0,066	0,389	0,058	0,310	0,076
AGE	0,040	0,019	0,003	0,041	0,004	0,021	0,003	0,040	0,004
ENFANTS	-0,045	-0,044	0,020	-0,044	0,030	-0,039	0,019	-0,035	0,026

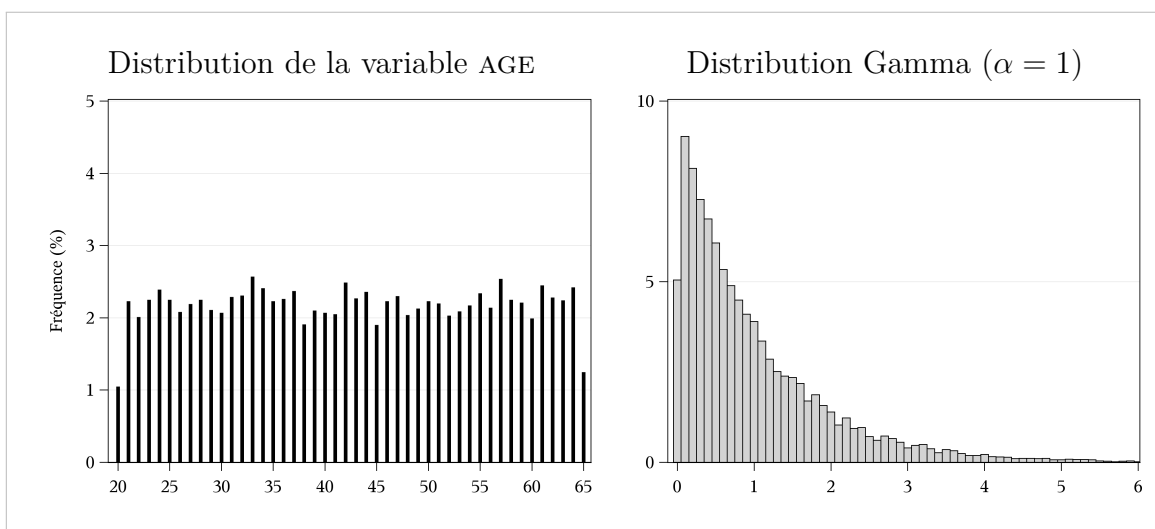
$(\bar{x})$ moyenne des coefficients sur 100 tirages

(σ) écart-type des coefficients sur 100 tirages

2.3.3 Simulation n° 7 & n° 8 : hypothèses d'hétérogénéité

Les deux simulations suivantes analysent les implications d'une modélisation erronée de l'hétérogénéité entre individus. Ce point est d'une importance cruciale pour la qualité de l'estimation. Il a été observé qu'une loi Gamma dont les deux paramètres sont identiques semble être un bon choix pour modéliser les différences entre individus, puisqu'elle permet de s'adapter à des distributions fort diverses. Néanmoins, il s'agit d'une hypothèse et il serait intéressant de savoir si une erreur peut biaiser la mesure des probabilités. Rappelons que dans le cadre de ces simulations, l'hétérogénéité inter-individuelle est imputée uniquement à la santé des individus, par simplification.

Graphique 3.10 – Distributions des variables corrélées (simulation n° 7)

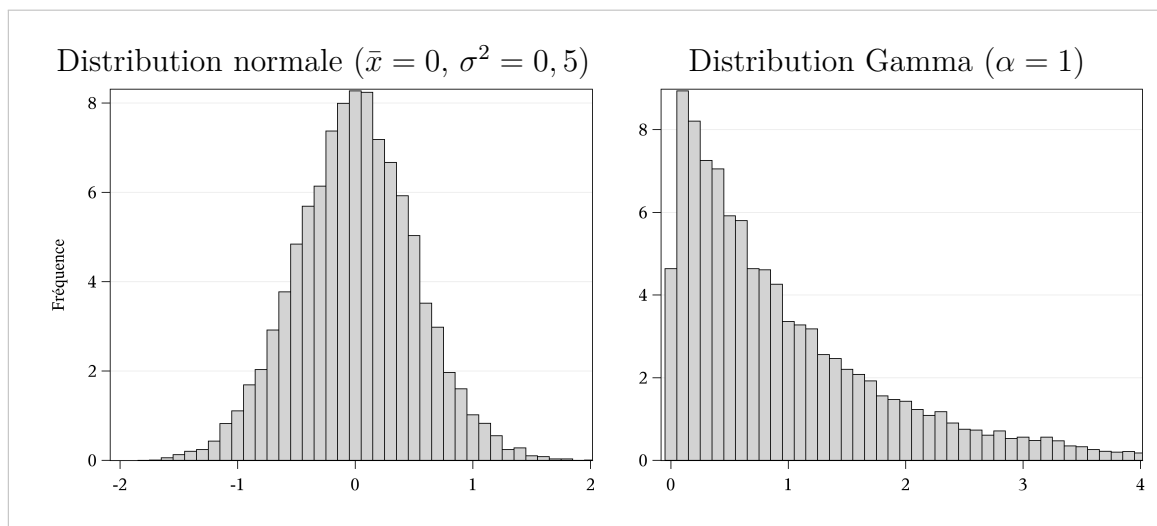


Pour cela, deux situations sont mises en avant. Dans la simulation n° 7, on considère que le facteur d'hétérogénéité est corrélé avec l'une des variables explicatives du modèle. Comme précédemment, deux variables corrélées ont été créées¹⁰, l'une étant uniformément distribuée et simulant la variable d'AGE (comprise entre 20 et 65) et l'autre, la variable SANTÉ suivant une distribution Gamma de moyenne unitaire, avec $\alpha = 1$. La distribution de ces deux variables est présentée sur le graphique 3.10 : un coefficient de corrélation de 0,54 est calculé entre elles.

10. La corrélation est toujours obtenue en simulant deux variables normales grâce à la décomposition de Cholesky. En prenant les fonctions de répartition des variables normales, on obtient donc deux variables de distribution uniforme. Puis, en utilisant la loi Gamma inverse, on donne une distribution Gamma à la seconde variable.

Ensuite, dans la simulation n° 8, c'est la question de la distribution de l'hétérogénéité qui se pose. L'estimation est-elle affectée lorsque celle-ci est en réalité normalement distribuée? Dans le cas d'une distribution normale, le paramètre u ne peut être entré multiplicativement dans la loi de Poisson, sous peine de générer des valeurs négatives de λ . Il doit donc être entré additivement dans la spécification de λ . Nous avons opté pour une distribution normale de moyenne nulle et de variance égale à 0,5. Augmenter inconsidérément la variance de la distribution normale conduirait à des valeurs d'hétérogénéité extrêmes et de ce fait, à des données d'absences ne reflétant plus la réalité. Le graphique 3.11 présente cette distribution normale ainsi qu'une distribution Gamma.

Graphique 3.11 – Distributions de l'hétérogénéité (simulation n° 8)



Les probabilités de présentisme estimées dans la simulation n° 7 sont présentés sur le graphique 3.12 (page 122). Lorsque le paramètre d'hétérogénéité est corrélé avec une variable explicative du modèle, il y a une surestimation des probabilités, notamment pour les plus faibles. Même si l'ordonnancement global des probabilités de présentisme est respecté, cette surestimation peut s'avérer gênante et oblige à nuancer certaines conclusions¹¹. Par ailleurs, comme on le voit dans le tableau 3.4 (page 122), le problème est surtout lié à une surestimation de la constante d'inflation. Les coefficients de l'équation d'inflation sont quant à eux plutôt sous-estimés. En comparaison, la qualité de l'estimation est davantage mise à mal dans l'équation de comptage.

11. Cependant, dans les applications empiriques qui suivent, la possibilité d'une éventuelle corrélation entre l'âge et la santé est limitée par l'exclusion des maladies de longue durée.

Graphique 3.12 – Probabilités de présentisme théoriques et estimées (simulation n° 7)

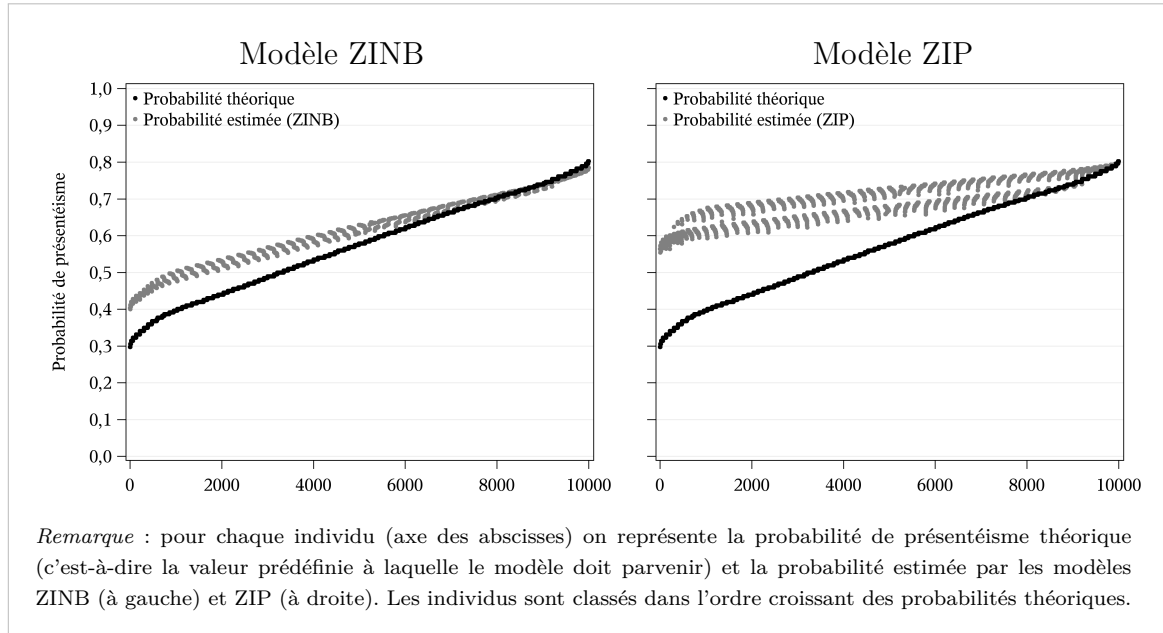


Tableau 3.4 – Coefficients estimés (simulation n° 7 & n° 8)

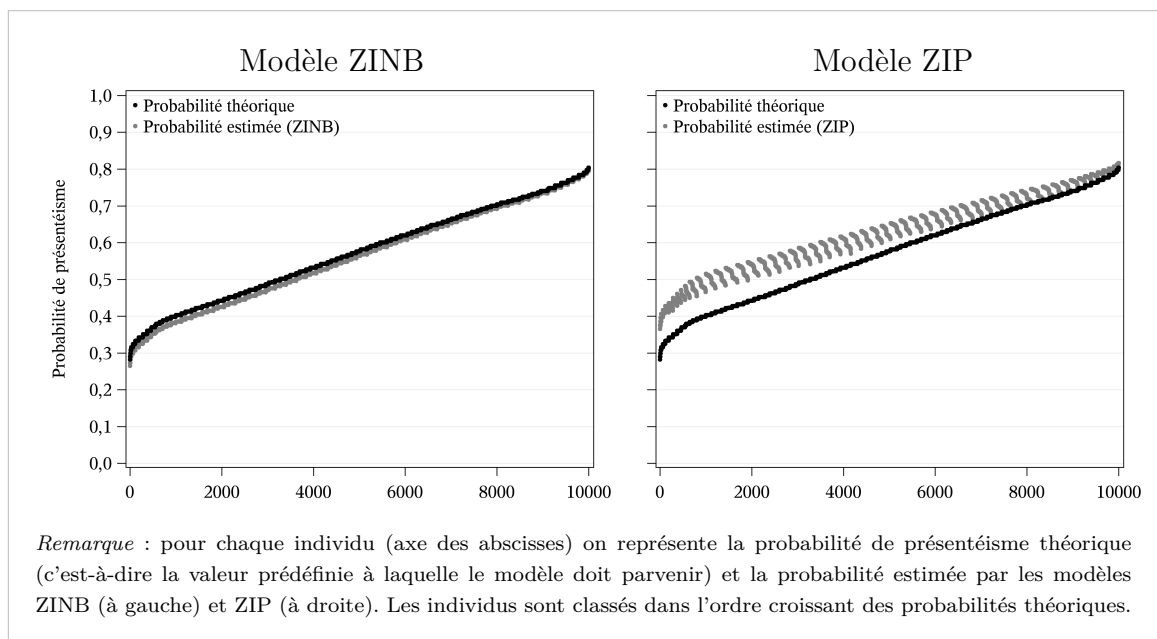
	Base	Simulation n° 7				Simulation n° 8			
		ZIP		ZINB		ZIP		ZINB	
		$(\bar{x})$	(σ)	$(\bar{x})$	(σ)	$(\bar{x})$	(σ)	$(\bar{x})$	(σ)
CONSTANTE	-1,300	-2,498	0,150	-2,961	0,149	-0,948	0,090	-1,216	0,094
HOMME	-0,600	-0,540	0,043	-0,605	0,050	-0,545	0,044	-0,596	0,048
AGE	0,040	0,071	0,003	0,078	0,003	0,037	0,002	0,040	0,002
ENFANTS	0,045	0,048	0,016	0,048	0,016	0,042	0,015	0,046	0,016
α				0,493	0,054			0,307	0,041
CONSTANTE	-1,500	0,107	0,199	-0,825	0,257	-0,931	0,145	-1,601	0,190
HOMME	0,300	0,340	0,055	0,280	0,064	0,353	0,067	0,306	0,077
AGE	0,040	0,014	0,004	0,028	0,005	0,032	0,003	0,041	0,003
ENFANTS	-0,045	-0,035	0,020	-0,034	0,023	-0,044	0,020	-0,042	0,023

$(\bar{x})$ moyenne des coefficients sur 100 tirages

(σ) écart-type des coefficients sur 100 tirages

D'autre part, les résultats de la simulation n° 8 mettent en évidence qu'une distribution normale du paramètre d'hétérogénéité n'a pas d'impact majeur sur la qualité de l'estimation (graphique 3.13). La souplesse de la loi Gamma intégrée au modèle ZINB semble capable de fournir des coefficients de qualité, même lorsque le paramètre d'hétérogénéité est normalement distribué. À condition, sans doute, que les paramètres de la distribution permettent une modélisation réaliste de l'espérance d'absence. Nous avons réalisé la même simulation toujours avec une moyenne nulle, mais avec une variance égale à 1. Les probabilités demeurent bien ordonnées, mais sont légèrement sous-estimées par le modèle ZINB. Même dans cette situation extrême elles seraient donc acceptables *a minima*. Cependant, à cause des valeurs extrêmes du paramètre d'hétérogénéité, les données d'absences perdent tout réalisme : plus de 95% des individus ont alors une valeur d'absence inférieure ou égale à 4.

Graphique 3.13 – Probabilités de présentéisme théoriques et estimées (simulation n° 8)

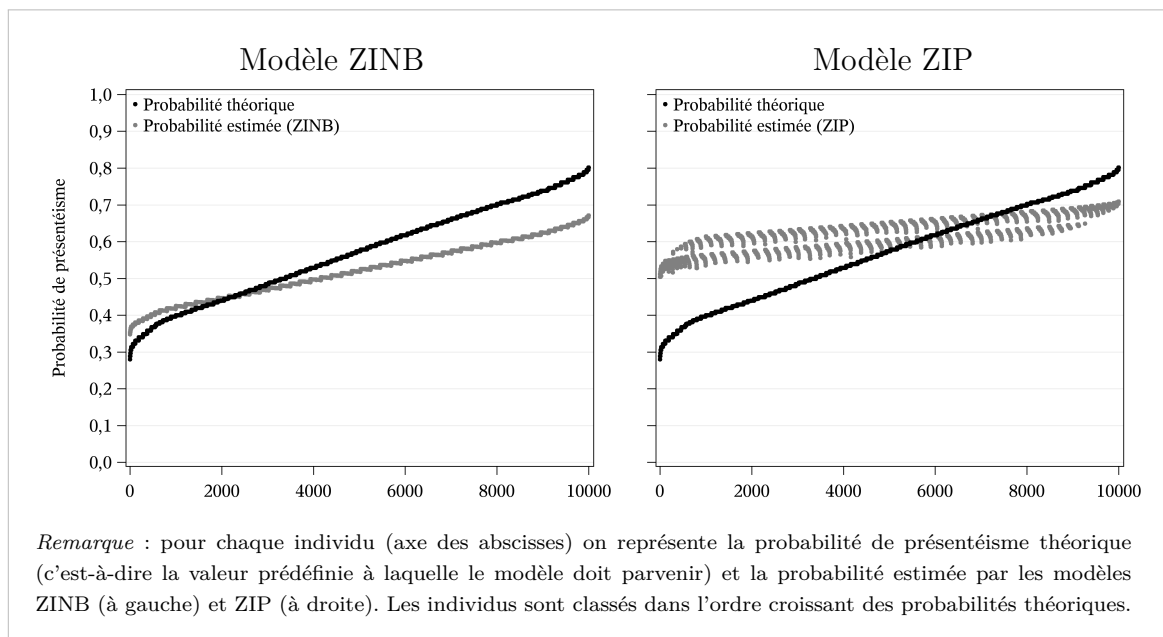


2.3.4 Simulation n° 9 : présentéisme et maladies graves

Les deux prochaines simulations évaluent des objections intuitives évoquées plus haut. Jusqu'ici, nous avons considéré que tous les salariés étaient en mesure de décider s'ils se rendaient ou pas à leur travail, alors qu'ils sont malades. Or, en réalité, l'état de

santé de certains d'entre eux peut être trop dégradé, ou la maladie qu'ils connaissent trop grave pour qu'un tel choix soit possible : ceux-ci n'ont d'autre alternative que de s'absenter pour se soigner. Au-delà d'un certain seuil, le présentéisme n'est donc plus possible. L'une de nos premières décisions pour contourner ce problème a été d'exclure du périmètre d'analyse les individus ayant de très longues maladies. Mais si l'on admet qu'une certaine part de la population restante est encore trop malade pour être présentéiste, comment le modèle réagit-il ? Dans la simulation n° 9, nous nous concentrons donc sur 20% des individus dont l'état de santé est le plus dégradé. Ces 2000 individus, n'ont pas d'autre choix que de s'absenter, même si leur propension au présentéisme est élevée : leur valeur d'absence est générée par le processus poissonnien standard.

Graphique 3.14 – Probabilités de présentéisme théoriques et estimées (simulation n° 9)



Dans le modèle ZINB, les coefficients sont évalués en-deçà de leur valeur prédéfinie (tableau 3.5 page 125), ce qui produit une légère sous-estimation des probabilités de présentéisme (graphique 3.14). Seules les probabilités les plus faibles sont légèrement surévaluées. Cependant, les conclusions que l'on peut tirer d'un tel modèle restent tout à fait pertinentes *a minima*. L'incidence des facteurs de présentéisme est en effet minorée par rapport à la réalité. Dans ce cas, l'effet réel serait encore plus élevé. Cette

thèse prétendant simplement offrir aux entreprises un outil fiable en première analyse, le résultat observé est donc tout à fait acceptable.

Tableau 3.5 – Coefficients estimés (simulation n° 9)

	Base	ZIP		ZINB	
		$(\bar{x})$	(σ)	$(\bar{x})$	(σ)
CONSTANTE	-1,300	-0,685	0,087	-1,138	0,093
HOMME	-0,600	-0,473	0,039	-0,549	0,044
AGE	0,040	0,040	0,002	0,046	0,002
ENFANTS	0,045	0,033	0,015	0,039	0,015
α	1,000			0,579	0,048
CONSTANTE	-1,500	0,002	0,127	-0,942	0,186
HOMME	0,300	0,274	0,051	0,185	0,061
AGE	0,040	0,009	0,002	0,023	0,003
ENFANTS	-0,045	-0,029	0,020	-0,024	0,024

$(\bar{x})$ moyenne des coefficients sur 100 tirages

(σ) écart-type des coefficients sur 100 tirages

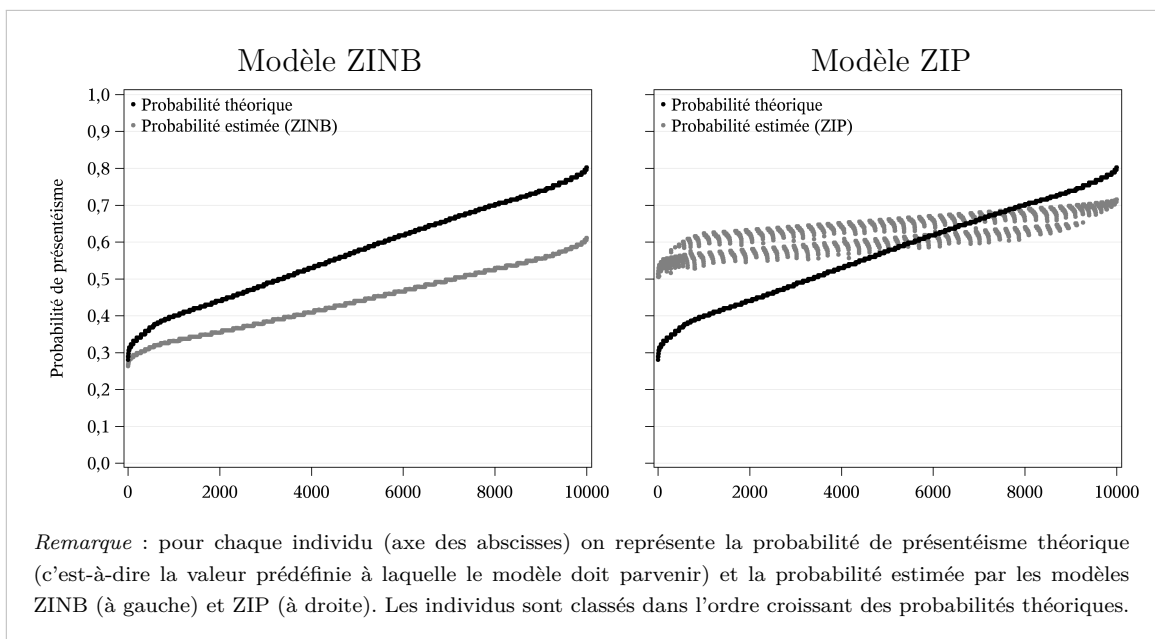
2.3.5 Simulation n° 10 : présentéisme et valeurs positives d'absence

Nous avons testé une autre hypothèse qui porte directement sur le comportement de présentéisme. Dans les simulations précédentes, les individus présentéistes obtenaient systématiquement une valeur d'absence nulle. Ces valeurs nulles sont l'expression la plus manifeste du présentéisme et le modèle s'en sert pour définir le profil statistique des présentéistes. Cependant, on pourrait admettre qu'un individu ayant deux jours d'absence, par exemple, aurait été absent pendant 6 jours dans une situation normale. Il peut par exemple être retourné au travail avant une guérison complète. Cet individu a effectivement été présentéiste pendant 4 jours, mais sa valeur d'absence observée n'est pas nulle. La question est alors de savoir si le modèle reste pertinent lorsque cette hypothèse initiale est relâchée.

Une nouvelle règle fait alors son entrée dans le modèle. Dans la simulation n° 10, les individus considérés comme présentéistes n'obtiennent pas systématiquement une valeur nulle. Pour les 20 % des présentéistes ayant la plus mauvaise santé, l'absence n'est pas nulle, mais la valeur en est générée aléatoirement sur la base de leur espérance

individuelle. Cependant, parce qu'ils sont présentéistes, leur espérance d'absence est tout de même minorée (divisée par 3).

Graphique 3.15 – Probabilités de présentéisme théoriques et estimées (simulation n° 10)



Encore une fois, les probabilités de présentéisme sont sous-estimées par le modèle ZINB (graphique 3.15). Ce que nous évoquions plus haut reste vrai : notre objectif est en définitive de fournir une première analyse du présentéisme. Toute inférence sur les coefficients obtenus est au moins juste *a minima* (tableau 3.6 page 127). Une statistique intéressante émerge aussi des simulations. Si l'on se concentre sur 25% de la population ayant les probabilités de présentéisme (théoriques) les plus élevées (les plus présentéistes, donc), on découvre que sur un tirage donné, environ 28,6% de ces individus ont également une valeur d'absence positive. Dans cette situation, le présentéisme ne se manifeste donc pas uniquement au travers de valeurs nulles.

Tableau 3.6 – Coefficients estimés (simulation n° 10)

	Base	ZIP		ZINB	
		$(\bar{x})$	(σ)	$(\bar{x})$	(σ)
CONSTANTE	-1,300	-0,613	0,108	-1,215	0,124
HOMME	-0,600	-0,525	0,047	-0,625	0,054
AGE	0,040	0,031	0,002	0,036	0,002
ENFANTS	0,045	0,041	0,016	0,048	0,018
α	1,000			0,889	0,113
CONSTANTE	-1,500	0,003	0,136	-1,466	0,301
HOMME	0,300	0,288	0,061	0,182	0,097
AGE	0,040	0,01	0,002	0,027	0,005
ENFANTS	-0,045	-0,027	0,023	-0,022	0,034

$(\bar{x})$ moyenne des coefficients sur 100 tirages

(σ) écart-type des coefficients sur 100 tirages

2.4 Cas particulier des données en panels

Pour conclure cette batterie de simulations, nous voudrions tester l'efficacité des modèles en panels. Bien que ceux-ci ne soient pas au centre de cette thèse, il peuvent s'avérer utiles pour nos travaux futurs et il est intéressant d'offrir un aperçu de nos premiers résultats. N'étant pas implémentés par défaut dans les logiciels statistiques classiques, ces modèles en panels impliquent des méthodes d'estimation plus lourdes et plus longues. Le contexte de la simulation n° 11 a donc été considérablement allégé : l'analyse porte désormais sur 1 000 individus observés sur cinq périodes successives (5 000 observations). Il y a toujours trois variables explicatives : la variable HOMME (invariable dans le temps), la variable AGE (augmentant de 1 à chaque période) et une variable SALAIRE uniformément distribuée. L'intérêt d'introduire cette dernière est d'observer l'effet d'une variable explicative qui soit davantage fluctuante suivant les différentes périodes : d'abord fixée pour chaque individu à une valeur comprise entre 1 et 2,5, elle connaît une augmentation comprise en 0 et 3% à chaque période. Enfin un niveau de santé est affecté à chaque individu sous la forme d'une variable SANTÉ suivant une distribution Gamma dont le paramètre est $\alpha = 0,5$. Cet état de santé demeure fixe sur les 5 périodes, ce qui implique une relation entre les valeurs d'absence

successives d'un même individu. Dans le même esprit que les simulations précédentes, l'espérance λ pour un individu i à une période t est définie par l'équation suivante :

$$\lambda_{it} = \exp(\mu_{it}) \times \text{SANTÉ} \quad (3.16)$$

avec

$$\begin{aligned} \mu_{it} = & 1 \times (1 + \epsilon_1) - 0,6 \times (1 + \epsilon_2) \times \text{HOMME} \\ & + 0,04 \times (1 + \epsilon_3) \times \text{AGE} - 0,3 \times (1 + \epsilon_4) \times \text{SALAIRE} \end{aligned} \quad (3.17)$$

De même la probabilité de présentéisme p_{it} est quant à elle calculée comme suit :

$$p_{it} = \frac{\exp(\eta_{it})}{1 + \exp(\eta_{it})} \quad (3.18)$$

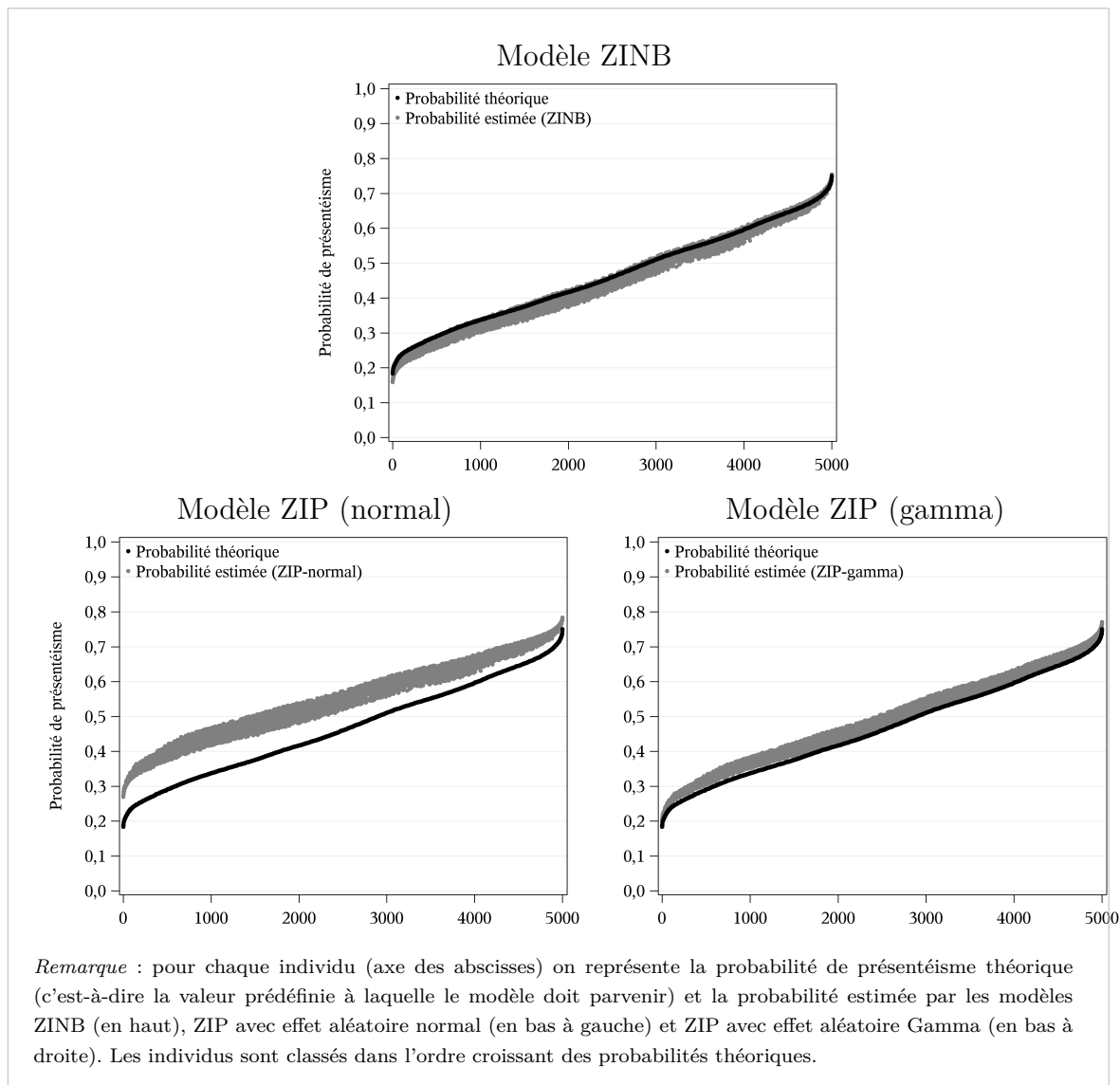
avec

$$\begin{aligned} \eta_{it} = & - 1,6 \times (1 + \epsilon_5) + 0,3 \times (1 + \epsilon_6) \times \text{HOMME} \\ & + 0,04 \times (1 + \epsilon_7) \times \text{AGE} - 0,3 \times (1 + \epsilon_8) \times \text{SALAIRE} \end{aligned} \quad (3.19)$$

Le déroulement de la simulation reste le même que précédemment, si ce n'est que trois modèles sont testés : le modèle ZINB en coupe transversale et deux modèles ZIP en panels dont les effets aléatoires suivent respectivement une distribution normale et une distribution Gamma.

Même si les résultats de cette simulation n°11 sont à interpréter avec vigilance, on s'aperçoit que le modèle ZIP en panels faisant l'hypothèse d'une distribution normale de l'effet aléatoire surestime les probabilités de présentéisme tandis que les deux autres modèles sont davantage ajustés aux probabilités théoriques (graphique 3.16 page 129). Le modèle ZINB en coupe transversale, tout en ne tenant pas compte de l'aspect répétitif des données, produit une estimation raisonnable des probabilités. Peut-on cependant aller jusqu'à dire qu'il peut être utilisé directement sur des données en panels ? Cette simulation, d'une portée limitée tant sur la méthodologie de simulation que d'estimation ne permet pas de trancher la question.

Graphique 3.16 – Probabilités de présentéisme théoriques et estimées (simulation n° 11)



Conclusion

Ce chapitre était particulièrement consacré au traitement des valeurs nulles dans les données d'absence, qui constituent le matériau essentiel de notre mesure du présentéisme. Nous mettons en lumière qu'elles ne tirent pas leur origine d'un processus unique caractérisé par l'absence de maladie, mais qu'elles ont une double origine. D'une

part, elles peuvent bien entendu être le fait d'un individu dont la santé n'a pas nécessité d'arrêt de travail. En termes de modélisation, un simple modèle de comptage suffit à générer ces zéros qui peuvent être considérés comme « naturels ». Un autre processus se superpose cependant et produit d'autres valeurs nulles. De ce processus dit « structurel » ou « stratégique », ne peuvent d'ailleurs naître que des zéros. Ils proviennent d'individus appliquant une règle comportementale en s'interdisant de prendre des jours d'absence, même lorsque leur santé le justifierait. Ce sont les présentéistes.

Les modèles à inflation de zéros sont conçus pour appréhender cette double origine. Parmi les individus ayant une valeur nulle d'absence, ces modèles distinguent ceux qui font partie du régime structurel ne produisant que des valeurs nulles de ceux qui relèvent du régime normal générant des zéros « circonstanciels » ainsi que des valeurs positives. Ces modèles estiment une probabilité d'appartenir au premier régime que nous interprétons comme une probabilité de présentéisme. Cette analyse est non seulement confortée par l'étude approfondie de la variable d'absence, mais également par les utilisateurs des modèles *zero-inflated* dans la littérature. Ces derniers mettent notamment en évidence qu'interpréter la probabilité estimée par le modèle comme une probabilité de non-occurrence (i.e. de ne pas être absent) est erroné. Il s'agit bien d'une probabilité d'appartenir au groupe d'individus produisant des valeurs nulles « inhabituelles ».

Dans la seconde partie de ce chapitre, nous avons entamé l'application des modèles de comptage à des données. Cette mise en œuvre s'est faite sur des données simulées : tous les paramètres étant sous contrôle dans ce contexte, il est possible d'obtenir une mesure des biais engendrés par certaines prises de distances avec les hypothèses théoriques initiales. Les enseignements à tirer de ces simulations de Monte-Carlo sont essentiellement de trois ordres pour le modèle ZINB. Premièrement, le modèle fonctionne parfaitement en présence d'hétérogénéité non observée. Même lorsque la distribution n'est pas une loi Gamma, le modèle est capable d'estimer correctement les coefficients. L'existence d'une corrélation entre le terme d'hétérogénéité et une variable explicative produit quand à lui une légère perte de qualité de l'estimation.

Deuxièmement, au vu des résultats, l'utilisateur des modèles *zero-inflated* est incité à la prudence en matière de spécification. L'omission d'une variable, notamment quantitative, peut impliquer une certaine variabilité dans les résultats. Pour le modèle ZINB,

cet effet est cependant assez limité. De même, la coexistence dans le modèle de deux variables corrélées ne semble pas poser de problème majeur.

Troisièmement, lorsque les hypothèses initiales sont relâchées, par exemple si les individus en mauvaise santé n'ont pas la possibilité d'être présentéistes ou si le présentéisme peut également se manifester par des valeurs faiblement positives d'absence, les probabilités sont estimées *a minima*. C'est là l'un des enseignements importants, transversal à presque toutes les simulations. Lorsqu'il existe des différences entre les probabilités estimées et les probabilités réelles, les écarts apparaissent toujours à cause d'une sous-estimation. Ainsi, lorsqu'un effet est mesuré par le modèle il est non seulement existant, mais sans doute plus prononcé dans la réalité.

Chapitre 4

Application empirique : deux études de cas

Introduction

Un projet de fond oriente cette thèse depuis ses premiers paragraphes. Nous avons toujours eu à l'esprit de souligner à quel point l'absence-maladie et le présentéisme constituent deux comportements connexes, car deux réponses possibles face à un incident de santé. Dès le premier chapitre, nous observions que la recherche autour de l'absence a constitué le terreau scientifique dans lequel est né et s'est développé le concept de présentéisme. Cette thématique y a d'abord émergé comme un élément central de réflexion sur les conséquences possibles de la « lutte contre l'absentéisme ». Mais le présentéisme s'est aussi construit concomitamment à la redécouverte de l'absence comme une problématique de santé, et non comme un simple dysfonctionnement organisationnel qu'il faudrait à tout prix résoudre.

Par la suite, le deuxième chapitre a énoncé cette question du présentéisme dans un vocabulaire plus statistique. En examinant de plus près les données d'absence, leur nature particulière et les moyens utilisables pour leur modélisation, il est peu à peu apparu

que ce type de données renferme davantage d'informations que la somme annuelle des absences individuelles. Elles contiennent des indications précieuses sur la propension des salariés au présentéisme. Mais, en l'absence d'un outil d'identification permettant de démêler des comportements fortement imbriqués au sein des données d'absence, l'intérêt de celles-ci pour l'étude du présentéisme n'a jusqu'ici guère attiré l'attention. La difficulté de la mesure statistique du présentéisme est sans doute également l'une des raisons expliquant la méconnaissance du phénomène en entreprise.

L'instrument de mesure que nous proposons a été scruté par le détail dans le troisième chapitre. Le présentéisme y a été envisagé comme une cause d'hétérogénéité que nous avons appelée comportementale et se manifestant dans les données d'absence par une surreprésentation des valeurs nulles. Le modèle binomial négatif à inflation de zéros (ZINB) est un outil capable de fournir à la fois une mesure de l'espérance d'absence, sur laquelle sont classiquement fondés les modèles de comptage, et une probabilité d'appartenance à un groupe d'individus ayant pour habitude d'être présents au travail malgré la maladie, les présentéistes.

Jusqu'ici, les données d'absence ont été étudiées pour elles-mêmes. Des statistiques descriptives ont mis en évidence leurs spécificités, par exemple une distribution asymétrique, une importante concentration ou encore une sensibilité aux épisodes épidémiques. En dernier lieu, nous avons simulé des données d'absence. Grâce au contrôle de l'ensemble des paramètres explicatifs qu'autorisent les simulations de Monte-Carlo, les forces et les limites du modèle ZINB ont été mises en lumière.

Avec les deux derniers chapitres, s'ouvre un mouvement essentiel de cette thèse. Le modèle ZINB va à présent être confronté à de réelles données d'entreprise. L'un des intérêts majeurs de cette méthodologie est de pouvoir être appliquée aisément à des situations concrètes d'entreprises, à partir d'une base de données reprenant les enregistrements individuels des absences des salariés. Tout au long de ces deux chapitres nous nous intéresserons d'une part au cas d'une banque de couverture régionale, et d'autre part à celui d'un centre hospitalier également situé dans le Nord-Pas-de-Calais. Les données disponibles pour ces deux organisations vont d'abord être décrites par le détail. Non seulement leurs cœurs de métier sont distincts et impliquent des contrastes comportementaux vis-à-vis de l'absence-maladie et du présentéisme, mais leurs modes de gestion sont également différents, la banque étant issue du secteur privé, l'hôpital s'inscrivant dans le contexte du service public.

Ensuite, nous avons fait le choix de scinder en deux parts la description des résultats économétriques. Ce chapitre sera spécifiquement consacré à une approche technique des résultats : présentation et interprétation des sorties économétriques (section 2), statistiques descriptives sur les résultats et ajustement des modèles (section 3), intérêt d'un modèle ZINB sans variables explicatives (section 3.4), exploitation de la dimension longitudinale (section 4). Tout en proposant progressivement des clefs de lecture et de compréhension des résultats, l'idée maîtresse est encore principalement de souligner l'intérêt et la pertinence du modèle pour la mesure du présentéisme. En revanche, le cinquième et dernier chapitre sera quant à lui dédié à l'analyse de ces résultats dans une perspective de meilleure connaissance du présentéisme. Nous ne manquerons pas alors de revenir à l'état des lieux dressé dans le premier chapitre afin de situer nos propres résultats vis-à-vis de ceux qui existent déjà dans la littérature, bien qu'ils aient été obtenus grâce à des méthodes différentes.

1 Présentation des bases de données

Dans un premier temps nous décrivons les deux contextes dans lesquels se déroulera l'analyse des comportements de présentéisme et d'absence-maladie. L'objet n'est pas d'effectuer une étude comparative entre ces deux contextes : il serait en effet tentant d'extrapoler nos conclusions et d'en faire une comparaison entre secteur privé et public. Nous ne prétendons pas ici à la généralisation, mais aborderons les deux organisations avec leurs spécificités, leurs politiques propres et l'environnement de travail qu'elles procurent à leurs salariés. Notre méthodologie a essentiellement une visée opérationnelle, autorisant un diagnostic local pertinent dans les murs de l'entreprise. Bien entendu, comme ceux de toute autre étude de cas, nos résultats alimentent la connaissance générale du présentéisme et de ses mécanismes, s'ajoutant aux autres preuves empiriques de la littérature.

Pour « faire connaissance » avec ces micro-données, nous développerons trois éléments pour chacune des deux organisations. D'abord, bien que ce point ait déjà été abordé auparavant, nous reviendrons sur les données d'absence et donnerons quelques informations supplémentaires sur leur structure générale et leur évolution. Ensuite, nous décrirons la structure démographique des deux organisations ainsi que les variables démographiques qui seront intégrées aux régressions à venir. Pour finir, nous ferons de

même pour les variables organisationnelles, c'est-à-dire celles qui sont liées au contrat ou à l'emploi des salariés.

1.1 Établissement bancaire

La banque régionale que nous étudions (nommée « établissement bancaire » ci-après) est une entité régionale (Nord-Pas-de-Calais) d'un groupe bancaire d'envergure nationale et internationale. Si sa stratégie globale et sa politique de développement sont déterminées par des structures supra-régionales, cette entreprise dispose d'une indépendance en termes de gestion. Elle est dirigée par un directoire et un conseil de surveillance. Les décisions portant sur les ressources humaines sont prises dans le cadre de ce directoire, dont l'un des membres est spécifiquement chargé de la fonction RH.

En 2008, 71,5% des 2 708 salariés¹ sont employés dans le réseau commercial de la banque. Cette fonction de « banque de détail » est sans doute la plus connue : il s'agit de la mise à disposition de moyens de paiement, de la commercialisation de produits d'épargne ou de crédit, de produits d'assurance, etc. Cependant, l'entreprise exerce également comme banque d'affaires et d'investissement auprès d'entreprises, d'associations ou de collectivités.

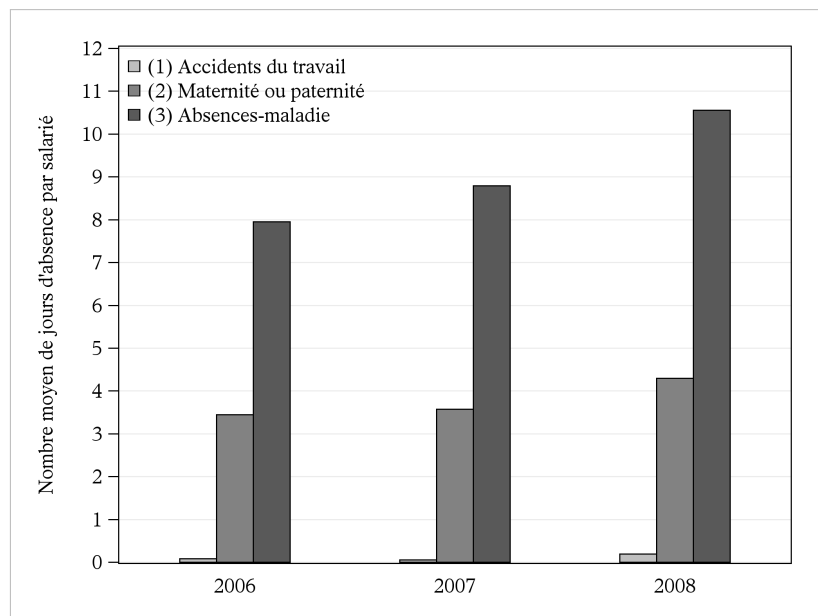
Dans sa forme actuelle, l'entreprise est relativement jeune, puisqu'elle est née en 2007 de la fusion de plusieurs entités infra-régionales. Cet évènement sera à conserver en mémoire dans l'analyse des comportements qui nous intéressent, puisqu'il a occasionné d'importantes mobilités de personnel et la suppression d'un certain nombre d'emplois. Il implique également que lors de l'année subséquente à la fusion, qui est également la période d'analyse dans cette thèse, de nombreux ajustements organisationnels ont dû avoir lieu, ne serait-ce qu'à cause de la diversité des méthodes de gestion ou des habitudes de travail qui existaient dans les anciennes structures. Par ailleurs, dans la continuité de la fusion, l'établissement bancaire a engagé une réorganisation de ses services administratifs. Celle-ci consistait principalement en des mesures de départs volontaires (retraite, création d'entreprise, projet personnel) ou de mobilité interne des salariés, notamment vers des emplois du réseau commercial.

1. Ces 2 708 salariés représentent 2 344 ETP (équivalent temps plein).

1.1.1 Absences-maladie et autres absences

Ensemble des absences. Dans ce contexte de fusion, l'établissement bancaire s'est interrogé sur l'explication à donner à une progression du nombre d'absences-maladie. Entre l'année 2006, précédant la fusion, et l'année 2008, suivant la fusion, le nombre moyen de journées d'absence-maladie a en effet progressé de 32,8%. En moyenne, les salariés se sont absentés 10,6 jours en 2008 contre 7,9 en 2006. Même si nous nous intéressons principalement à l'absence-maladie, ce constat est valable pour l'ensemble des catégories d'absence (graphique 4.1). Tous motifs confondus, la somme des jours d'absence correspondrait à environ 180 salariés à temps plein absents tout au long de l'année 2008. Le graphique 4.1 montre l'évolution des trois composantes essentielles de l'absence dans cette entreprise : la maladie est la cause majeure d'absence (60,7% de l'ensemble des absences en 2008), suivi de la maternité (24,7%). Les accidents du travail ne représentent quant à eux qu'une très faible part des absences (1,1%). D'autres types d'absences sont également enregistrés par l'établissement bancaire mais ne sont pas abordés ici².

Graphique 4.1 – Évolution des absences selon leur type (Banque de 2006 à 2008)



2. Il s'agit notamment des absences pour événements familiaux (baptême, mariage, enfant malade, décès, etc.) et les congés sans solde (congé sabbatique, congé création d'entreprise, etc.).

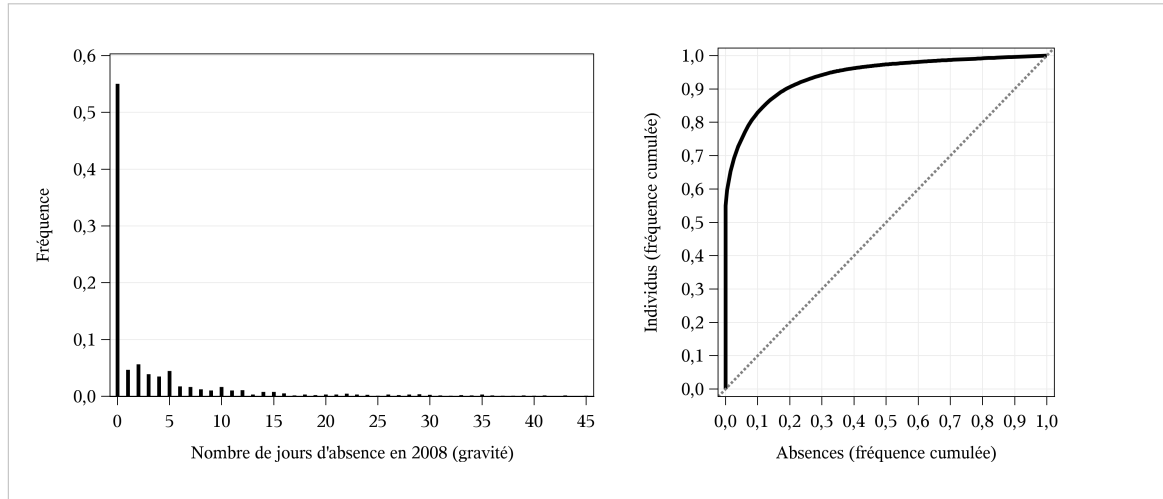
L'entreprise compare également son propre niveau d'absence à celui du groupe au niveau national. Elle s'appuie en cela sur l'indicateur classique qu'est le « taux d'absentéisme »³ : si en 2006, il était de 6,52% en moyenne pour les entités antérieures à la fusion contre 7,06% sur l'ensemble du groupe, en 2008 le taux régional dépasse le taux national (7,71% contre 7,47%), notamment pour l'absence-maladie (4,69% contre 4,16%).

L'élément essentiel de la base de données que nous étudierons est donc le nombre de jours d'absence enregistré pour chaque individu. La stratégie d'analyse de cette variable dépendante implique un certain nombre de choix méthodologiques. D'abord, comme indiqué précédemment (cf. chapitre 2), bien que la base de données contienne un large panel de motifs d'absence, nous ne nous intéresserons dans cette thèse qu'aux seules absences-maladie. Ensuite, pour l'établissement bancaire, nous ne retenons que les absences de l'année 2008. En effet, la qualité de leur recueil nous semble renforcé par la mise en place d'un système d'information unifié au cours de l'année 2007 : auparavant, les absences étaient enregistrées séparément dans les entités qui ont fusionné. Pour les années 2006 et 2007, les variables d'absence sont donc des combinaisons d'information : or, les habitudes de recueil des données, les périmètres d'analyse et la qualité des données étant hétérogènes, des simplifications ont parfois dû être pratiquées. Nous écartons donc ces informations pour le cœur de l'analyse. Le graphique 4.2 (page 138) rappelle la distribution des jours d'absence-maladie ainsi que leur concentration importante sur quelques individus dont on peut présumer que l'état de santé est très mauvais.

Absences-maladie de courte durée. En raison de cette forte concentration et pour les raisons évoquées dans le chapitre 2, nous choisissons également de limiter notre périmètre d'analyse en excluant les individus ayant plus de 45 jours d'absence. Comme nous l'avons évoqué plus haut, ceci nous permet de ne conserver que les situations dans lesquelles les individus ont effectivement la possibilité de choisir entre absence-maladie et présentisme. En effet, les absences de longue durée procèdent généralement de maladies plus lourdes et presque toujours incompatibles avec le travail. Ce choix rend également possible une analyse en classes d'âge plus équilibrée, puisque l'incidence de ces longues maladies est plus élevée chez les seniors. Le seuil de 45 jours nous a semblé particulièrement pertinent. Sur ce périmètre, lorsqu'une absence-maladie survient, elle

3. Rappelons que ce taux correspond au rapport entre le nombre moyen de jours d'absence par salarié et le nombre de jours normalement travaillés, soit 260 jours dans cette entreprise.

Graphique 4.2 – Distribution et concentration des jours d’absence (Banque en 2008)



ture en moyenne 5,4 jours dans l’établissement bancaire et 7,4 jours dans le centre hospitalier, soit approximativement le nombre de jours travaillés pendant une semaine dans chacun des deux contextes. Dans l’établissement bancaire, comme dans le centre hospitalier, tenir à l’écart de l’analyse les individus totalisant plus de 45 jours d’absence sur l’année revient à exclure environ 5% de la population ⁴ totalisant environ deux tiers de l’ensemble des jours d’absence-maladie ⁵.

Après cette modification, le périmètre d’analyse se réduit donc à 2 574 individus (contre 2 708 initialement), les salariés s’absentant en moyenne 3,7 jours dans l’année. Notre stratégie de réduction du périmètre d’analyse aux absences inférieures ou égales à 45 jours s’avère également utile pour la comparaison entre classes d’âge. Sur ce périmètre, la moyenne d’absence des seniors est similaire à celle des autres salariés (environ 3,7 jours). La prise en compte de l’ensemble des absences aurait produit un réel déséquilibre, puisque, dans ce cas, la moyenne d’absence des seniors (plus de 50 ans) est de 17,3 jours contre 8,1 pour les autres salariés. La moyenne des seniors serait alors alourdie par quelques individus victimes de maladies graves. **L’ensemble des statistiques qui vont désormais être produites pour l’établissement bancaire utilisent cette population comme assiette de calcul.**

4. 4,95% dans l’établissement bancaire et 5,02% dans le centre hospitalier.

5. 66,68% dans l’établissement bancaire et 60,6% dans le centre hospitalier.

1.1.2 Structure démographique

Outre le relevé individuel des absences, les premières variables présentes dans la base de données obtenue de l'établissement bancaire sont de type démographique (tableau 4.1). Elles produisent des informations sur les salariés eux-mêmes (âge, sexe, niveau de diplôme), mais également sur le ménage dans lequel ils vivent (situation familiale, nombre d'enfants). À ces variables démographiques classiques, nous ajoutons une variable binaire inférée des autres données d'absences : le fait d'avoir eu un congé maternité ou paternité dans l'année. Cette variable s'est montrée particulièrement intéressante pour l'analyse, et ce pour deux raisons. En premier lieu, elle explique une part importante de l'absence-maladie des femmes. Or, comme des arrêts liés aux pathologies de la grossesse n'ont pas tout à fait le même statut que les autres absences-maladie, il nous a semblé important d'intégrer cette variable pour isoler leur effet, qui est toujours significatif. L'effet pur de la variable SEXE n'en sera alors que plus pertinent. En second lieu, cette variable informe sur l'arrivée d'un enfant au foyer, avec la nouvelle organisation qu'elle implique.

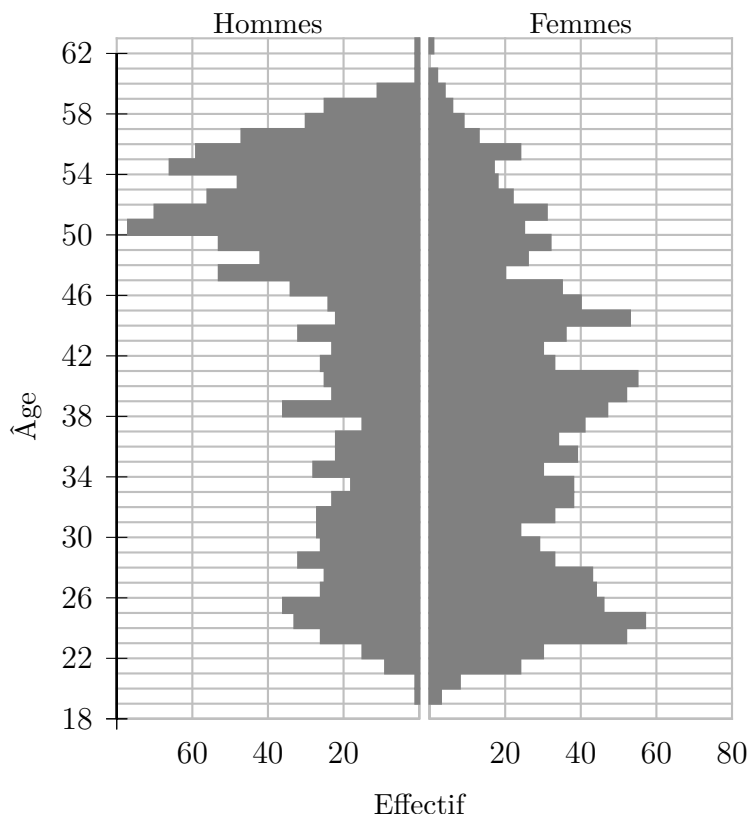
Tableau 4.1 – Variables démographiques (Banque en 2008)

	Hommes	Femmes	Ensemble
Âge moyen	42,86	37,08	39,99
Nombre moyen d'enfants	1,52	1,08	1,30
Salariés sans enfant	29,07%	41,11%	35,04%
Salariés avec jeune enfant (< 1 an)	2,93%	7,36%	5,13%
Salariés en couple	69,93%	58,26%	64,14%
Salariés titulaires d'une licence	13,34%	14,17%	13,75%
Salariés titulaires d'un master	15,19%	10,96%	13,09%
Nombre d'individus	1297	1277	2574

Comme le montre la pyramide des âges (graphique 4.3 page 140), bien que le nombre d'hommes et de femmes soit quasiment identique, la répartition dans les classes d'âge varie considérablement entre les sexes. Du côté des hommes, on note une importante surreprésentation des seniors (37,9% des hommes ont 50 ans ou plus), tandis que du côté des femmes, les jeunes générations sont mieux représentées (44,7% des femmes ont 35 ans ou moins). Ceci explique en partie les chiffres du tableau 4.1 : les femmes, dont

la moyenne d'âge est plus faible, ont également moins d'enfants et ont plus rarement une vie familiale établie (en couple).

Graphique 4.3 – Pyramide des âges (Banque en 2008)



Un premier croisement est également possible avec les données d'absence-maladie (inférieures ou égales à 45 jours). La moyenne d'absence des femmes est plus élevée, puisqu'elles s'arrêtent en moyenne 4,7 jours contre 2,7 pour les hommes. La part des femmes jamais absentes en 2008 est également plus faible que celle des hommes (51,1% contre 65,5%). Comme nous l'avons indiqué, les différences de niveau d'absence entre classes d'âge s'estompent : la moyenne d'absence des seniors est désormais légèrement inférieure à celle des autres salariés (3,5 contre 3,7 jours). Leur comportement est aussi marqué par un plus grand nombre de salariés jamais absents (62,2% contre 56,4% pour les autres), ce qui est une constante, quel que soit le périmètre choisi.

1.1.3 Structure de l'emploi

D'autres variables, que nous qualifierons d'organisationnelles (tableau 4.2), décrivent le poste occupé par un salarié ainsi que certains éléments contractuels. La majeure partie de la main d'œuvre (70,7%) est employée dans les fonctions commerciales. Deux métiers sont en relation quotidienne avec la clientèle : les conseillers commerciaux et les gestionnaires de clientèle. Ces deux métiers représentent la moitié de l'effectif (49,8%) et ont une moyenne d'absence légèrement supérieure à celle des autres salariés (4,3 jours contre 3,1 jours).

La population est répartie en trois grandes catégories hiérarchiques : les techniciens⁶ (33,1%), les agents de maîtrise⁷ (42,9%) et les cadres. Près d'un quart de la population (24%) entre dans cette dernière catégorie. La moyenne d'absence des cadres est inférieure à celle des non-cadres (2,4 jours contre 4,1 jours). Dans cette frange de la population, la plupart des individus (67,6%) sont en charge d'une équipe de travail, en particulier dans le contexte des agences bancaires (42,3% des cadres sont directeurs d'agence).

Tableau 4.2 – Variables organisationnelles (Banque en 2008)

	Hommes	Femmes	Ensemble
Ancienneté moyenne	17,87	12,62	15,27
Salariés du réseau commercial	71,01%	70,40%	70,71%
Salariés techniciens	21,90%	44,56%	33,14%
Salariés cadres	33,92%	13,86%	23,97%
Salariés ayant une responsabilité d'équipe	23,59%	8,69%	16,20%
Salariés en CDD	7,48%	16,37%	11,89%
Salariés sans indemnisation	10,10%	14,88%	12,47%
Salariés ayant connu une mobilité	9,56%	7,52%	8,55%
Nombre d'individus	1297	1277	2574

Les salariés sous contrat à durée déterminée représentent 11,9% de la population. Il sont le plus souvent embauchés sur des postes de conseillers commerciaux (dans 78,4% des

6. Essentiellement les conseillers commerciaux et les chargés de gestion bancaire en back-office.

7. Surtout les gestionnaires de clientèle et les chargés d'affaires pour les professionnels.

cas) ou de secrétariat (dans 9,5% des cas). Ces postes en CDD, sont majoritairement occupés par des femmes (68,3%).

L'entreprise applique également une règle particulière d'indemnisation de l'absence-maladie pour les salariés ayant plus d'un an d'ancienneté. En effet, dans le cadre classique de l'assurance-maladie, un salarié absent pour raison de santé bénéficie d'une indemnisation de l'absence-maladie égale à 50% du salaire de base après trois jours de carence. Cependant, pour les salariés ayant plus d'un an d'ancienneté, l'entreprise perçoit elle-même les indemnités journalières et les complète (subrogation) : ces salariés ne supportent pas de délai de carence et touchent 100% de leur salaire de base. Dans la population, 12,5% des salariés ne bénéficient pas de ce régime. Parmi eux, 62,3% sont embauchés en CDD.

Enfin, au cours de la fusion, certains salariés (8,55% de la population) ont connu une mobilité professionnelle ayant nécessité une reconversion complète, par exemple par le passage d'un métier administratif à un métier commercial. Ce type de mobilité, qui se manifeste par un changement de « filière » d'emploi, est susceptible d'avoir des implications sur les comportements d'absence et/ou de présentéisme.

1.2 Centre hospitalier

Le deuxième terrain de travail est un centre hospitalier régional du Nord-Pas-de-Calais (1500 lits). Comme tout établissement public hospitalier, son fonctionnement général est régi par la loi de réforme de l'hôpital datant de 2009⁸. À la tête du centre hospitalier se trouve un directeur général (nommé par décret) ayant autorité sur l'ensemble du personnel. Il préside un directoire regroupant des directeurs et des praticiens dont chacun un titulaire d'un portefeuille d'activité (dont les ressources humaines). Un conseil de surveillance fixe la stratégie générale de l'établissement et détient un pouvoir de contrôle sur sa gestion. Enfin, une commission médicale représente le corps des médecins : son président (également vice-président du directoire) établit le projet médical de l'établissement avec le directeur général.

L'établissement est subdivisé en pôles qui disposent chacun d'une certaine autonomie : le chef de pôle (un praticien hospitalier) a autorité en matière médicale, mais aussi administrative et technique. Dans cet hôpital, environ un tiers de la population (30,2%)

8. Loi n° 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires (loi HPST).

est employé dans le pôle administratif ou dans le pôle logistique. Pour les pôles de soins, la gériatrie, la maternité et les urgences sont les pôles les plus importants en termes d'effectifs (respectivement 10,8%, 8,3% et 7,7% du personnel en 2010). Ce fonctionnement en pôles d'activité a été mis en place dans le cadre du plan Hôpital 2007 qui a également introduit la tarification à l'activité.

L'originalité de la base de données que nous avons obtenue de cet établissement hospitalier est sans conteste sa profondeur temporelle. Elle porte en effet sur 34 années au total, de 1978 à 2012. Notre analyse exploitera plus spécifiquement une période de 20 ans comprise entre 1991 et 2010. Avant cette période, les données nous ont semblé être de plus faible qualité⁹ et après cette période, les informations, notamment les données d'absence, demandaient encore à être consolidées. Notons également que la base de données n'inclut pas les médecins du centre hospitalier : seul le personnel paramédical (infirmières, sage-femmes, etc.), médico-technique (manipulateurs en radiologie, préparateurs en pharmacie, etc.) et administratif est concerné.

La base de données nous est parvenue sous la forme d'un extrait brut du système d'information et a nécessité un travail de réagencement des données. Au cours de celui-ci, l'une des règles de décision était de conserver pour chaque année la situation d'un individu au 31 décembre, comme c'est le cas dans l'établissement bancaire. L'une de nos premières actions a également été d'exclure pour chaque année l'ensemble des salariés ayant le statut spécifique de « malade de longue durée » (invalidité).

1.2.1 Absences-maladie et autres absences

Ensemble des absences. La base de données comprend notamment les enregistrements relatifs à trois motifs d'absence distincts : la maladie, les accidents du travail ou maladies professionnelles et la maternité. En 2010, le nombre moyen de jours d'absence pour cause de maladie était de 7,9 jours dans le centre hospitalier, mais comme l'indique le graphique 4.4 présentant l'évolution des absences sur la période 1987-2012, le niveau d'absence était relativement faible cette année-là (9,6 jours en moyenne sur la période).

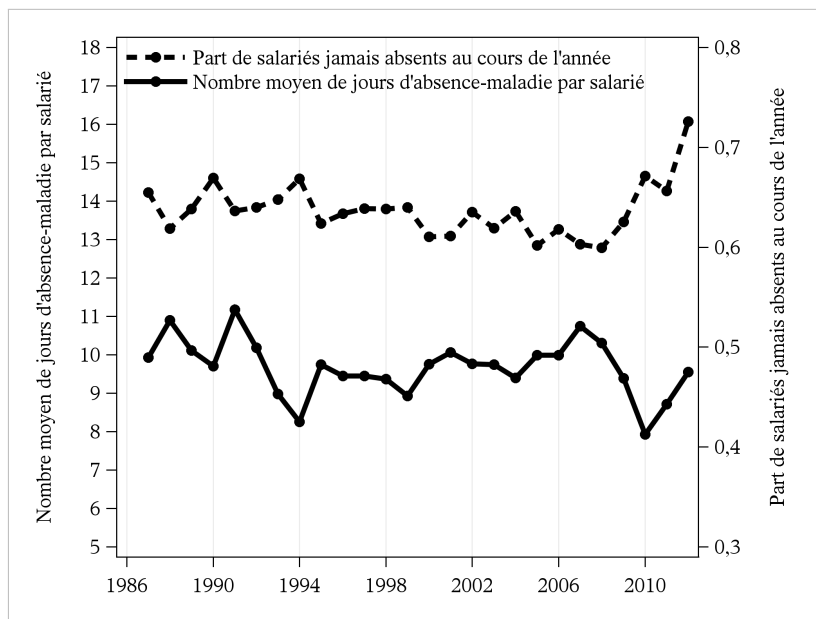
Peu de chiffres sont disponibles sur le niveau d'absence dans l'ensemble de la fonction publique hospitalière, mais d'après une enquête sur les données sociales hospitalières, la

9. Ceci peut être dû à l'évolution des méthodes de recueil ou à l'inclusion de nouveaux motifs ou périmètres d'absence.

moyenne d'absence pour maladie y était de 9,4 jours en 2008 et représentait 45,3% des absences¹⁰. Sur la même période, la moyenne d'absence était de 10,3 jours au centre hospitalier, soit légèrement plus que dans le reste de la fonction publique hospitalière. La maladie représentait 43,5% de l'ensemble des absences (20,3% pour la maternité contre 21,9% dans la fonction publique hospitalière)¹¹. Le niveau des absences pour cause de maternité était comparable (4,9 jours en moyenne au centre hospitalier contre 4,5 jours dans l'ensemble de la fonction publique).

Au centre hospitalier la situation des absences se caractérise également par une part élevée de salariés n'ayant aucune absence pendant une année donnée (graphique 4.4). En 2010, 67,1% des salariés n'ont enregistré aucune absence. Sur la période 1987-2012, la part moyenne des salariés jamais absents est de 63,7%. Cette part est toujours plus élevée pour les hommes que pour les femmes (73,7% en moyenne pour les hommes contre 61,1% pour les femmes).

Graphique 4.4 – Évolution des absences-maladie (Hôpital de 1987 à 2012)



10. Une synthèse des résultats de cette enquête sont disponibles en ligne à l'adresse suivante : <http://www.sante.gouv.fr/donnees-sociales-hospitalieres-2008.html>.

11. Notons que dans l'enquête évoquée, l'assiette totale des absences est composée comme suit : absences-maladie, accidents du travail et maladies professionnelles, absences maternité et longues maladies. La statistique concernant le centre hospitalier adopte la même structure.

Il est intéressant d'ouvrir une parenthèse concernant l'évolution des absences sur cette longue période. Parmi les éléments traditionnellement avancés dans la littérature pour expliquer le niveau d'absence, on peut citer pêle-mêle l'influence des politiques managériales (Grinyer & Singleton, 2000; Rocheleau & Renaud, 2003), l'incidence des maladies épidémiques (DARES, 2013), voire un taux de chômage élevé (Askildsen et al., 2005). Certaines circonstances peuvent en effet constituer des clefs de lecture pour les variations les plus marquées du graphique 4.4. Mêmes si certaines coïncidences paraissent frappantes, les remarques qui suivent n'ont pas statut de preuve et sont purement indicatives.

La courbe de niveau d'absence est d'abord marquée par deux pics en 1991 et en 2007 auxquels succèdent trois années successives de baisse du niveau d'absence. Sur les deux décennies, les années 1991 et 2007 correspondent également à deux réformes majeures de l'hôpital public. Le pic d'absence de 1991 coïncide avec la mise en place d'une grande réforme hospitalière instaurant une planification plus contraignante et visant à une réduction des coûts de fonctionnement des hôpitaux¹². De même l'année 2007, voit l'instauration de la tarification à l'activité (Plan Hôpital 2007), qui a de nombreuses résurgences organisationnelles et professionnelles, par exemple liées à la complexité des nouveaux outils mis en place (Pepin & Moisdon, 2010).

D'autre part, les creux enregistrés pour les années 1994 et 2010 coïncident avec des années où le taux de chômage était très élevé (9,3% en 1994 et 9% en 2010)¹³. Néanmoins, le chômage atteint son paroxysme lors de l'année 1997 (taux annuel de 9,5%), qui n'est pas spécialement marquée sur la courbe des absences.

Par ailleurs, les années 1994, 2004 et 2010 sont également des années où l'épidémie de grippe a été peu virulente dans le Nord-Pas-de-Calais¹⁴. En revanche, en 2001, l'épidémie de gastro-entérite a été particulièrement forte¹⁵.

Absences-maladie de courte durée. Dans le chapitre 2, après avoir exploré la distribution et la concentration des absences du centre hospitalier, nous avons déjà expliqué pourquoi il nous semblait nécessaire d'exclure de l'analyse les individus pré-

12. Loi n° 91-748 du 31 juillet 1991 portant réforme hospitalière.

13. Source : INSEE.

14. Plus précisément 37 161 cas en 1994, 62 504 en 2004 et 91 533 en 2010 contre une moyenne de 217 372 sur la période 1987-2012. Source : réseau Sentinelles de l'INSERM (<http://websenti.u707.jussieu.fr/sentiweb/>).

15. 428 057 cas contre 301 804 en moyenne sur la période.

sentant plus de 45 jours d'absence. Pour 2010, année sur laquelle portera l'analyse en coupe transversale, la population étudiée comprendra donc 4 090 individus, au lieu de 4 306 initialement : 5,02% des individus les plus malades sont exclus, cumulant 60,6% de l'ensemble des jours d'absence. De même que précédemment, il est intéressant de constater que le niveau d'absence des seniors tend à s'égaliser comparativement aux autres classes d'âge : de 11,9 jours (contre 6,9 jours pour les autres) sur le périmètre large il passe à 4,1 jours (contre 3,1) sur les absences de courte-durée. **L'ensemble des statistiques descriptives à venir dans cette section seront calculées sur ce périmètre.**

1.2.2 Structure démographique

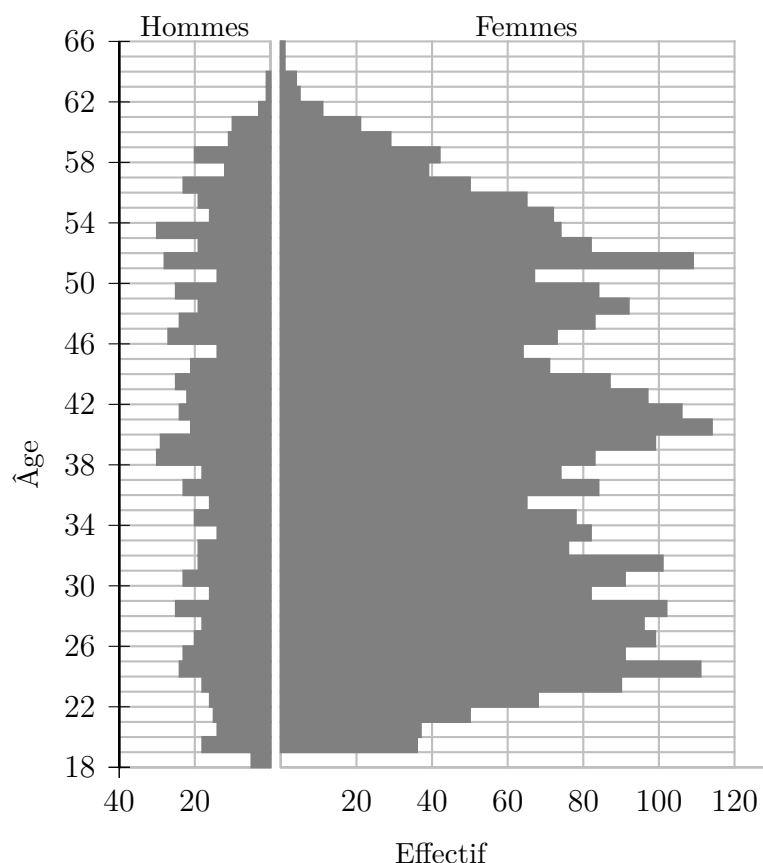
Ainsi que pour l'établissement bancaire, les variables explicatives disponibles peuvent être subdivisées en deux catégories : les variables démographiques et les variables organisationnelles. Les variables entrant dans la première catégorie sont sensiblement identiques à celles de l'établissement bancaire. Le tableau 4.3, en plus de présenter quelques statistiques descriptives pour 2010, donne une idée de l'évolution de ces variables au cours des dernières années.

Tableau 4.3 – Variables démographiques (Hôpital)

	1995	2000	2005	2010
Âge moyen	34,62	36,32	38,63	38,78
Nombre moyen d'enfants	1,23	1,32	1,43	1,31
Salariés sans enfant	39,22%	36,12%	31,22%	34,57%
Salariés avec enfant jeune (< 1 an)	4,01%	3,63%	3,76%	4,18%
Salariés ayant des enfants à charge	24,36%	24,65%	23,05%	22,05%
Salariés en couple	59,95%	59,00%	61,75%	59,07%
Nombre d'individus	3715	4073	3940	4090
dont part d'hommes	22,5%	22,2%	21,6%	20,8%

On constate que la féminisation du personnel n'est pas un fait nouveau, bien qu'il se soit accentué ces dernières années. Les hommes représentent approximativement un cinquième de l'effectif, ce qui est comparable au niveau national de la fonction publique hospitalière. Au 31 décembre 2010, celle-ci comptait approximativement un

Graphique 4.5 – Pyramide des âges (Hôpital en 2010)



million d'agents (hors médecins), dont 80,2% de femmes¹⁶. La pyramide des âges du centre hospitalier (graphique 4.5) illustre bien cette forte féminisation et procure une autre information : la répartition des âges est relativement homogène pour les deux sexes. Les moyennes d'âge des hommes et des femmes sont d'ailleurs similaires. Bien que l'âge moyen global présente une légère hausse sur les 20 dernières années, les seniors ne sont pas présents en surnombre (20,5% de la population en 2010 contre 26,6% dans la fonction publique hospitalière) mais sont légèrement plus nombreux chez les hommes (23,5% contre 19,7% chez les femmes). En première analyse, les salariés de l'entreprise ont pris en moyenne 3,3 jours d'absence-maladie. Ce chiffre est légèrement plus élevé pour les seniors (4,1 jours d'absence pour les 50 ans et plus contre 3,1 pour les autres) ainsi que pour les femmes (3,5 jours contre 2,4 pour les hommes).

16. L'ensemble des statistiques portant sur la fonction publique hospitalière produites dans ce chapitre peuvent être trouvées dans le *Rapport annuel sur l'état de la fonction publique - Édition 2012* publié par la Direction Générale de l'Administration et de la Fonction Publique (<http://www.fonction-publique.gouv.fr/publications/rapport-annuel-0>).

1.2.3 *Structure de l'emploi*

L'autre catégorie de variables décrit les individus en fonction des spécificités de leur poste de travail et des éléments principaux de leur contrat de travail (tableau 4.4 page 149). La plus grande partie du personnel est consacrée au cœur du métier hospitalier, puisque 69% des salariés occupent un emploi lié aux soins médicaux (68,7% dans l'ensemble de la fonction publique hospitalière). Ce personnel soignant est réparti en plusieurs grands corps de métiers (cf. graphique 4.6 page 149)¹⁷, parmi lesquels trois se distinguent. Premièrement, les infirmiers (29,7% de la population en 2010) constituent la clef de voûte de la filière soignante : leur métier implique la réalisation de gestes médicaux et une relation spécifique aux patients passant par un accompagnement quotidien. Deuxièmement, les aide-soignants (20,5% de la population en 2010) sont les assistants directs des infirmiers et, sous leur responsabilité, réalisent des soins-spécifiques (e.g. toilette des patients) ainsi que certaines tâches d'entretien et de désinfection. Troisièmement, les agents de service hospitaliers qualifiés (ASH)¹⁸ représentent 13,2% de la population en 2010. Leur charge est principalement liée à l'entretien et au nettoyage des locaux, voire à l'activité hôtelière. Ces trois grands corps collaborent avec des métiers plus spécifiques liés à la rééducation (ergothérapeute, kinésithérapeute, diététicien, etc.), aux activités médico-techniques (manipulateurs de radiologie, préparateurs en pharmacie, etc.) ou à l'accompagnement psychologique.

Des cadres de santé gèrent ces équipes, soit au niveau local (unité de soins), soit au niveau du pôle (cadre supérieur de santé). Leurs tâches impliquent une responsabilité d'équipe (management des soins) : ils organisent le travail des services et veillent au suivi des patients ainsi qu'à la qualité des soins. Avec leurs adjoints et les cadres des fonctions administratives, ils représentent approximativement 4% des effectifs.

Outre leur métier, les salariés du centre hospitalier peuvent être répartis selon leur statut (titulaire ou stagiaire, CDI, CDD). Le centre hospitalier que nous étudions est marqué par un grand nombre de salariés contractuels, en particulier sous contrat à durée déterminée (17,8% de la population en 2010). Le tableau 4.4 (page 149) montre une forte fluctuation de la part de CDD suivant l'année de référence. On peut penser que le recrutement d'agents contractuels est fortement lié aux besoins occasionnels et

17. Les informations sur les métiers proposés dans cette section sont disponibles en ligne à l'adresse suivante : <http://www.metiers-fonctionpubliquehospitaliere.sante.gouv.fr>.

18. Aussi appelés agents de bio-nettoyage.

Graphique 4.6 – Évolution de la structure des emplois (Hôpital de 1987 à 2012)

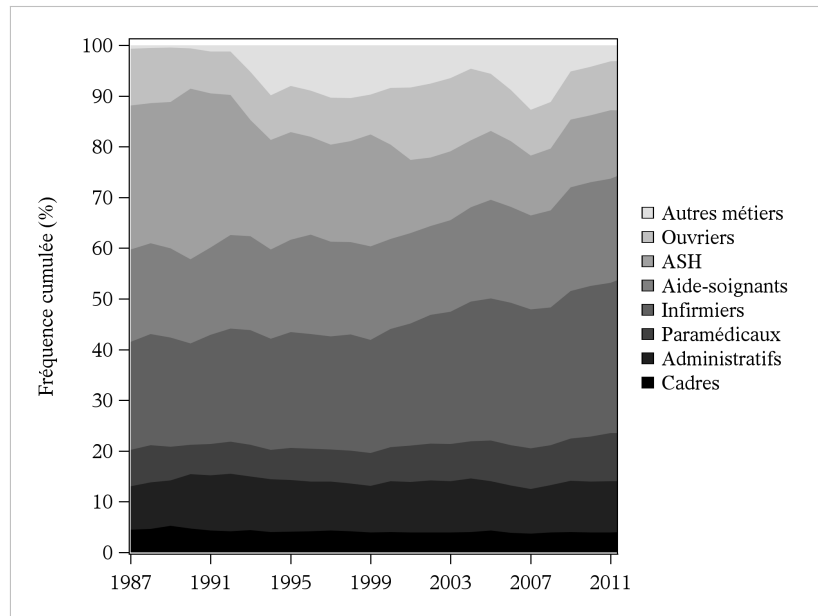


Tableau 4.4 – Variables organisationnelles (Hôpital)

	1995	2000	2005	2010
Ancienneté moyenne (entreprise)	10,04	11,36	13,30	13,45
Ancienneté moyenne (au poste)	3,82	5,02	5,60	4,31
Personnel administratifs	11,06%	10,88%	10,79%	11,25%
Personnel de soins	67,08%	64,77%	66,78%	69,00%
Salariés titulaires	68,13%	65,36%	72,66%	65,33%
Salariés cadres	4,09%	4,00%	4,29%	3,96%
Salariés en CDD	17,66%	21,19%	14,37%	17,78%
Salariés ayant connu une mobilité	11,92%	10,12%	10,81%	12,10%
Nombre d'individus	3715	4073	3940	4090
dont hommes	838	921	873	875

aux ressources du centre hospitalier. Parmi les salariés en CDD, la plupart occupent un poste d'ASH (42,1%), d'aide-soignant (14,3%) ou d'administration (13,62%). La part des salariés titulaires (c'est-à-dire agents de la fonction publique) est relativement faible dans ce centre hospitalier : 65,3% en 2010 contre 81,7% dans le reste de la fonction publique hospitalière¹⁹.

Enfin, une variable liée à la mobilité indique si l'individu a connu une mobilité professionnelle au cours de l'année qui précède l'analyse, c'est-à-dire, si le salarié, déjà en poste au centre hospitalier, a changé récemment d'affectation : 12,1% des salariés sont dans ce cas en 2010. Les régressions intégreront par ailleurs une variable indiquant si l'ancienneté dans l'emploi est inférieure à une année, que le salarié ait déjà été en poste au centre hospitalier ou qu'il s'agisse de son premier emploi dans l'établissement.

2 Régressions en coupe transversale

Pour les deux organisations qui viennent d'être présentées, les seules statistiques descriptives fournies jusqu'ici portaient sur l'absence-maladie des salariés ou la part de salariés jamais absents. Mais comme nous avons eu l'occasion de l'expliquer, une statistique sur la catégorie des salariés jamais absents pendant l'année de référence, n'est pas directement une statistique de présentéisme. En effet, dans cette catégorie, des individus en très bonne santé (jamais malades) cohabitent avec des salariés qui ont l'habitude de travailler étant malade.

Jusqu'ici, rien n'a donc encore été dit sur le présentéisme au travail. Pour les raisons que nous avons copieusement développées dans les chapitres précédents, la régression d'un modèle binomial négatif à inflation de zéros (ZINB) est l'option que nous proposons pour en savoir plus sur ce comportement. La tâche essentielle de ce chapitre est de l'appliquer à ces deux bases de données.

Dans cette section, après avoir fourni quelques éléments sur les méthodes de spécification des modèles, nous présentons les résultats d'estimation. Même si les quatre modèles de comptage décrit plus haut vont être évoqués, le modèle ZINB retiendra principalement notre attention. Nous tenterons de rendre plus évidente la compréhension et l'interprétation des résultats de ce modèle, en nous dotant de quelques clefs

19. Outre les salariés titulaires (65,3% de la population) et les salariés en CDD (17,8%), le centre hospitalier emploie une partie du personnel en CDI de droit privé (soit 16,9% de la population).

de lecture de la sortie économétrique. Dans la section suivante, nous fournirons davantage de détails sur ce qu'il est possible d'inférer de nos résultats, en particulier les probabilités individuelles de présentéisme.

2.1 Spécification des modèles

Plusieurs critères ont présidé au choix des variables explicatives qui sont intégrées dans les modèles ci-après. La condition préalable à la spécification des modèles était de mieux connaître les comportements d'absence et de présence au travail, grâce à la littérature et en mettant à profit notre propre expérience. Au fur et à mesure de l'avancée de notre recherche, deux profils-types, correspondant aux deux phénomènes étudiés, l'absence-maladie et le présentéisme ont peu à peu émergé. Quelques variables susceptibles d'expliquer le niveau d'absence se détachent par exemple dans la littérature. Les femmes, les seniors, les parents, les salariés indemnisés ou encore qui exercent un métier pénible, semblent davantage susceptibles d'être absents. De même, certaines variables sont plus évocatrices de présentéisme : le niveau de responsabilité, la sécurité dans l'emploi, mais également l'âge ou la charge familiale.

Sans nous limiter exclusivement à ces variables explicatives, nous avons ensuite cherché celles qui s'intégraient le mieux au modèle et dont la combinaison produisaient le meilleur ajustement avec la variable dépendante. En d'autres termes nous avons privilégié les plus significatives et celles qui, associées, maximisaient la vraisemblance. Cette méthode d'estimation, classique des modèles de comptage, est brièvement décrite en annexe (annexe 2) : les fonctions de log-vraisemblance des modèles étudiés y sont également indiquées. La valeur de la log-vraisemblance est chiffrée à chaque régression : plus cette log-vraisemblance ($\mathcal{L}$) est élevée (donc proche de zéro, puisqu'elle est toujours négative), mieux le modèle est ajusté. Dans la mesure où la log-vraisemblance augmente systématiquement lorsque le nombre de paramètres augmente, on peut également s'appuyer sur un autre indicateur qui privilégie à la fois le modèle le mieux ajusté et le plus simple (celui qui comporte le moins de paramètres). Il s'agit du *Akaike Information Criterion* ($AIC = -2\mathcal{L} + 2k$, avec k le nombre de paramètres) : plus ce dernier indicateur est faible, meilleur est le modèle.

Les modèles présentés ne comprennent donc que les variables significatives à un seuil de 5%. Bien qu'une variable non significative ait un intérêt en soi (par exemple la non-significativité de la variable FEMME dans l'équation de comptage pour l'établis-

ment hospitalier), la lecture et l'interprétation des modèles nous ont parues plus aisées lorsqu'elles n'étaient pas incluses.

Une exception a cependant été faite pour la variable `EXPOSITION`, qui est incluse quelle que soit sa significativité. Il s'agit d'une variable technique dont l'interprétation n'a pas d'intérêt en soi : elle rend possible une analyse à temps d'exposition égal. En effet, l'exposition au risque d'absence diffère d'un individu à l'autre, pour deux raisons. D'une part, un salarié peut travailler à temps partiel ou à temps complet. Avoir trois jours d'absence dans une semaine donnée n'a pas la même signification pour un salarié qui travaille toute la semaine que pour quelqu'un ne travaillant que trois jours. D'autre part, parmi plusieurs salariés à temps plein, le contrat de travail de certains peut n'avoir commencé qu'en cours d'année ou s'être arrêté avant la fin de l'année. Pour tenir compte de ces différences, la variable `EXPOSITION` est donc construite de la façon suivante : elle est le produit entre la proportion de temps de travail sur une semaine et la proportion de présence dans l'entreprise sur l'année. Autrement dit, un individu travaillant à temps plein et présent toute l'année aura une valeur de 1. Un salarié à mi-temps obtiendra une valeur de 0,5 ($= 0,5 \times 1$) et un salarié à mi-temps dont le contrat a débuté le 1^{er} juillet (milieu de l'année) une valeur de 0,25 ($= 0,5 \times 0,5$).

2.2 Résultats des régressions

Les résultats des régressions des quatre modèles de comptage sont présentés pour l'établissement bancaire et le centre hospitalier sur les deux pages suivantes (respectivement page 153 et page 154). Pour chaque modèle, plusieurs informations y figurent. D'abord, les éléments classiques d'une régression économétrique : le coefficient de chaque paramètre, sa significativité (à 0,1%, 1% ou 5%) et l'erreur-type, c'est-à-dire l'écart-type de chaque coefficient. Ensuite, on trouve deux indicateurs d'ajustement propres à la méthode d'estimation (la log-vraisemblance et l'AIC) ainsi qu'une donnée spécifique au modèle binomial négatif et à son extension *zero-inflated* : l'estimation du paramètre α , noté `DISPERSION`, qui caractérise la forme de la distribution du facteur d'hétérogénéité (suivant une loi Gamma). Ces informations seront largement analysées et commentées dans les pages qui suivent.

Tableau 4.5 – Estimation des modèles standards (Banque)

Paramètre	Poisson		Binomial négatif	
	Coefficient	Err.-type	Coefficient	Err.-type
CONSTANTE	0,257 **	0,094	0,284	0,345
FEMME	0,486 ***	0,023	0,484 ***	0,096
AGE	0,003 *	0,001	0,007	0,005
JEUNE ENFANT (< 1 an)	0,736 ***	0,033	0,647 **	0,207
CDD	-0,657 ***	0,060	-0,716 ***	0,210
HIÉRARCHIE	-0,209 ***	0,009	-0,246 ***	0,036
EXPOSITION	1,686 ***	0,075	1,673 ***	0,227
DISPERSION			4,671 ***	0,187
Log-vraisemblance (AIC)	-13000 (26014)		-5066 (10147)	

Significativité : <0,1%=*** <1%=** <5%=*

Tableau 4.6 – Estimation des modèles *zero-inflated* (Banque)

Paramètre	ZIP		ZINB	
	Coefficient	Err.-type	Coefficient	Err.-type
<i>Équation de comptage :</i>				
CONSTANTE	1,410 ***	0,100	1,138 ***	0,315
FEMME	0,241 ***	0,023	0,272 ***	0,079
AGE	0,012 ***	0,001	0,016 ***	0,004
JEUNE ENFANT (< 1 an)	0,753 ***	0,034	0,816 ***	0,150
CDD	-0,366 ***	0,063	-0,446 *	0,181
HIÉRARCHIE	-0,093 ***	0,009	-0,119 ***	0,031
EXPOSITION	0,533 ***	0,079	0,540 *	0,230
DISPERSION			1,262 ***	0,106
<i>Équation d'inflation :</i>				
CONSTANTE	0,690 *	0,317	0,015	0,360
FEMME	-0,455 ***	0,089	-0,445 ***	0,110
AGE	0,027 ***	0,005	0,037 ***	0,006
UN SEUL ENFANT	-0,361 ***	0,107	-0,412 **	0,134
BAC+3/+5	0,474 ***	0,105	0,532 ***	0,124
SANS INDEMNISATION	0,605 **	0,430	0,682 **	0,214
RESPONSABILITÉ	0,599 ***	0,126	0,630 ***	0,143
MOBILITÉ	-0,307 *	0,149	-0,417 *	0,190
EXPOSITION	-1,623 ***	0,242	-1,731 ***	0,280
Log-vraisemblance (AIC)	-7459 (14949)		-4943 (9919)	

Significativité : <0,1%=*** <1%=** <5%=*

Tableau 4.7 – Estimation des modèles standards (Hôpital)

Paramètre	Poisson		Binomial négatif	
	Coefficient	Err.-type	Coefficient	Err.-type
CONSTANTE	-0,355 ***	0,048	-0,821 ***	0,239
AGE	0,015 ***	0,001	0,020 ***	0,005
JEUNE ENFANT (< 1 an)	0,993 ***	0,031	1,124 ***	0,243
ASH	0,319 ***	0,028	0,172	0,147
AIDE-SOIGNANT	0,570 ***	0,019	0,590 ***	0,122
EXPOSITION	0,804 ***	0,042	1,127 ***	0,207
DISPERSION			9,175 ***	0,327
Log-vraisemblance (AIC)	-23339 (46691)		-6554 (13123)	
Significativité : <0,1%=*** <1%=** <5%=*				

Tableau 4.8 – Estimation des modèles *zero-inflated* (Hôpital)

Paramètre	ZIP		ZINB	
	Coefficient	Err.-type	Coefficient	Err.-type
<i>Équation de comptage :</i>				
CONSTANTE	1,956 ***	0,054	1,658 ***	0,217
AGE	0,010 ***	0,001	0,013 ***	0,003
JEUNE ENFANT (< 1 an)	0,486 ***	0,031	0,628 ***	0,140
ASH	0,150 ***	0,028	0,225 *	0,109
AIDE-SOIGNANT	0,304 ***	0,019	0,380 ***	0,080
EXPOSITION	-0,108 *	0,047	-0,106	0,191
DISPERSION			1,331 ***	0,100
<i>Équation d'inflation :</i>				
CONSTANTE	1,121 ***	0,266	0,837 **	0,290
FEMME	-0,431 ***	0,093	-0,447 ***	0,100
AGE	0,019 ***	0,004	0,022 ***	0,005
SANS ENFANTS	0,387 ***	0,095	0,402 ***	0,103
INFIRMIER	0,321 ***	0,083	0,319 ***	0,091
CADRE	1,000 ***	0,222	1,029 ***	0,232
CDD	0,548 ***	0,145	0,585 ***	0,154
TITULAIRE	-0,305 **	0,113	-0,318 *	0,123
ANCIENNETÉ FAIBLE (< 1 an)	0,364 **	0,116	0,388 **	0,123
EXPOSITION	-0,960 ***	0,172	-1,039 ***	0,185
Log-vraisemblance (AIC)	-10186 (20405)		-6362 (12757)	
Significativité : <0,1%=*** <1%=** <5%=*				

Notons que les sorties économétriques d'une régression à inflation de zéros sont toujours divisées en deux vecteurs de variables. La première partie, appelée équation de comptage, mesure l'espérance conditionnelle d'absence et ses facteurs (à l'instar d'un modèle standard), tandis que la seconde partie, appelée équation d'inflation caractérise le phénomène de présentéisme, autrement dit la probabilité d'appartenir au groupe d'individus ayant systématiquement une valeur nulle.

2.3 Éléments d'interprétation

Se pose à présent la question de la lecture et de l'interprétation des coefficients contenus dans les sorties économétriques. Leur signe et leur poids relatif constituent déjà de bons éléments d'appréciation. Partons du cas du centre hospitalier (tableau 4.8 page 154) : dans l'équation d'inflation, le coefficient de la variable FEMME est de -0,447 et celui de la variable CADRE de 1,029. Le fait d'être une femme a donc un effet négatif (au sens statistique) sur le présentéisme, c'est-à-dire que les femmes sont moins susceptibles d'être présentéistes, tandis que le fait d'être cadre a un effet positif. En valeur absolue, l'incidence propre au statut de cadre est plus importante que celle du sexe. En d'autres termes, le poids de la variable CADRE dans l'équation du présentéisme est plus élevé que celui de la variable FEMME.

D'autre part, même si la régression estime un coefficient pour toute variable intégrée dans le modèle, chacune n'est pas nécessairement significative. Notre décision de n'inclure que les variables significatives dispense le lecteur de comparer la significativité des variables²⁰. Lorsqu'on évoque la significativité, c'est la question du risque d'erreur qui est posée. Nous n'avons conservé ici que les variables pour lesquelles ce risque était inférieur à 5%.

Ceci constitue le premier niveau de lecture, commun aux deux équations du modèle. Mais quelques clefs de lecture supplémentaires autorisent une approche plus précise des coefficients, suivant le vecteur de variables explicatives duquel ils proviennent.

20. Les *p-values* sont ici obtenues sur la base d'un test de Wald, c'est-à-dire le rapport élevé au carré entre le coefficient et son erreur-type, puis lecture du niveau de significativité du résultat dans une table de χ^2 .

2.3.1 Équation de comptage

Les coefficients de l'équation de comptage expriment l'influence des variables explicatives incluses sur l'espérance d'absence dans le régime normal (Lawal, 2012), c'est-à-dire le groupe des non-présentéistes, que nous appelons espérance conditionnelle d'absence. Il s'agit de l'espérance λ du modèle standard qui, pour un individu donné, s'écrit :

$$\lambda_i = \exp(X_i^T \beta)$$

Si, dans le cas du centre hospitalier, on exprime cette fois λ en fonction des coefficients obtenus dans la régression, on obtient :

$$\begin{aligned}\lambda &= \exp(1,658 + 0,013 \times \text{AGE} + 0,628 \times \text{ENFANT} + \dots - 0,106 \times \text{EXPOSITION}) \\ \lambda &= e^{1,658} \times e^{0,013 \times \text{AGE}} \times e^{0,628 \times \text{ENFANT}} \times \dots \times e^{-0,106 \times \text{EXPOSITION}}\end{aligned}$$

Ce qui est mesuré ici est donc l'effet d'une modification de l'une des variables sur l'espérance conditionnelle d'absence et non sur l'absence elle-même. Dans l'ensemble de cette thèse, les variables incluses dans l'équation de comptage sont soit binaires, soit quantitatives discrètes. Dans le cas d'une variable muette (binaire), une sous-population est observée comparativement au reste de la population. Par exemple, dans l'équation de comptage du centre hospitalier, le coefficient de la variable AIDE-SOIGNANT marque la variation dans l'espérance d'absence pour un aide-soignant en comparaison à la référence²¹. Ici, l'espérance conditionnelle d'absence d'un aide-soignant est de 46% plus élevée que pour les autres salariés, sachant que $e^{0,380} = 1,46$. Puisqu'il s'agit de l'espérance conditionnelle d'absence, ce constat n'est vrai que dans le régime normal d'absence.

Si l'on s'intéresse à une variable quantitative, par exemple la variable AGE, l'interprétation est quelque peu différente. Le coefficient représente l'évolution de l'espérance conditionnelle d'absence résultant de l'augmentation marginale de l'âge. Autrement dit, quel est l'effet d'une année d'âge supplémentaire sur l'espérance λ ? Dans le centre hospitalier, un an de plus correspond à une augmentation de 1,3% de l'espérance conditionnelle d'absence ($e^{0,013} = 1,013$). Ainsi, pour un écart d'âge de 20 ans entre deux

21. Autrement dit, en comparaison à un salarié qui n'exerce pas ce métier et dont le métier n'est pas lui-même pris en compte dans l'équation.

individus, λ sera multipliée par 1,3 ($= e^{0,013 \times 20}$). Du fait de l'âge, un « salarié moyen » de 50 ans aurait donc une espérance d'absence de 30% plus élevée qu'un individu de 30 ans, toutes choses égales par ailleurs.

2.3.2 Équation d'inflation

L'équation d'inflation s'interprète quant à elle de la même manière qu'une régression logistique. Il s'agit ici d'évaluer l'influence d'une variable explicative sur la propension d'un salarié à appartenir au régime de valeurs nulles structurelles, c'est-à-dire au groupe des présentéistes. On peut par exemple calculer un rapport des chances pour la variable CDD (coefficient égal à 0,585) : il est de 1,79 ($= e^{0,585}$). Un rapport des chances supérieur à 1 (donc un coefficient positif) implique une augmentation des chances, tandis qu'un rapport inférieur à 1 (et un coefficient négatif) indique une diminution des chances. Un salarié ayant un contrat précaire a donc 1,79 fois plus de chances de venir travailler étant malade, en comparaison de ceux qui ne sont pas dans cette situation. Cependant, cela ne signifie pas que la probabilité d'un salarié en CDD soit 1,79 fois plus élevée que celle d'un autre salarié. La valeur de la probabilité dépend en effet du statut des autres variables, ce qui complique l'interprétation directe. Prenons deux exemples concrets au centre hospitalier.

Considérons d'une part la situation d'une aide-soignante de 30 ans ayant un enfant, non-titulaire de son poste et dont l'ancienneté est supérieure à un an. Si cette aide-soignante n'est pas en CDD, sa probabilité de présentéisme est de 50,3%. Au contraire, si elle est en CDD, sa probabilité est de 64,5%. D'autre part, un infirmier de 30 ans sans enfants, non-titulaire de son poste et de plus d'un an d'ancienneté aura une probabilité de présentéisme de 76,5% s'il est en CDI contre 85,4% s'il est en CDD. Le rapport des chances (RC) peut être déterminé soit en prenant l'exponentielle du coefficient (cf. ci-dessus), soit en effectuant un rapport de probabilités :

$$RC = \frac{0,645/(1 - 0,645)}{0,503/(1 - 0,503)} = 1,79 \quad \text{dans le premier cas}$$

$$RC = \frac{0,854/(1 - 0,854)}{0,765/(1 - 0,765)} = 1,79 \quad \text{dans le second cas}$$

Or, si l'on raisonne cette fois en termes de probabilité, le multiplicateur des probabilités de présentéisme d'un salarié d'un CDD et d'un salarié en CDI est de 1,28 dans le premier exemple, tandis qu'il n'est que de 1,12 dans le second exemple. L'effet de la variable CDD ne peut donc s'interpréter précisément en termes d'écarts de probabilité que lorsque le statut des autres variables est bien identifié. Il est relatif à une situation de départ. L'intérêt essentiel du rapport des chances est qu'il demeure identique dans toutes les situations. Le même raisonnement est valable pour une variable quantitative comme l'âge. Ici, un écart de 20 ans implique un rapport des chances de 1,55 ($= e^{0,022 \times 20}$).

3 Analyse des résultats

Après avoir présenté les résultats des régressions et décrit la manière de les interpréter, nous accédons à présent à un autre niveau d'analyse. Dans un premier temps, nous reviendrons sur la signification du paramètre DISPERSION estimé dans les régressions binomiales négatives. Cette information sera ensuite utilisée pour répondre à la question de la pertinence des modèles : le modèle ZINB est-il le plus adéquat ? Comment s'ajuste-t-il aux données empiriques ? Une batterie de tests et des graphiques d'ajustement permettront de clarifier ces points. Un indicateur-clef des modèles ZINB sera également étudié. En effet, à partir des coefficients de l'équation d'inflation, il est possible de calculer pour chaque individu une probabilité de présentéisme. De simples statistiques descriptives mettront en lumière l'intérêt de cet indicateur. Enfin, nous montrerons que les informations fournies par le modèle ZINB sont intéressantes et restent cohérentes même en l'absence de variables explicatives (section 3.4), ce qui s'avère précieux lorsque les données à disposition sont pauvres.

3.1 Hétérogénéité non observée

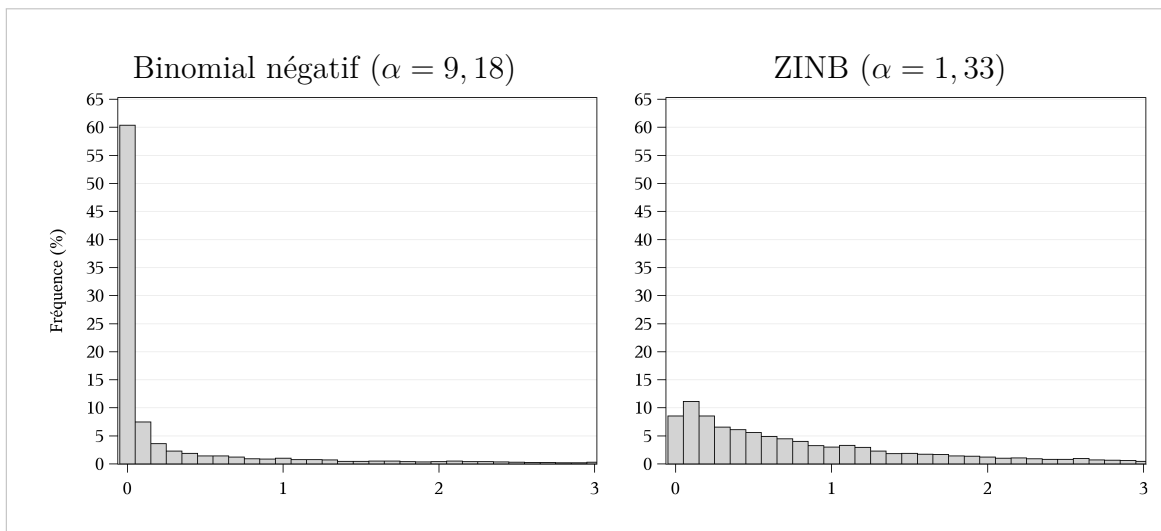
Si la lecture des coefficients paraît désormais relativement claire, il en est un dont l'interprétation n'a pas encore été évoquée dans ce chapitre. Il s'agit du paramètre noté DISPERSION dans les tableaux de résultats (pages 153 et 154). Rappelons que ce paramètre, propre à la régression binomiale négative et au modèle ZINB donne une indication sur la forme distributionnelle du facteur d'hétérogénéité. Celui-ci suit une

loi Gamma dont la valeur des deux paramètres est égale à l'inverse du paramètre estimé dans la régression, traditionnellement noté α .

Ce paramètre α , dans le cas du centre hospitalier (resp. de l'établissement bancaire) et égal à 9,18 (resp. 4,67) dans le modèle binomial négatif et à 1,33 (resp. 1,26) dans le modèle ZINB. Nous verrons plus bas que de ces valeurs peuvent être tirées quelques conclusions sur le choix du modèle le plus adapté. Mais voyons d'abord ce qu'elles signifient du point de vue des différences inter-individuelles. Le terme d'hétérogénéité qu'intègre le modèle binomial négatif est en effet destiné à contourner la propriété de l'égalité entre moyenne et variance inhérente à la loi de Poisson, et permet de modéliser des espérances individuelles d'absence potentiellement très différentes. Dans le cas qui nous intéresse, cette hétérogénéité non observée est notamment liée à l'état de santé qui varie parfois fortement d'un individu à l'autre.

Le graphique 4.7 présente deux distributions d'une variable aléatoire suivant une loi Gamma²². Dans le graphique de gauche, les paramètres de la loi Gamma prennent la valeur $\frac{1}{\alpha} = \frac{1}{9,18}$ tandis que dans le graphique de droite ils prennent la valeur $\frac{1}{\alpha} = \frac{1}{1,33}$. Ces deux situations correspondent à la forme des termes d'hétérogénéité respectivement dans le modèle binomial négatif et le modèle ZINB du centre hospitalier.

Graphique 4.7 – Distribution du paramètre d'hétérogénéité (Hôpital)



22. Reposant chacune sur un tirage de 100 000 observations.

Dans le cas de la régression binomiale négative, le modèle cherche à expliquer le grand nombre de valeurs nulles par le fait que les individus sont en très bonne santé. De ce fait, pour environ 60% des individus la valeur d'hétérogénéité (multiplicative de l'espérance conditionnelle d'absence) est comprise entre 0 et 0,1. Cependant, certains individus obtiendront des valeurs d'absence très élevées (la valeur du dernier centile avoisine 16). En d'autres termes, le modèle binomial négatif, qui considère que tous les zéros (empiriques) proviennent d'un processus unique, se trompe sur les sources de l'hétérogénéité. Le résultat apparent est un paramètre α très élevé, dans les deux contextes d'études.

Le modèle ZINB, pour lequel l'origine d'une partie des zéros est attribuable à l'hétérogénéité comportementale, ne commet pas la même erreur. En comparaison au modèle binomial négatif, le paramètre α évoque une hétérogénéité inter-individuelle plutôt faible. Dans les deux entreprises, il est beaucoup plus proche de zéro que le modèle binomial négatif, $\alpha = 0$ correspondant à une situation poissonnienne. Ceci se manifeste par une distribution moins asymétrique du terme d'hétérogénéité : environ 65% des valeurs sont comprises entre 0 et 1 et les valeurs maximales sont relativement peu élevées (la valeur du dernier centile avoisine 5).

3.2 Ajustement général des modèles

Ce paramètre α constitue un premier indicateur pour la sélection des modèles. Il figure parmi une batterie d'instruments permettant d'une part de choisir le modèle de comptage le plus adapté aux caractéristiques des données d'absences (la surdispersion et l'inflation de zéros, en particulier) et, d'autre part, celui dont les valeurs estimées s'ajustent le mieux à la distribution empirique.

3.2.1 Sélection des modèles

Deux questions se posent à ce stade. Le modèle binomial négatif doit-il être préféré au modèle de Poisson ? Un modèle *zero-inflated* est-il mieux ajusté qu'un modèle standard ? Nous avons déjà répondu à ces questions de manière théorique, mais il est à présent possible de raisonner à la lumière des micro-données d'entreprise.

Surdispersion. Si l'on reprend le processus de sélection des modèles de comptage, la première étape est d'envisager une régression de Poisson. Les résultats des modèles de Poisson présentés ci-dessus ont deux caractéristiques. D'abord, les coefficients sont relativement proches de ceux proposés par le modèle binomial négatif, exception faite des constantes. Ceci confirme que l'estimation n'est pas biaisée en elle-même. Cependant, les erreurs-types sont considérablement sous-évaluées en comparaison au modèle binomial négatif, ce qui conduit à une significativité très élevée pour l'ensemble des coefficients. Le modèle de Poisson peut alors conduire à des inférences erronées. La raison de cette sous-évaluation est à rechercher dans la surdispersion des données (Winkelmann, 2008). Comme l'indiquent Cameron et Trivedi (1998) dans leur ouvrage de référence sur les modèles de comptage, nous disposons de trois instruments pour savoir s'il y a effectivement surdispersion et donc si le modèle binomial négatif est à privilégier au détriment du modèle de Poisson.

D'abord, on peut simplement comparer l'espérance et la variance du modèle de Poisson à la moyenne et à la variance empiriques. Dans les deux entreprises, la variance empirique est largement supérieure à la moyenne empirique : pour l'établissement bancaire, la variance empirique est par exemple de 56,5 avec une moyenne de 3,7. Or, si la moyenne des espérances d'absence estimées par le modèle de Poisson est effectivement égale à la moyenne empirique, la variance estimée est, quant à elle, largement inférieure à la variance empirique. Le modèle semble donc incapable de prendre en compte l'importante dispersion des données.

La deuxième aide à la décision repose sur la valeur du paramètre α du modèle binomial négatif. Si celui-ci tend vers zéro, la variance du modèle binomial négatif ($= \lambda + \alpha\lambda^2$) est égale à l'espérance, ce qui revient à estimer un simple modèle de Poisson. De ce fait, si le paramètre α est significativement non-nul, le modèle binomial négatif est à privilégier (Erdman et al., 2008). Ce test est intégré à la régression économétrique. Dans les deux cas qui nous intéressent, le paramètre α est toujours élevé (4,67 et 9,18 respectivement pour la banque et l'hôpital) et toujours significativement différent de zéro.

Enfin, dans la mesure où les modèles sont emboîtés, on peut tout simplement effectuer un test du rapport de vraisemblance (RV) pour connaître le modèle à privilégier (Winkelmann, 2008). Avec $\mathcal{L}_1$ la log-vraisemblance du modèle de Poisson et $\mathcal{L}_2$ celle

du binomial négatif ce test se calcule comme suit :

$$RV = -2(\mathcal{L}_1 - \mathcal{L}_2) \quad (4.1)$$

Pour les deux entreprises, le résultat du test est significatif, indiquant que l'hypothèse nulle (selon laquelle le meilleur modèle est celui de Poisson) est rejetée.

Notons que la même analyse est aussi vraie pour comparer le modèle ZINB au modèle ZIP, qui sont également emboîtés. Dans le modèle ZINB, le paramètre α est toujours significativement non-nul (1,26 et 1,33 respectivement pour la banque et l'hôpital). Le test du rapport de vraisemblance confirme également que le modèle ZINB est toujours à privilégier au modèle ZIP.

Inflation de zéros. La pertinence d'un modèle *zero-inflated* comparativement à un modèle standard peut être évaluée avec des techniques similaires. Bien entendu, tout notre travail repose sur l'utilisation des modèles à expansion de zéro, puisque sans eux la mesure du présentéisme n'est pas possible. Cela ne nous autorise pas cependant à négliger les arguments statistiques qui permettent de les privilégier aux modèles standards.

Une première indication nous est donnée par l'évolution du paramètre α entre le modèle binomial négatif et le modèle ZINB. Dans les deux entreprises, celui-ci diminue considérablement avec l'usage d'un modèle *zero-inflated*. Dans la section précédente nous avons montré que des conclusions radicalement différentes sur la distribution de l'hétérogénéité non-observées résultaient de cette variation. De plus, d'après Erdman et al. (2008), le fait que la constante de l'équation d'inflation dans le modèle ZINB soit fortement négative est également un signe que le modèle binomial négatif est sans doute préférable, ce qui n'est pas le cas ici.

Une stratégie plus précise pour évaluer le gain inhérent au modèle ZINB serait de pouvoir calculer un test du rapport de vraisemblance. Hélas, celui-ci ne convient pas lorsque les modèles ne sont pas emboîtés (Winkelmann, 2008). Au fond, on pourrait se contenter de comparer les deux log-vraisemblances et s'apercevoir que celle du modèle ZINB est (légèrement) supérieure à celle du binomial négatif, ce qui évoque un meilleur ajustement. Mais la solution classique est d'utiliser le test de Vuong (V), qui repose sur les écarts entre les contributions individuelles aux log-vraisemblances des

modèles. Notons $\delta_i = \ell_{1,i} - \ell_{2,i}$ la différence entre la contribution d'un individu i à la log-vraisemblance du modèle ZINB ($\ell_{1,i}$) et sa contribution à la log-vraisemblance du modèle binomial négatif ($\ell_{2,i}$). Le test de Vuong rapporte la moyenne des écarts ($\bar{\delta}$) à l'écart-type de ces écarts $\sigma(\delta) = \sqrt{\sum_{i=1}^n (\delta_i^2) - \bar{\delta}^2}$. Plus précisément, il est calculé comme suit :

$$V = \frac{\bar{\delta}}{\sigma(\delta)/\sqrt{n}} \quad (4.2)$$

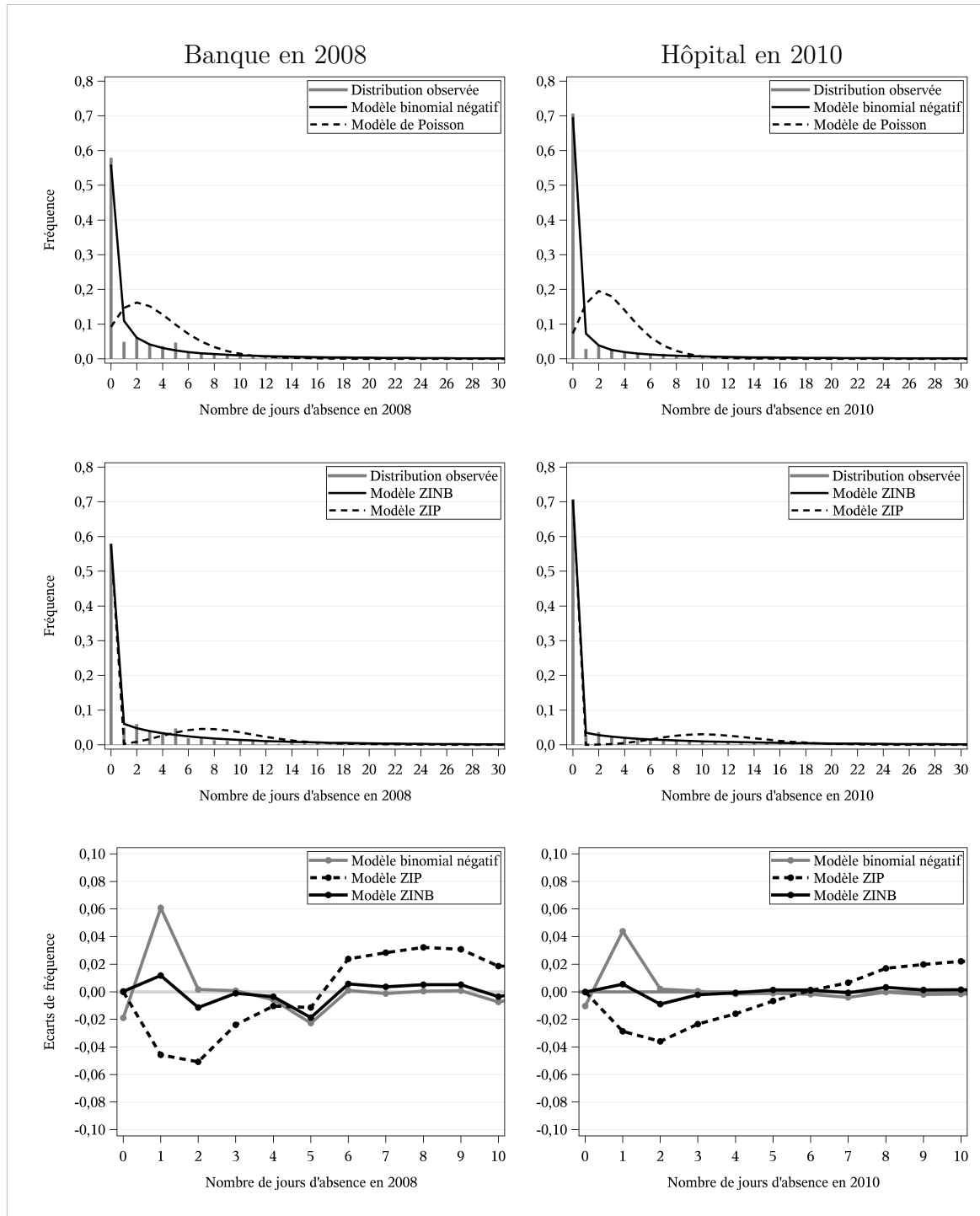
La valeur du test de Vuong suivant une loi normale centrée, une valeur significativement positive milite en faveur du modèle ZINB, et inversement, une valeur significativement négative en faveur du modèle binomial négatif. Dans les deux établissements, on s'aperçoit que le test de Vuong conduit à privilégier d'une part le modèle ZIP au détriment du modèle de Poisson (22,39 pour l'établissement bancaire et 29,02 pour l'hôpital) et d'autre part le modèle ZINB plutôt que le modèle binomial négatif (respectivement 8,55 et 10,83). Tous les tests sont significatifs à 0,1%.

3.2.2 Statistiques d'ajustement

Pour comparer l'ajustement des quatre modèles de comptage aux données empiriques, la première stratégie est de comparer les indicateurs d'ajustement proposés avec les résultats (pages 153 et 154), à savoir la log-vraisemblance et l'AIC. Plus la valeur de la log-vraisemblance (resp. l'AIC) est proche de zéro (resp. faible), plus pertinent est le modèle.

Prenons l'exemple du centre hospitalier. Le constat est simple : il y a d'abord un gain à passer d'un modèle de Poisson ($\mathcal{L}=-23339$) à un modèle ZIP ($\mathcal{L}=-10186$). À sa manière, le modèle ZIP tient compte de la surdispersion des données : il fait l'hypothèse que celle-ci sera entièrement résorbée par la prise en compte d'un second régime générateur de valeurs nulles. L'origine de la quasi-totalité des zéros est alors expliquée par ce régime structurel. De la même façon le passage d'un modèle de Poisson à un modèle binomial négatif ($\mathcal{L}=-6555$) possède un véritable intérêt en terme de log-vraisemblance. Mais cette fois, le modèle ne cherche à expliquer l'inflation de valeurs nulles que par l'hétérogénéité inter-individuelle. Le paramètre α atteint alors une valeur extrême. La meilleure valeur de log-vraisemblance est obtenue par le modèle ZINB ($\mathcal{L}=-6362$), qui prend en compte les deux sources de surdispersion : l'hétérogénéité comportementale et l'hétérogénéité inter-individuelle inobservée.

Graphique 4.8 – Ajustement des modèles



Ces conclusions sont plus parlantes lorsqu'elles sont exprimées par des représentations graphiques. Pour chaque modèle, il est possible de calculer la probabilité d'occurrence d'une valeur donnée d'absence. En comparant ces probabilités estimées aux fréquences d'occurrence réelles, on obtient un bon aperçu de l'adéquation des modèles aux données empiriques. Les résultats de ce traitement apparaissent dans le graphique 4.8 (page 164) : qu'il s'agisse des modèles standards (première ligne) ou de leur extension *zero-inflated* (deuxième ligne), la conclusion est la même. Les modèles de type binomial négatif parviennent toujours à un meilleur ajustement, en particulier sur les valeurs nulles. Mais le meilleur modèle est avec certitude le modèle ZINB qui se confond quasiment avec les valeurs observées. Cela s'observe plus aisément si l'on calcule les écarts entre la fréquence observée et la fréquence estimée pour une valeur donnée (troisième ligne du graphique 4.8). Le modèle ZINB est toujours le mieux ajusté, en particulier dans le cas du centre hospitalier.

Le test du χ^2 de Pearson peut être un outil intéressant pour évaluer l'ajustement d'un modèle donné à la distribution empirique. Notons μ_i l'espérance d'absence de l'un des modèles du comptage étudié. Dans le cas d'un modèle de Poisson $\mu_i = \lambda_i$, mais pour les modèles *zero-inflated* on utilise naturellement l'espérance globale du modèle et non l'espérance conditionnelle :

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\sigma^2} \quad (4.3)$$

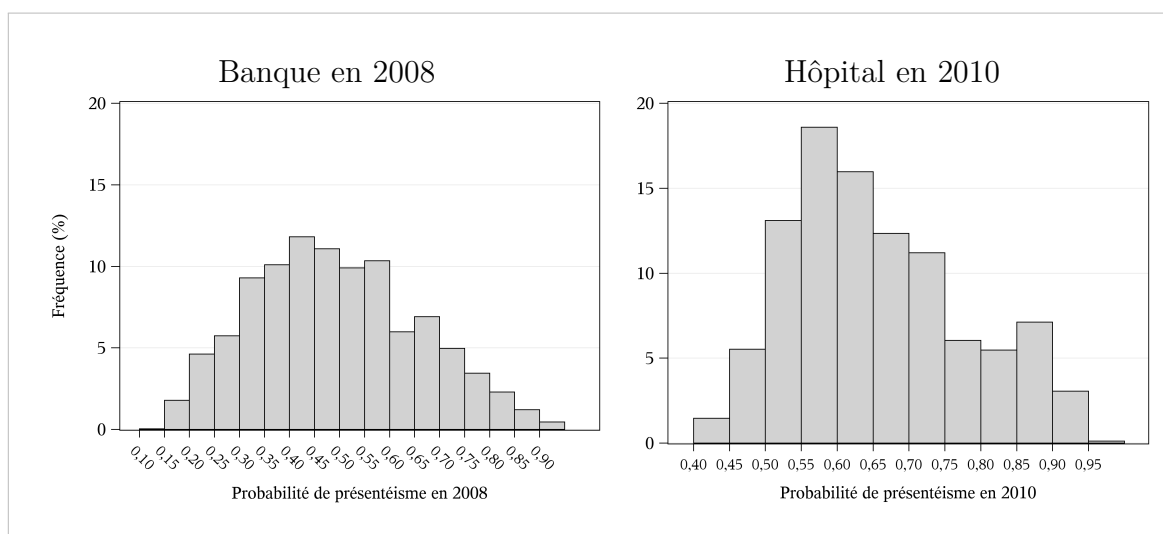
où y_i est le nombre de jours d'absence observé pour l'individu et σ^2 la variance du modèle (cf. chapitres 2 et 3). L'hypothèse nulle du test est que les deux distributions sont en bonne adéquation. Dans les deux établissements que nous étudions, celle-ci peut toujours être rejetée dans le cas du modèle de Poisson et du modèle ZIP. L'hypothèse d'adéquation du modèle binomial négatif ne peut quant à elle être rejetée, mais nous lui préférons le modèle ZINB pour les raisons évoquées ci-dessus. Notons par ailleurs que le rapport entre cette statistique et le nombre de degrés de liberté dans le modèle (nombre d'observations d'où l'on soustrait le nombre de paramètres du modèle) peut être considéré comme un indicateur de surdispersion : s'il est significativement supérieur à 1, il y a surdispersion des données. Dans nos deux cas, lorsqu'un modèle poissonnien est utilisé, il marque toujours une forte surdispersion, que ce soit dans l'établissement bancaire (13,3 pour le modèle de Poisson et 2,4 pour le ZIP) ou dans le centre hospitalier (17,7 et 2,0). En revanche, il est toujours proche de 1 lorsqu'un modèle binomial négatif ou ZINB est utilisé.

3.3 Probabilités de présentéisme

À partir des coefficients de l'équation d'inflation, il est possible de calculer pour chaque individu une probabilité d'appartenir au groupe des présentéistes. Ce calcul se faisant indépendamment de la valeur d'absence observée, il permet l'extension du profil de présentéisme, évalué grâce à l'inflation de valeurs nulles, à l'intégralité de la population. Cet indicateur reflète pour chaque individu, comparativement à la situation de l'ensemble de la population, la propension à venir travailler étant malade pendant une année donnée. Il s'agit donc d'obtenir une information de « risque » individuel à partir du profil de présentéisme obtenu grâce à la modélisation.

Le graphique 4.9 représente les distributions des probabilités de présentéisme dans l'établissement bancaire et dans le centre hospitalier. D'emblée, on peut être interpellé par la distribution relativement à droite au centre hospitalier, tandis que les valeurs sont mieux réparties entre 0 et 1 dans l'établissement bancaire. La probabilité moyenne est donc plus élevée au centre hospitalier (65,6% contre 49,2%) et l'amplitude des probabilités plus faible (0,56 contre 0,8). En avançant la discussion des résultats (chapitre 5), cela pourrait être expliqué par le métier différent des deux entreprises. Les métiers de la santé, en particulier dans le secteur public, sont systématiquement désignés comme particulièrement propices au présentéisme (Aronsson & Gustafsson, 2005 ; Demerouti et al., 2009 ; Dew et al., 2005).

Graphique 4.9 – Distribution de la probabilité de présentéisme



Dans les deux cas, la moyenne et la médiane sont quasiment équivalentes, ce qui semble indiquer une distribution relativement égalitaire autour des valeurs centrales. De façon générale, les probabilités sont fortement regroupées autour de ces valeurs centrales, puisque dans l'établissement bancaire (resp. l'hôpital), 50% de la population a une probabilité comprise entre 36,6% et 60,2% (resp. entre 56,5% et 73,2%) pour une probabilité médiane de 47,9% (resp. 63,6%).

Le fait que l'estimation de l'équation d'inflation, dont découlent les probabilités de présentéisme, repose sur les valeurs nulles empiriques, n'implique pas que les individus ayant une forte probabilité de présentéisme aient systématiquement une valeur nulle d'absence. Les tableaux 4.9 et 4.10 montrent que dans le tiers de la population ayant les probabilités de présentéisme les plus élevées²³, 26,9% des individus ont une valeur positive d'absence dans l'établissement bancaire (20,1% au centre hospitalier).

Tableau 4.9 – Probabilité de présentéisme et jours d'absence (Banque)

Jours d'absence	Probabilité de présentéisme			Total
	[0,00-0,41[	[0,41-0,55[	[0,55-1,00]	
Aucun	43,38%	56,25%	73,1%	57,89%
De 1 à 4	22,10%	19,59%	14,40%	18,61%
De 5 à 10	17,02%	13,70%	7,03%	12,47%
De 11 à 45	17,49%	10,46%	5,47%	11,03%
Total	100,00%	100,00%	100,00%	100,00%

Tableau 4.10 – Probabilité de présentéisme et jours d'absence (Hôpital)

Jours d'absence	Probabilité de présentéisme			Total
	[0,00-0,59[	[0,59-0,70[	[0,70-1,00]	
Aucun	64,76%	69,78%	79,92%	70,13%
De 1 à 4	13,15%	12,17%	8,79%	11,81%
De 5 à 10	12,16%	10,55%	6,28%	10,22%
De 11 à 45	9,93%	7,51%	5,02%	7,84%
Total	100,00%	100,00%	100,00%	100,00%

23. Dans les tableaux 4.9 et 4.10, les populations ont été divisées en trois, suivant les probabilités de présentéisme au 33^e centile (41% pour la banque et 59% pour l'hôpital) et au 66^e centile (respectivement 55% et 70%). Les effectifs des trois sous-populations sont donc égaux.

De même, le fait de n'avoir aucune absence pendant l'année de référence n'entraîne pas automatiquement une forte probabilité de présentéisme. Ainsi, dans l'établissement bancaire, parmi les salariés n'ayant pas été absents en 2008, 44% des salariés se situent dans le tiers des individus ayant les plus fortes probabilités de présentéisme, mais 31,4% dans le tiers intermédiaire et 24,6% dans le tiers inférieur.

Pour finir, il serait tentant de rechercher un seuil au-delà duquel les individus pourraient être considérés comme présentéistes. On pourrait bien sûr utiliser les valeurs centrales. Mais cela reviendrait simplement à scinder la population en deux part égales, étant donné que la médiane et la moyenne sont très proches. Une probabilité de 50% pourrait aussi constituer un seuil, mais au centre hospitalier, plus de 90% de la population se situe au-delà de cette valeur.

Une façon purement statistique de répondre à cette question serait de savoir combien d'individus se situent dans le régime normal d'absence et combien relèvent du régime structurel de valeurs nulles. Prenons l'exemple de la banque. On sait déjà que tous les individus ayant une valeur d'absence non-nulle relèvent du régime normal, soit 42,11% de la population. Sur les 57,89% restants (valeurs nulles), une partie se situe dans le régime normal (autrement dit le modèle binomial négatif) et l'autre dans le régime structurel. La question se pose alors autrement : combien de zéros « naturels » sont générés par le modèle binomial négatif ?

La réponse est donnée par une simple simulation. Dans le modèle ZINB estimé, l'espérance conditionnelle d'absence est de 6,86 et le paramètre α de 1,26. Un modèle binomial négatif de même espérance et dont le paramètre α est identique produit en moyenne²⁴ 16,56% de valeurs nulles. Ces valeurs nulles relèvent elles aussi du régime normal, comme les valeurs positives : 58,67% (16,56% + 42,11%) des individus s'inscriraient donc dans ce régime. Si l'on interprète rigoureusement les résultats du modèle ZINB, les 41,33% restants constitueraient donc le groupe des présentéistes. À partir de là, il serait possible d'extrapoler, en considérant que les 41,33% des individus ayant les probabilités de présentéisme les plus élevées sont à considérer comme présentéistes. Le même calcul effectué pour le centre hospitalier indique que 56,38% de la population relèverait du régime structurel (présentéistes).

24. Simulation réalisée sur un million de tirages.

3.4 Modèle sans variables explicatives

Avant d'aller plus loin, observons une autre propriété intéressante du modèle ZINB. Dans la mesure où nous nous sommes fixé l'objectif de mettre en lumière les facteurs explicatifs du présentéisme, nous avons directement intégré des variables explicatives dans le modèle afin de repérer des effets. Cependant, les informations fournies par un simple modèle sans variables explicatives s'avèrent également très précieuses. Un tel modèle renonce, certes, à expliquer les facteurs de l'absence et du présentéisme, mais sa capacité à évaluer l'hétérogénéité inter-individuelle vis-à-vis de l'absence et l'hétérogénéité comportementale (présentéisme) demeure intacte. À partir de la seule distribution des absences-maladies, il est donc possible d'obtenir une batterie d'indicateurs très informatifs, en particulier une estimation de la probabilité moyenne de présentéistes.

Le tableau 4.11 présente les résultats d'un modèle ZINB sans variables explicatives pour les deux entreprises. Logiquement, les valeurs du paramètre d'hétérogénéité augmentent par rapport aux modèles avec variables explicatives, puisque la part observable de l'hétérogénéité inter-individuelle n'est pas prise en compte. La log-vraisemblance de ces modèles est plus faible (plus éloignée de zéro) que celle des autres modèles avec variables explicatives (binomial négatif et ZIP), ce qui indique un ajustement plus faible.

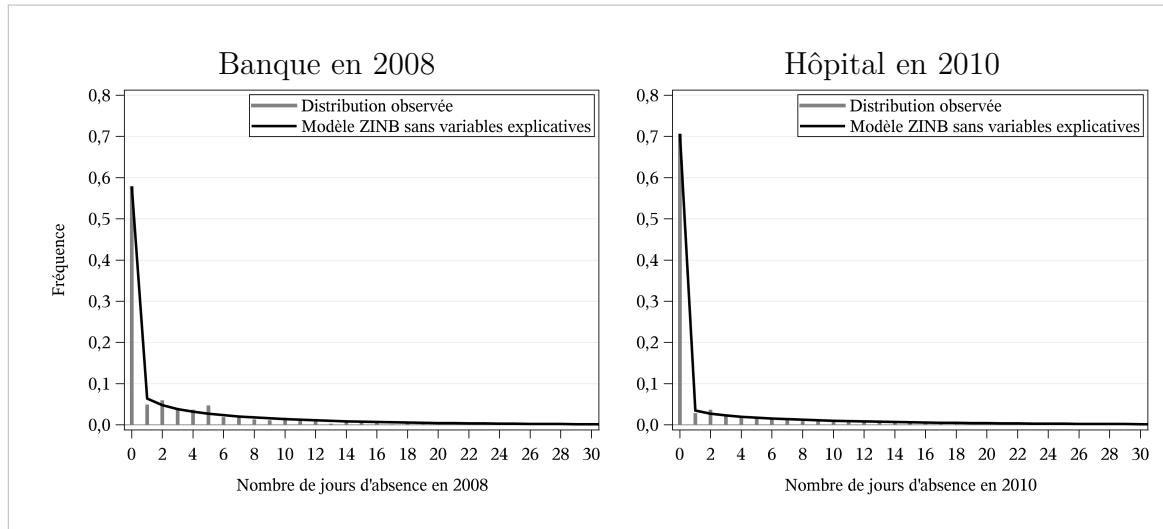
Tableau 4.11 – Estimation du modèle ZINB sans variables explicatives

Paramètre	Banque		Hôpital	
	Coefficient	Err.-type	Coefficient	Err.-type
CONSTANTE DE COMPTAGE	1,944 ***	0,047	2,250 ***	0,040
DISPERSION	1,558 ***	0,142	1,423 ***	0,109
CONSTANTE D'INFLATION	-0,117	0,077	0,636 ***	0,048
Log-vraisemblance (AIC)	-5112 (10230)		-6506 (13019)	

Significativité : <0,1%=*** <1%=** <5%=*

Cependant, au vu du graphique 4.10 (page 170), on s'aperçoit que l'ajustement du modèle ZINB sans variables explicatives aux données empiriques demeure tout à fait valable. Les imprécisions se situent aux mêmes valeurs que dans le cas d'un modèle avec variables et les mêmes raisons peuvent les expliquer.

Graphique 4.10 – Ajustement des modèles ZINB sans variables explicatives



Enfin, le modèle ZINB sans variables explicatives donne des estimations de l'espérance conditionnelle d'absence et de la probabilité de présentéisme voisines des valeurs moyennes observées précédemment. Au centre hospitalier, dans le cadre du modèle avec variables explicatives, l'espérance conditionnelle d'absence est de 9,280 et la probabilité moyenne de présentéisme de 65,66%. Sans variables explicatives, les valeurs moyennes du modèle restent très proches : respectivement 9,492 et 65,37%. Le même constat de proximité est valable pour l'établissement bancaire. Ceci constitue un résultat important de notre thèse et ouvre des voies intéressantes pour des recherches futures : à partir d'un matériau simple (la distribution des absences) on pourrait facilement estimer l'importance du phénomène de présentéisme dans un grand nombre d'entreprises.

4 Exploitation de la dimension longitudinale

L'intérêt et la stabilité de ces résultats est aussi mise en lumière par l'exploitation de la profondeur temporelle de la base de données. Ces informations longitudinales, rarement conservées par les entreprises ou peu fiables, sont disponibles dans l'établissement hospitalier que nous analysons. Dès lors, plusieurs vérifications sont possibles, qui font l'objet des dernières pages de ce chapitre. L'argument principal à retenir est

celui du croisement entre les probabilités calculées en 2010 et les absences antérieures des individus. Ensuite, nous opérons deux types de régressions sur ces données longitudinales. D'une part, le modèle spécifié ci-dessus pour l'année 2010 est appliqué sur chacune des années incluses dans la base de données. Il s'agit donc d'estimer un modèle en coupe transversale sur chaque année, successivement. D'autre part, comme nous l'avons vu lors des simulations de Monte-Carlo (chapitre 3), certains auteurs ont proposé des modèles de type *zero-inflated* adaptés aux données en panels. Une dernière section envisagera leur exploitation dans le contexte du centre hospitalier.

4.1 Probabilités de présentéisme et absences antérieures

Grâce au modèle ZINB, on obtient donc, pour chaque salarié, une probabilité de présentéisme exprimant la propension de celui-ci à venir travailler étant malade. Ces probabilités procèdent directement du profil des individus, de leur personnalité, de leur tempérament : pour certains d'entre eux, le présentéisme constitue l'option préférentielle, voire l'habitude. De ce fait, la tendance au présentéisme est un élément durablement inscrit dans leur comportement.

Une autre façon d'appréhender le présentéisme serait alors de s'intéresser aux absences antérieures de ces individus, et notamment à ceux qui n'ont pas connu d'absence sur une longue période. Si le modèle est pertinent, les individus ayant une probabilité de présentéisme élevée en 2010, devraient donc également avoir une moyenne d'absence très faible sur les années précédentes. Autrement dit, si le présentéisme est effectivement un trait durable et un comportement préférentiel de l'individu, il pourrait aussi se manifester par un faible nombre d'absence à moyen ou à long terme. La moyenne d'absence sur 10 ou 20 années est un indicateur très différent du nombre de jours d'absence sur une année donnée. En effet, plus longue est la période d'analyse, plus rares sont les zéros dits « circonstanciels » (ceux résultant du simple fait de ne pas avoir été malade). Pour le dire autrement, la probabilité de ne pas être malade sur 10 ou 20 ans est pratiquement nulle. Une moyenne d'absence faible sur une plus longue période est donc sans doute plus évocatrice de présentéisme qu'un indicateur annuel.

Un moyen de contrôler la pertinence des probabilités de présentéisme (en 2010) est donc le croisement avec l'historique d'absence des salariés sur les années antérieures. Pour ce faire, nous construisons deux populations : celle des salariés employés continuellement au centre hospitalier sur l'ensemble des 10 dernières années (2001-2010) et celles des

individus employés continuellement sur l'ensemble des 20 dernières années. On constate que, sur l'ensemble des 4090 salariés de 2010, 2081 ont travaillé continûment dans l'entreprise lors des 10 dernières années et 1135 lors des 20 dernières années.

Tableau 4.12 – Probabilité de présentéisme et moyenne d'absence sur 10 ans (Hôpital)

Moyenne d'absence	Probabilité de présentéisme (2010)			Total
	[0,00-0,59[	[0,59-0,70[	[0,70-1,00]	
[0]	5,75%	13,45%	21,19%	11,48%
]0-5]	35,63%	39,29%	45,76%	38,83%
[5-12,5]	27,17%	25,60%	20,90%	25,04%
[12,5-103]	31,45%	21,67%	14,69%	24,65%
Total	100,00%	100,00%	100,00%	100,00%

Population : $N = 2081$

Tableau 4.13 – Probabilité de présentéisme et moyenne d'absence sur 20 ans (Hôpital)

Moyenne d'absence	Probabilité de présentéisme (2010)			Total
	[0,00-0,59[	[0,59-0,70[	[0,70-1,00]	
[0]	2,23%	6,69%	9,21%	5,64%
]0-5]	38,46%	46,25%	52,30%	44,76%
[5,3-11,3]	25,31%	25,56%	22,18%	24,76%
[11,3-64]	34,00%	21,50%	16,32%	24,84%
Total	100,00%	100,00%	100,00%	100,00%

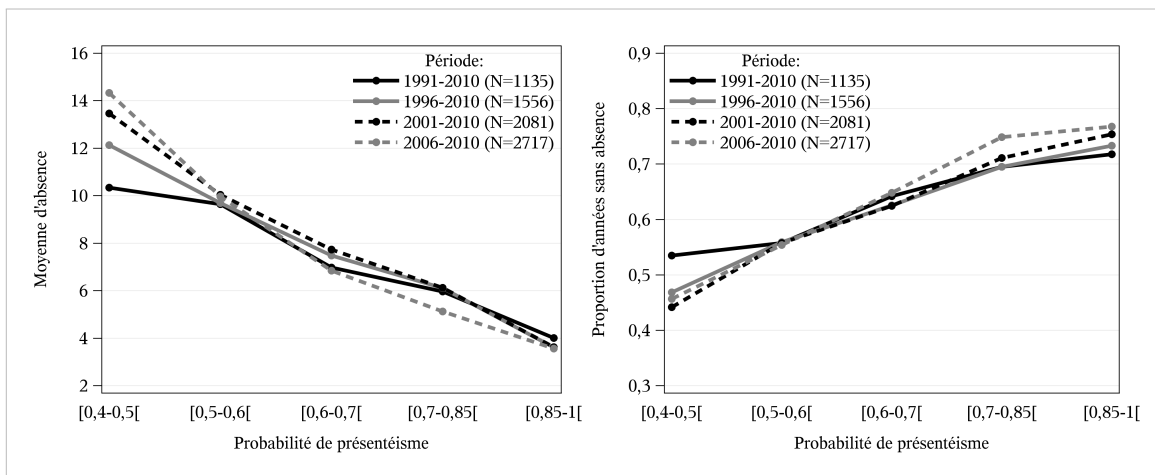
Population : $N = 1135$

Les tableaux 4.12 et 4.13 croisent respectivement le nombre moyen de jours d'absence sur 10 ans et sur 20 ans avec les probabilités de présentéisme produites par le modèle sur la seule année 2010. Ils sont construits de la manière suivante : la population est divisée en trois parts égales selon la probabilité de présentéisme (en colonnes). Le premier tiers possède une probabilité inférieure à 59% (33^e centile), le dernier tiers une probabilité supérieure à 70% (66^e centile) et le tiers intermédiaire une probabilité comprise entre ces deux bornes. En ligne, on trouve la moyenne d'absence sur la période. Pour cet indicateur, une première tranche regroupe les individus n'ayant pas eu d'absence sur la période. La médiane (dont la valeur est de 5 sur la période 2001-2010 et de 5,3 sur la

période 1991-2010) et le troisième quartile (12,5 et 11,3) ont été choisis pour les bornes suivantes.

Il existe clairement des points d'accumulation. Ainsi, dans le tableau 4.12, 67,0% des salariés dotés par le modèle d'une probabilité de présentéisme supérieure à 70% ont, en moyenne, moins de 5 jours d'absence (soit la valeur médiane) sur la période 2001-2010. Dans le même temps, 58,6% des salariés dont la probabilité de présentéisme est relativement faible (inférieure à 59%) ont une moyenne d'absence sur dix ans supérieure à la médiane. Un constat similaire est valable sur la période 1991-2010. Un test d'indépendance du χ^2 montre que les différences de répartition sont nettes sur les deux périodes²⁵. Comme on s'y attend, le test confirme que les principales contributions au χ^2 total se situent dans les quatre coins des tableaux : les individus ayant une faible probabilité de présentéisme sont sous-représentés dans les valeurs faibles d'absence et surreprésentés dans les valeurs élevées. En contraste, les individus ayant une forte probabilité de présentéisme sont sous-représentés dans les valeurs fortes d'absence et surreprésentés dans les valeurs faibles.

Graphique 4.11 – Comparaison avec les données longitudinales



Le graphique 4.11 présente d'autres types de croisement. Le graphique de gauche compare à nouveau les probabilités de présentéisme (réparties en tranches) avec les moyennes d'absence sur plusieurs périodes, tandis que le graphique de droite indique

25. Rappelons que le test du χ^2 se réfère à une répartition théorique basée sur les fréquences marginales. Sur la période 2001-2010, $\chi^2=105,97$ et sur la période 1991-2010, $\chi^2=45,12$. Dans les deux cas, le test est significatif à un seuil de 0,1% sur 6 degrés de liberté.

la proportion d'années sans absence sur les mêmes périodes. À nouveau, il apparaît que les probabilités de présentéisme estimées par le modèle sont très cohérentes avec l'historique d'absence. Les moyennes catégorielles d'absence tendent à diminuer progressivement lorsque les probabilités augmentent, tandis que la proportion d'années sans absence évolue dans le même sens, quelle que soit la période d'analyse.

La conclusion de ces croisements est donc la suivante : les probabilités estimées par le modèle sur la base des seules informations de 2010 sont cohérentes avec l'historique d'absence des salariés sur 10 ans, voire sur 20 ans.

4.2 Stabilité des régressions annuelles

D'autres indications sur la pertinence de l'analyse peuvent être données grâce à la dimension longitudinale de la base de données. Il serait en effet intéressant d'observer l'évolution des coefficients du modèle, tel qu'il est spécifié en 2010²⁶, lorsque celui-ci est appliqué à d'autres années de référence. Une variation erratique de l'ensemble des coefficients marquerait un défaut de stabilité du modèle et/ou un manque de robustesse des variables explicatives.

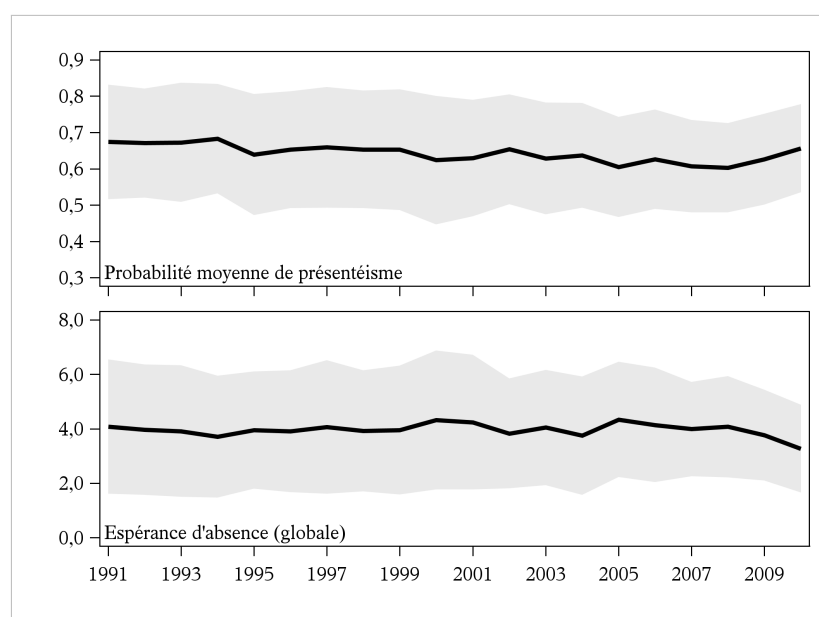
Les modèles estimés pour chaque année sur la période 1991-2010 sont présentés en annexe (page 236). Chaque modèle porte sur les salariés ayant moins de 45 jours d'absence pendant l'année de référence. Les moyennes respectives des coefficients ainsi que leur écart-type et leur coefficient de variation figurent dans le tableau 4.14 (page 175). La stabilité globale peut être énoncée selon deux critères : la constance de la significativité des variables et les variations de leurs coefficients. D'abord, on remarque que certaines variables sont nettement significatives quelle que soit l'année de référence : elles constituent les piliers du modèle. Dans l'équation de comptage, c'est le cas pour la constante, pour les variables AGE et JEUNE ENFANT ainsi que pour le paramètre de dispersion. Dans l'équation d'inflation, outre la constante et la paramètre d'exposition, il en est ainsi pour les variables FEMME, AGE (sauf en 1992 et en 1994), INFIRMIER (sauf en 2002) et CADRE. Les estimateurs de ces variables se caractérisent aussi par un coefficient de variation relativement faible sur la période.

26. Bien entendu, les variables d'intérêt ne sont pas nécessairement les mêmes d'une année sur l'autre. Il serait possible de chercher le « meilleur modèle » pour chaque année. Mais l'enjeu ici est d'observer le comportement des variables que nous utilisons en 2010.

Tableau 4.14 – Statistiques sur les coefficients des modèles annuels (Hôpital)

Paramètre	Coefficient (moyenne)	Écart-type	Coefficient de variation
<i>Équation de comptage :</i>			
CONSTANTE	1,509	0,148	0,098
AGE	0,014	0,003	0,215
JEUNE ENFANT (< 1 an)	0,653	0,106	0,163
ASH	0,201	0,058	0,287
AIDE-SOIGNANT	0,174	0,069	0,397
EXPOSITION	0,279	0,157	0,561
DISPERSION	0,957	0,147	0,153
<i>Équation d'inflation :</i>			
CONSTANTE	1,553	0,518	0,334
FEMME	-0,557	0,114	-0,205
AGE	0,018	0,007	0,401
SANS ENFANTS	0,405	0,086	0,212
INFIRMIER	0,332	0,085	0,256
CADRE	0,868	0,207	0,238
CDD	0,478	0,286	0,597
TITULAIRE	-0,194	0,199	-1,022
ANCIENNETÉ FAIBLE (< 1 an)	0,115	0,160	1,391
EXPOSITION	-1,724	0,303	-0,176

Graphique 4.12 – Stabilité des modèles annuels (Hôpital)



D'autres variables sont un peu moins régulières, en particulier les variables TITULAIRE et ANCIENNETÉ FAIBLE dans l'équation d'inflation : la première reste majoritairement significative, mais son coefficient varie fortement, tandis que la seconde est assez peu robuste sur une période étendue. Mais il est intéressant d'observer que les années où ces variables sont significatives, leurs coefficients ont un niveau très comparable. Dans l'équation de comptage, les coefficients des variables portant sur les métiers sont relativement stables et demeurent majoritairement significatifs. Le statut de la variable CDD dans l'équation d'inflation est plus spécifique. Comme il a été indiqué plus haut, la part des salariés en CDD est assez fluctuante, puisqu'elle dépend des besoins occasionnels du centre hospitalier. De plus, d'une année sur l'autre, il ne s'agit pas des mêmes individus. Même si le coefficient demeure relativement stable, le manque de significativité sur certaines années semble logique.

Au total, les coefficients demeurent relativement stables lorsque le modèle ZINB de 2010 est appliqué sur d'autres années. De ce fait, ils conduisent à des estimations de l'espérance d'absence et des probabilités de présentéisme qui restent elles-mêmes équilibrées. Le graphique 4.12 (page 175) présente la stabilité de l'espérance moyenne et de la probabilité moyenne estimées par le modèle sur la période, avec leur intervalle de confiance (écart-type). Du point de vue individuel, si l'on compare la moyenne des probabilités de chaque salarié sur l'ensemble de la période à son écart-type individuel (variations entre les années), on s'aperçoit également d'une faible dispersion. Dans la plupart des cas, la probabilité de présentéisme ne varie au maximum que de 5 points de pourcentage environ.

4.3 Modèles en panels

Pour finir ce chapitre, nous réalisons une incursion dans les modèles *zero-inflated* en panels. Les coupes transversales annuelles du centre hospitalier (de 1991 à 2010) sont tout simplement empilées, formant un ensemble non-cylindré de 83 124 observations pour 12 748 individus. Près d'un tiers des individus présents en 2010 (26,4%, soit 1 135 personnes) étaient déjà présents en 1991, au début de la période. De même, 48,3% de l'effectif de 2010 a été employé continûment dans l'entreprise sur la période 2001-2010. L'idée est toujours de dresser un panorama du présentéisme dans l'entreprise : or les absences de longue durée ont la particularité d'empêcher le travail, et donc le présentéisme, de façon rédhibitoire. Dans ce panel initial ne sont donc conservées que

les valeurs annuelles d'absence n'excédant pas 45 jours : si un individu dépasse ce seuil, une valeur manquante est générée cette année-là. Il reste alors 77 881 observations pour 12 696 individus.

L'inconvénient de cette stratégie d'exclusion, lorsqu'elle est utilisée sur des données de panels, est qu'elle produit des ruptures de continuité dans les résultats. Un individu présent sur l'ensemble de la période, sera exclu de l'analyse lors d'une année où il avait plus de 45 jours d'absence. Ceci peut s'avérer incommode si le projet est d'étudier l'évolution des probabilités ou des trajectoires individuelles. Mais dans la mesure où les probabilités de présentéisme dépendent uniquement du profil individuel et non de la valeur effective d'absence, rien ne nous empêche par la suite de calculer une probabilité pour les individus exclus à ce stade.

Plusieurs méthodologies sont explorées. D'une part, un simple modèle ZINB, similaire en tout point à celui appliqué aux données en coupe transversale, est utilisé (noté *a* dans le tableau 4.15 page 178). Ensuite, le même modèle est estimé, mais en intégrant cette fois des variables muettes temporelles dans l'équation de comptage (*b*), c'est-à-dire un effet fixe pour le temps. Enfin, deux modèles ZIP à effets aléatoire, qui sont, quant à eux, informés de l'aspect répétitif des données sont estimés : l'un postule que l'effet aléatoire suit une loi normale (*c*), l'autre une loi Gamma (*d*).

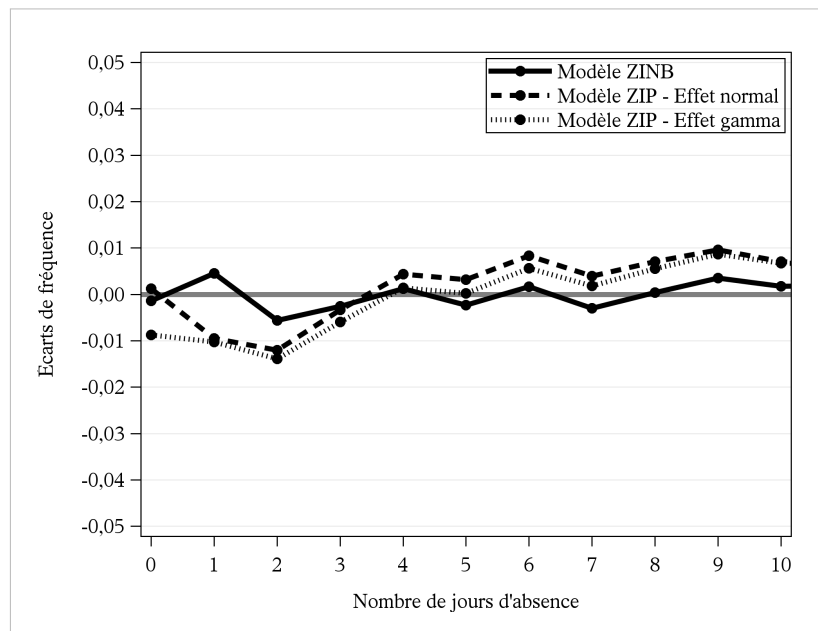
Les résultats sont présentés dans le tableau 4.15 (page 178). Deux constats émergent d'emblée : les coefficients sont remarquablement similaires suivant les différentes stratégies d'estimation et l'ensemble des variables sont significatives à un seuil de 0,1% (à une exception près). Dans les prochaines étapes de nos travaux sur les données de panels, il nous faudra déterminer clairement si cette apparente convergence est un signe de biais dans les modèles ou si elle est due à l'abondance d'information, propre aux données longitudinales. Comme nous l'avons indiqué, la répétition des observations pour chaque salarié renseigne mieux sur l'hétérogénéité inter-individuelle non-observée, ce que confirme un paramètre α en forte baisse.

Cependant quelques informations permettent d'être optimiste sur l'avenir de ces premières estimations. D'une part, les valeurs obtenues sont typiques : elles reflètent parfaitement les résultats obtenus précédemment, lors des régressions annuelles. D'autre part, les modèles s'ajustent bien aux données empiriques (cf. graphique 4.13 page 178) : les écarts sont au plus d'un point de pourcentage avec les fréquences observées et le modèle ZINB ne s'en éloigne pas de plus d'un demi-point.

Tableau 4.15 – Estimation des modèles en panels (Hôpital)

Paramètre	ZINB (a)	ZINB (b)	ZIPN (c)	ZIPG (d)
<i>Équation de comptage :</i>				
CONSTANTE	1,547 ***	1,428 ***	1,273 ***	1,496 ***
AGE	0,013 ***	0,013 ***	0,012 ***	0,011 ***
JEUNE ENFANT (< 1 an)	0,649 ***	0,653 ***	0,578 ***	0,572 ***
ASH	0,210 ***	0,199 ***	0,158 ***	0,184 ***
AIDE-SOIGNANT	0,171 ***	0,171 ***	0,255 ***	0,250 ***
EXPOSITION	0,285 ***	0,266 ***	0,438 ***	0,389 ***
DISPERSION	0,961 ***	0,957 ***	0,358 ***	0,423 ***
<i>Équation d'inflation :</i>				
CONSTANTE	1,622 ***	1,774 ***	1,683 ***	1,547 ***
FEMME	-0,553 ***	-0,556 ***	-0,521 ***	-0,544 ***
AGE	0,018 ***	0,019 ***	0,017 ***	0,017 ***
SANS ENFANTS	0,403 ***	0,404 ***	0,391 ***	0,422 ***
INFIRMIER	0,329 ***	0,342 ***	0,337 ***	0,348 ***
CADRE	0,875 ***	0,870 ***	0,926 ***	0,941 ***
CDD	0,433 ***	0,433 ***	0,464 ***	0,527 ***
TITULAIRE	-0,230 ***	-0,229 ***	-0,226 ***	-0,159 ***
ANCIENNETÉ FAIBLE (< 1 an)	0,050 *	0,081 **	0,057 *	0,128 ***
EXPOSITION	-1,736 ***	-1,758 ***	-1,665 ***	-1,650 ***
Log-vraisemblance	-133431	-133313	-168094	-167358
N=77881	Significativité : <0,1%=*** <1%=** <5%=*			

Graphique 4.13 – Ajustement des modèles en panels



En dernier lieu, à partir de ces résultats, il serait intéressant d'observer l'évolution des probabilités de présentéisme et de l'espérance d'absence, par exemple pour les 1 135 personnes présentes sur l'ensemble de la période 1991-2010. Les trajectoires individuelles étant trop spécifiques, c'est-à-dire trop dépendantes des caractéristiques de l'individu, il faudrait parvenir à une classification en plusieurs catégories homogènes. Un outil exploitable en ce sens est la classification hiérarchique et la construction d'un dendrogramme : l'idée est de répartir les individus en plusieurs classes suivant les distances qui les séparent pour un ensemble de variables. Nous avons réalisé une classification hiérarchique (méthode de Ward) sur les 1 135 salariés présents tout au long de la période, à partir de leurs probabilités de présentéisme et de leurs espérances conditionnelles d'absence pour chaque année. Le nombre de classes attendu est fixé à 30 classes, chiffre suffisamment élevé pour ne pas brider le processus. Les quatre premières classes produites par la classification hiérarchique comprennent 61,2% des individus : leurs profils d'évolution respectifs (moyennes) sont présentées en annexe 5. Ils ne présentent pas d'éléments suffisamment discriminants pour l'analyse, l'évolution des probabilités moyennes étant le plus souvent stable et l'espérance croissante mais située à différents paliers. La classification hiérarchique semble également sensible aux spécificités annuelles, qui ne sont pas l'objet de notre analyse : la classe n° 2 par exemple, regroupe les individus qui présentent une inflexion de l'espérance d'absence en 2010 accompagnée d'une progression du présentéisme.

Conclusion

Les premières confrontations du modèle ZINB à des données réelles mettent en évidence son intérêt pour l'analyse du présentéisme ainsi que la cohérence des mesures qu'il autorise. Bien que les résultats ne seront détaillés et mis en dialogue avec la littérature existante que dans un second temps (chapitre 5), les éléments apportés dans les paragraphes précédents continuent de mettre en lumière la pertinence du modèle. Nous avons eu l'occasion de montrer que ce constat de pertinence est également vrai en l'absence de variables explicatives, ce qui ouvre des perspectives même lorsque l'information disponible est limitée. De façon générale, ce chapitre conduit à trois enseignements majeurs.

Premièrement, le modèle ZINB est parfaitement ajusté aux données empiriques. Différents indicateurs (log-vraisemblance, dispersion), une batterie de tests d'ajustement (rapport de vraisemblance, test du χ^2 , test de Vuong) ou, plus simplement, une analyse graphique des résultats plaident tous en faveur du modèle ZINB. Non seulement, celui-ci se distingue des autres modèles de comptage, mais tout laisse à penser que le cheminement intellectuel qui nous a conduit à le choisir et les hypothèses qui ont été posées sur la nature des données d'absence sont justifiés.

Deuxièmement, les probabilités de présentéisme que le modèle permet de calculer ont, d'une part, une distribution réaliste et, d'autre part, sont cohérentes avec l'historique d'absence des individus. De façon générale, les probabilités au centre hospitalier sont plus élevées que celles observées dans l'établissement bancaire : la moyenne est supérieure et l'amplitude ne couvre pas les valeurs de probabilité les plus faibles. Ceci est parfaitement cohérent avec une affirmation récurrente de la littérature selon laquelle le secteur de la santé serait particulièrement touché par le présentéisme. De plus, si l'on exploite la profondeur temporelle qui caractérise la base de données du centre hospitalier, on s'aperçoit que les probabilités calculées en 2010 sont cohérentes avec les enregistrements des absences passées des salariés. Dans la catégorie des probabilités les plus élevées, on retrouve en effet des individus dont la moyenne d'absence était particulièrement faible pendant les 10 ou 20 années précédentes ou dont la propension à ne pas s'absenter était plus élevée.

Troisièmement, que l'on estime des modèles empilés grâce à l'incursion dans les techniques de panels ou que le même modèle soit appliqué successivement sur chaque année comprise entre 1991 et 2010, les coefficients des variables sont remarquablement significatifs et stables. Globalement, cela nous permet non seulement de penser que les résultats du modèle ZINB demeurent stables, même s'il est appliqué à des populations partiellement différentes, mais également que les variables explicatives qui entrent dans la spécification du modèle constituent des facteurs expliquant adéquatement et durablement le comportement de présentéisme. Cela nous permet d'aborder maintenant avec confiance l'analyse de ces facteurs explicatifs du présentéisme.

Chapitre 5

Des facteurs explicatifs du présentéisme

Introduction

Si l'objet central des trois derniers chapitres était de déterminer une méthode de mesure pertinente du comportement de présentéisme, il est temps, à présent, de revenir à ce qui justifie la mise en place de ces moyens statistiques. L'autre objectif de notre travail est en effet de mieux connaître le présentéisme et d'en caractériser les facteurs. Cet ultime chapitre est à lire en balance avec celui qui ouvrait notre thèse et dont l'enjeu était de rapporter précisément ce que l'on sait déjà du présentéisme. Dans les lignes qui suivent, nous détaillons ce que nous pouvons nous-mêmes affirmer à propos de ce comportement, en nous appuyant sur les régressions effectuées précédemment. Nos résultats seront toujours mis en dialogue avec les informations déjà existantes dans la littérature : celles-ci sont le plus souvent vérifiées et complétées, mais aussi parfois remises en question par nos propres conclusions.

L'esprit de ce chapitre n'est donc pas d'opposer deux types de connaissances obtenues par deux stratégies différentes, mais bien de parvenir à une description du présentéisme

enrichie par la diversité méthodologique. L'exploitation du modèle ZINB sur d'autres cas d'entreprises, ou mieux sur de nombreux cas d'entreprises, permettrait des avancées en ce sens, sans oblitérer l'intérêt de travaux qualitatifs autorisant sans doute un niveau de précision plus élevé.

Au moment d'entrer dans le détail des éléments explicatifs du présentéisme, il est utile de reformuler la question originelle qui a été le déclencheur de cette thèse, puisque nous nous apprêtons à y répondre. Les salariés d'une même entreprise se répartissent *a priori* en deux catégories : ceux qui se sont absentés pendant l'année de référence et ceux qui n'ont pas connu d'absence. Dans la plupart des cas, les entreprises s'interrogent sur les caractéristiques des premiers, dans la perspective éventuelle de limiter leur niveau d'absence. Il s'agit là de l'approche classique, et cohérente au demeurant, du management des ressources humaines : comprendre le phénomène et trouver des leviers d'action (Gallois, 2005 ; Stankiewicz & Geuze, 2007).

L'analyse proposée dans ce chapitre, et plus largement dans cette thèse, n'exclut absolument pas cette question du niveau d'absence, mais au contraire l'éclaire d'une lumière nouvelle. En effet, les « non-absents », ne sont pas seulement des personnes en bonne santé ou des salariés refusant de prendre des absences injustifiées, mais également des personnes qui s'interdisent un arrêt de travail malgré un état de santé nécessitant du repos. Ce comportement, le présentéisme, remet radicalement en question le problème de l'absence de la main d'œuvre et devrait également renouveler la réflexion et les pratiques managériales. Pourquoi certains salariés sont-ils présents étant malades ? Veulent-ils se montrer dévoués envers l'entreprise ? Sont-ils trop investis dans leur travail ? Cèdent-ils à la nécessité du service ? À des nécessités personnelles ? Les questions irrésolues restent plus nombreuses que les moyens disponibles pour y répondre. Bien que des éléments généraux soient donnés dans la littérature, la question nous a semblé d'une telle importance qu'il nous semblait incontournable de l'opérationnaliser, d'y répondre dans le cadre restreint de l'entreprise elle-même. D'où la méthode économétrique proposée.

Pour exposer les résultats obtenus grâce à cet outil, une structure comparable à celle utilisée dans le premier chapitre sera retenue dans un premier temps : facteurs organisationnels (section 1) et individuels (section 2) seront distingués et abordés successivement. La forme de ces deux premières sections est identique : après un passage en revue des variables explicatives, elles intègrent chacune l'analyse détaillée d'une variable spécifique se distinguant comme un facteur essentiel de présentéisme dans sa catégorie :

la responsabilité d'équipe pour les facteurs organisationnels (section 1.2) et l'âge des salariés pour les facteurs individuels (section 2.2). L'étude de ces variables sera davantage approfondie que les autres éléments explicatifs et des régressions économétriques complémentaires (menées au centre hospitalier pour la question de l'âge et dans l'établissement bancaire pour celle de la responsabilité d'équipe) y seront consacrées. Enfin, à la lumière des résultats sur le présentéisme, nous présenterons les enseignements du modèle ZINB à l'égard de l'absence et de l'absentéisme des salariés.

Afin de décrire au mieux l'effet d'un facteur donné sur la propension au présentéisme, les indicateurs présentés dans le chapitre précédent seront systématiquement exploités. Ainsi, nous rappellerons pour chaque coefficient le rapport des chances (dans les deux entreprises) et nous nous appuierons sur un exemple concret pour illustrer l'écart (relatif) de probabilités¹.

Préambule : schéma d'analyse

Avant d'entrer dans l'interprétation et l'approfondissement des sorties économétriques obtenues dans le chapitre précédent, nous voudrions porter un éclairage particulier sur les résultats. D'après nous, une lecture méthodique des coefficients estimés, associée à une réflexion sur la signification des variables incluses dans les régression, fournit un argument supplémentaire sur la pertinence de notre méthode d'analyse.

Revenons d'abord aux motifs plausibles d'une valeur nulle d'absence. Pourquoi un individu pourrait-il ne pas s'absenter au cours d'une année donnée ? Deux possibilités s'imposent à l'évidence : soit l'individu, comparativement à la moyenne, est plus résistant physiquement, il n'est jamais ou très peu malade, soit il se rend à son travail malgré la maladie et adopte ainsi une règle comportementale de présentéisme. Comme nous l'avons vu précédemment, aucun autre scénario ne peut efficacement expliquer une telle fréquence de valeurs nulles dans les données d'absence.

1. En effet, comme nous l'avons expliqué dans le chapitre précédent, les comparaisons de sous-populations en termes de probabilités ne sont pas rigoureuses d'un point de vue statistique. Les probabilités pour un coefficient donné dépendent en effet du statut des autres variables de la régression. Nous utilisons donc également un autre indicateur, le rapport des chances, qui reste pertinent quelle que soit la situation de départ. Un rapport des chances supérieur à 1 (c'est-à-dire un coefficient positif) implique une augmentation des chances, tandis qu'un rapport inférieur à 1 (c'est-à-dire un coefficient négatif) indique une diminution des chances.

Or, l'une des objections qui pourrait être soulevée à l'encontre du modèle ZINB, tel qu'il est exploité dans ce travail, est qu'il ne mesure pas le présentéisme, mais simplement la bonne santé ou la résistance physique des individus. Le régime structurel serait alors non pas celui des présentéistes, mais celui des salariés qui ne sont jamais malades ou dont l'état de santé est supérieur à la moyenne. Si cette objection s'avérait fondée, on s'attendrait donc à trouver dans l'équation d'inflation des coefficients eux-mêmes compatibles avec une bonne santé. Le cas le plus évident est celui de la variable AGE. Si l'équation d'inflation fournit simplement une mesure de bonne santé et non de présentéisme, le coefficient de cette variable devrait être négatif, la santé se dégradant naturellement avec l'âge.

Pour la variable d'âge, comme pour la plupart des variables estimées dans l'équation d'inflation, cette objection ne tient pas. Le coefficient de la variable AGE est significativement positif. Cela signifie que l'apparition du phénomène observé augmente avec l'âge. Il ne peut donc s'agir d'un bon état de santé, puisqu'il est établi que l'état de santé se détériore lorsque l'âge augmente et que le taux d'incidence de nombreuses maladies progresse, parallèlement. Cette dégradation de la santé avec l'âge est d'ailleurs bien traduite par une augmentation de l'espérance d'absence dans l'équation de comptage. Le présentéisme demeure donc la seule interprétation plausible de l'équation d'inflation. Seule une telle règle comportementale permet d'expliquer correctement l'existence d'un régime de valeurs nulles.

D'autres variables peuvent aussi être analysées à la lumière de ce raisonnement. Par exemple, toutes choses égales par ailleurs, pourquoi les salariés sans indemnisation (coefficient égal à 0,682 dans l'établissement bancaire) auraient-ils un meilleur état de santé que les autres ? Rien ne permet de penser que ces individus aient un meilleur accès au soin où qu'ils soient davantage soucieux de leur santé. Au contraire, la question financière est souvent mise en lien avec un renoncement au soin (Dourgnon et al., 2012). Le raisonnement vaut aussi pour les salariés en CDD, dont les salaires sont relativement moins élevés. Même lorsqu'il s'agit de variables concernant les métiers des individus (e.g variable INFIRMIÈRE), le seul état de santé ne peut expliquer les différences de coefficients.

1 Facteurs organisationnels de présentéisme

Le première catégorie de facteurs que nous abordons regroupe les éléments liés spécifiquement à l'emploi de l'individu, que nous distinguons des facteurs relatifs à l'individu lui-même. Par nature, le présentéisme se situe en effet au carrefour entre la sphère privée et de la sphère professionnelle. Il est au centre d'une sorte de dialectique entre ces deux environnements. D'un côté, l'individu est amené à prendre une décision sur un événement privé, la maladie, qui a des conséquences sur son activité professionnelle. Mais il prend également en considération son contexte de travail pour évaluer si, oui ou non, il peut prendre un temps de soin et de repos. D'ailleurs, s'il présente une importante propension au présentéisme, ce contexte de travail prend vraisemblablement plus d'importance. Dans cette section, nous détaillons dans un premier temps quatre facteurs organisationnels ayant du poids dans cette décision. Puis, au travers d'une analyse plus approfondie, une place spéciale est réservée à un autre facteur : la responsabilité d'équipe (cf. 1.2). Elle produit des effets importants tant pour le responsable d'équipe lui-même que pour ses subordonnés.

1.1 Effets des variables liées à l'emploi

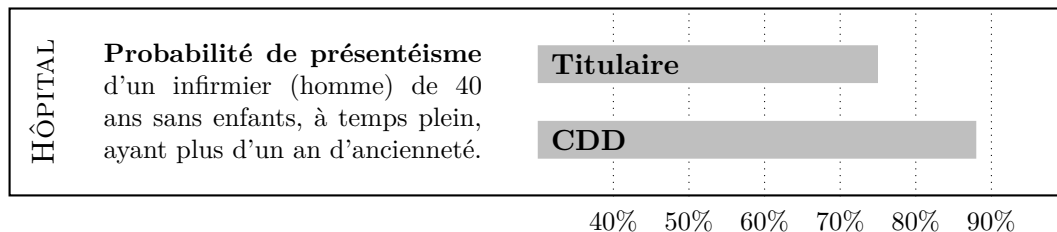
Dans un premier temps, nous chercherons à déterminer l'impact des facteurs organisationnels sur le présentéisme. Les trois premières variables explicatives évoquées dans cette partie relèvent davantage des éléments contractuels qui lient l'individu à son entreprise : la nature de son contrat de travail, les règles qui s'appliquent pour l'indemnisation de l'absence et les éléments relatifs à l'entrée dans l'emploi puis à l'évolution dans l'entreprise. Les aspects intrinsèques à l'emploi sont traités dans un second temps. Il s'agit du contenu des activités professionnelles ainsi que de la responsabilité d'encadrer une équipe. Ce dernier élément fera l'objet d'une étude spécifique parce qu'il nous semble être une clef de compréhension du comportement de présentéisme au sein d'une équipe de travail.

1.1.1 *Nature du contrat*

L'entrée d'un salarié dans un emploi en contrat à durée déterminée (CDD) s'accompagne souvent d'indications sur les conditions d'une éventuelle pérennisation du contrat

de travail. Parmi celles-ci, le nombre d'absences-maladie occupe une place de choix : au centre hospitalier que nous étudions, au-delà d'un certain niveau d'absence, le passage au contrat à durée indéterminée (CDI) ou une titularisation deviennent impossibles. D'ailleurs, que ces conditions de pérennisation soient explicites ou non, un salarié en CDD est souvent enclin à donner une bonne image de lui-même : l'effort qu'il déploie au profit de l'entreprise se mesure parfois à sa propension à l'absence. On peut donc s'attendre à ce que ces salariés en CDD soient plus susceptibles de renoncer à un arrêt-maladie pour préserver leur chances d'accéder à un emploi en CDI. Comme l'indiquent Grinyer et Singleton (2000), ils ont à opérer un arbitrage entre deux risques : un risque de manquer une opportunité professionnelle si leur niveau d'absence est jugé trop élevé et un risque pour leur santé s'ils ne se soignent pas.

Au centre hospitalier, le fait d'être en CDD est effectivement en relation positive avec le présentéisme. Le coefficient de cette variable est d'ailleurs l'un des plus élevé (0,585) et le rapport des chances est de 1,8 pour les salariés en CDD, comparativement aux autres, titulaires de leur poste ou en CDI. Si on confronte les titulaires aux salariés contractuels (CDI et CDD), on s'aperçoit par ailleurs que les premiers ont une probabilité plus faible que les seconds. L'écart est donc encore plus important entre un salarié en CDD et un salarié titulaire, comme le montre l'exemple suivant :



Ceci semble confirmer qu'un CDD a bien le statut d'antichambre d'un CDI ou d'une titularisation et qu'être présentéiste est envisagé comme un moyen d'augmenter ses chances en ce sens. Dès lors, il paraît intéressant aux salariés se trouvant dans cette situation de placer les contraintes professionnelles au-dessus de toutes les autres (y compris au-dessus de leur santé) afin de donner une image d'implication et de fiabilité. Il s'agit là d'une autre facette du présentéisme. Ce comportement est conçu comme un gage de loyauté d'un salarié envers l'organisation, comme une manière de manifester ses efforts et sa motivation. Nous confirmons en cela les résultats de Simpson (1998). Cette propension au présentéisme est d'ailleurs encore plus manifeste lorsqu'un salarié en CDD est placé dans un contexte de concurrence avec d'autres individus en CDD. Au

centre hospitalier, plus le nombre de « concurrents » en CDD augmente, plus le salarié (lui-même en CDD) est enclin au présentéisme, comme nous l'avons testé séparément².

Cette stratégie est sans doute temporaire, puisque nos résultats montrent également que l'accès à une position de titulaire inverse cette tendance. Le signe de la variable TITULAIRE est négatif, indiquant que les salariés titulaires sont moins touchés par le présentéisme (rapport des chances de 0,73). La sécurité de l'emploi, souvent entrevue comme un aboutissement par ceux qui s'engagent dans la fonction publique, aurait alors pour effet de rééquilibrer les priorités, autrement dit d'envisager plus sereinement la possibilité de se soigner avant de retourner au travail.

De façon plus générale, c'est la question de la sécurité dans l'emploi qui est ici en jeu. La décision de ne pas s'absenter en cas de maladie est souvent liée à un jugement du salarié sur l'avenir du poste qu'il occupe, que ce jugement soit basé sur des critères subjectifs (Johns, 2011) ou objectifs. Ainsi, Askildsen et al. (2005) ont-ils par exemple montré la sensibilité du niveau d'absence au taux de chômage. De même Hansen et Andersen (2008) soulignent que lorsque le taux de chômage augmente, la propension au présentéisme progresse également. Il y a donc bien un rapport entre les perspectives professionnelles et les comportements d'absence, qui se manifeste ici par le présentéisme de salariés en situation incertaine.

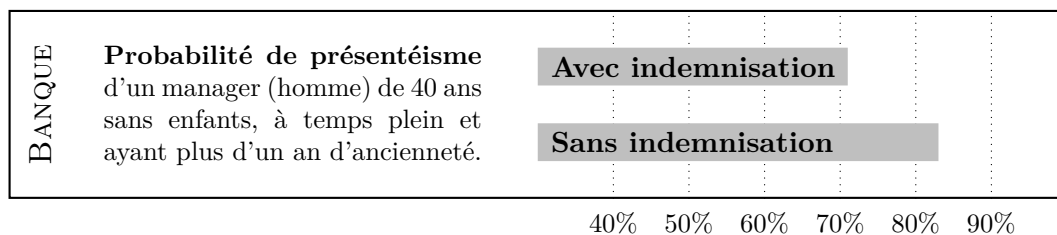
1.1.2 Coût financier de l'absence

En France, dans le régime général d'assurance-maladie, lorsqu'un salarié est absent pour des raisons de santé, il est indemnisé d'environ 50% de son salaire de base, après un délai de carence de trois jours. Cette indemnisation est en outre réservée aux salariés ayant travaillé un certain nombre d'heures au cours des trois mois précédant l'absence. Certaines personnes, jugeant l'absence trop coûteuse du fait de la perte totale ou partielle de leur salaire, peuvent être amenées à y renoncer, ce qui n'est pas sans inconvénients pour l'employeur, comme nous en avons discuté au début de cette thèse. Pour ces raisons, et en lien avec sa politique sociale, l'employeur peut contribuer à une prestation d'assurance-maladie complémentaire, visant par exemple à réduire le délai de carence ou à compléter les indemnités légales à hauteur du salaire.

2. Pour ce faire, nous avons observé l'interaction entre la variable CDD et une variable indiquant le nombre d'autres salariés en CDD dans l'équipe de travail. Le coefficient est positif et significatif à un seuil de 1%.

C'est le cas pour l'établissement bancaire. L'entreprise perçoit les indemnités journalières du salarié malade et les complète de façon à lui maintenir l'intégralité de son salaire et à prendre en charge le délai de carence. C'est le principe de la subrogation. Toutefois cet avantage n'est accordé qu'aux salariés de plus d'un an d'ancienneté, les autres ne bénéficiant que de la couverture maladie du régime général. Une manière de ne pas subir les pertes liées à cette indemnisation partielle pour les individus de moins d'un an d'ancienneté serait d'être présentéistes : on peut donc s'attendre à ce que le risque soit plus élevé pour ces personnes. D'un point de vue national, les conclusions de Claes (2011) vont dans ce sens : le présentéisme est plus élevé dans les pays où les politiques d'indemnisation de l'absence-maladie sont les moins généreuses.

Nous avons estimé l'impact de cet avantage sur la probabilité de présentéisme. Nos résultats montrent que les salariés ne bénéficiant pas de l'indemnisation totale ont une probabilité de présentéisme beaucoup plus importante : il s'agit même du coefficient le plus élevé (le rapport des chances est égal à 2). Voici un exemple :



Cette conclusion est congruente avec celle de Böckerman et Laukkanen (2010). Dans leur travail, certains salariés disposent annuellement de trois jours d'absence au cours desquels leur salaire est intégralement maintenu. Ces personnes sont moins présentéistes que les autres. De façon générale, le coût financier de l'absence, parfois généré volontairement dans le cadre de politiques de lutte contre l'absentéisme, a une influence tant sur le niveau d'absence des salariés qu'il réduit (Barmby, 2002) que sur le présentéisme qu'il accroît (Claes, 2011). Un salarié qui estime que trois journées sans solde (délai de carence) seraient trop difficiles à supporter, en particulier s'il est seul ou si son ménage n'a pas d'autres revenus, est donc logiquement amené à renoncer à l'arrêt-maladie alors qu'il n'aurait pas agi de la sorte avec une indemnisation complète.

Ces résultats concernant les politiques d'indemnisation de l'absence renvoient plus largement à un présentéisme pour raisons financières. Les articles d'Aronsson et al. (2000, 2005) évoqués dans le premier chapitre trouvent un lien entre les besoins financiers de la personne et le présentéisme : plus ce besoin est important, plus élevé est la propension

au présentéisme. Même si nos régressions n'intègrent pas directement de variable captant cet effet, nos résultats sur l'indemnisation partielle de l'absence semblent confirmer ce résultat.

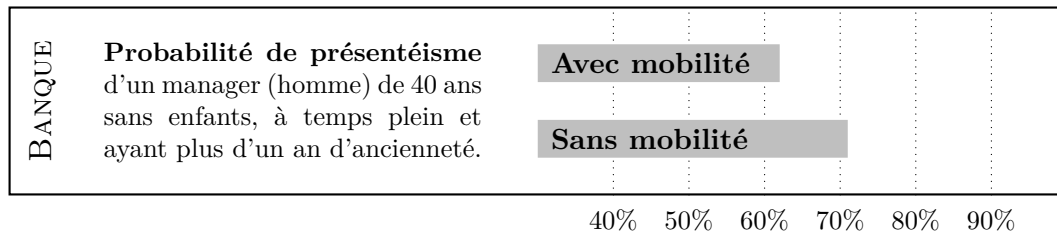
1.1.3 Entrée et évolution dans l'emploi

L'hypothèse selon laquelle l'installation d'un salarié dans un nouveau poste peut également être une source de présentéisme, paraît relativement raisonnable. La prise de poste implique en effet un investissement particulier du salarié, lié à l'appropriation de nouvelles activités, à l'insertion dans un nouvel environnement relationnel, etc. Un nouvel emploi peut susciter de l'enthousiasme, une réelle satisfaction, lorsqu'il est conforme au choix du salarié, mais peut aussi générer un mécontentement, lorsqu'il procède d'une mobilité subie. Le lien avec l'absence et le présentéisme semble alors possible. Mais il se trouve que la littérature est étrangement silencieuse à ce propos.

Dans chaque cas étudié, une variable porte sur cette thématique, même si leur sens n'est pas tout à fait similaire. Au centre hospitalier, la variable muette FAIBLE ANCIENNETÉ indique si le salarié possède moins d'un an d'ancienneté dans l'emploi. Elle teste davantage une entrée « choisie » dans un nouveau poste, qu'elle résulte d'un recrutement externe ou d'une mobilité interne. L'hypothèse initiale semble confirmée. En effet, les personnes dans cette situation sont plus susceptibles de venir travailler étant malades (rapport des chances de 1,5). La majorité de ces salariés ont moins de 30 ans (pour 78,1% d'entre eux), ce qui suggère qu'une grande partie de l'effet serait en lien avec une entrée dans un premier poste. Or, le début de la vie active est naturellement marqué par un manque d'expérience et une phase de découverte impliquant un investissement conséquent dans les activités professionnelles, et pouvant donc conduire à un renoncement à l'absence-maladie.

En contraste, la variable MOBILITÉ incluse dans l'équation d'inflation de la banque, mesure davantage une mobilité subie. En effet, au cours de l'année qui précède l'étude, l'entreprise a connu de fortes restructurations liées à la fusion de trois organisations infra-régionales. Dans ce contexte, de nombreux salariés se sont vus imposer des mobilités de filière, impliquant un changement radical de métier. Ainsi, certaines personnes ont-elles été déplacées de métiers commerciaux vers ceux de la gestion bancaire. On peut aussi par exemple noter le cas d'une mobilité entre une unité gérant les crédits vers celle gérant les ressources humaines, etc. Par ailleurs, ces mobilités profession-

nelles se sont souvent accompagnées d'une mobilité géographique régionale, puisque de nombreux services ont été centralisés. Le présentéisme des salariés concernés est sensiblement plus faible (rapport des chances de 0,66). On peut également proposer le cas suivant :



Ici le présentéisme prend un autre visage. Certains salariés peuvent le considérer comme un moyen de montrer leur engagement (comme c'était déjà le cas des salariés en CDD) ou leur loyauté envers une organisation qui les valorise. Lorsque cette relation de confiance est mise à mal, lorsque le salarié se sent trahi ou mis en danger par l'évolution de l'organisation (Hausknecht et al., 2008), renoncer au présentéisme et choisir l'absence-maladie devient alors une manière de contester les décisions de l'entreprise à son propos. Il existerait donc une sorte de contrat moral entre l'entreprise et le salarié : en contrepartie de son investissement, qu'il croit attester par le présentéisme, le salarié attend de son entreprise qu'elle le protège en cas de changements économiques ou organisationnels.

Ceci est vrai lorsqu'on mesure le présentéisme après la restructuration. En revanche, lorsque la mesure du présentéisme se situe en amont ou au cours de la réorganisation, la perspective de possibles mobilités professionnelles pourrait avoir l'effet inverse. C'est la conclusion de Caverley et al. (2007) : leur étude a lieu au cours d'une restructuration impliquant une réduction significative du personnel. Les salariés connaissent un présentéisme élevé à mettre en lien avec la crainte d'un licenciement prochain, comme s'ils cherchaient à prouver leur implication et leurs efforts en demeurant coûte que coûte à leur poste de travail.

Dans l'établissement bancaire, la mesure du présentéisme se situe chronologiquement en aval de la restructuration (une année après). Les résultats indiquent clairement que les individus ayant subi une mobilité de filière sont moins présentéistes. Une mobilité forcée réduirait donc cette forme extrême d'engagement que constitue le présentéisme. Nous avons cependant voulu vérifier si l'effet sur le présentéisme était différent en cours de restructuration : ceci est possible, puisque nous disposons des données d'absence pour

l'année de la restructuration (2007). Le même modèle appliqué à cette année-là indique un coefficient positif (quoique faiblement : 0,06) mais non-significatif. Même si l'effet n'est pas tranché, il suggère que le présentéisme n'a pas la même importance suivant l'étape de restructuration dans laquelle se situe l'entreprise.

1.1.4 Nature des tâches

Les articles parcourus dans le premier chapitre laissent entendre que le présentéisme ne se manifeste pas identiquement suivant la catégorie d'emploi ou, autrement dit, suivant la nature des tâches effectuées par les salariés. Intuitivement, il semble en effet que le présentéisme devrait être plus prégnant dans les professions où les conséquences de l'absence-maladie sont les plus lourdes. Par exemple, on peut penser que dans les métiers commerciaux, où le manque à gagner d'une journée d'absence se traduit directement en chiffre d'affaires perdu, le présentéisme est plus élevé que dans des métiers administratifs, où le travail s'accumule et peut être accompli au retour du salarié.

Tel est aussi le cas pour les salariés exerçant une profession en lien avec les besoins primaires d'autres personnes. Dans la littérature, un consensus se dessine par exemple autour des métiers du soin (Aronsson & Gustafsson, 2005 ; Elstad & Vabø, 2008). Au centre hospitalier, l'effet du métier d'infirmière sur le niveau de présentéisme est sans doute à interpréter en ce sens. Le rapport de chance d'être présentéiste pour une infirmière comparativement aux autres emplois est de 1,38. Cet emploi requiert en effet l'exercice d'actes médicaux et une relation particulière et suivie avec les patients. Selon Demerouti et al. (2009) la proximité quotidienne avec les patients conduit les infirmières à considérer que leurs tâches ne sauraient être reportées ou déléguées, les conduisant au présentéisme. Elstad et Vabø (2008) renvoient également le présentéisme des soignants à un ensemble de « normes professionnelles » informelles qui font de la présence au chevet des malades, une sorte d'obligation morale pour les individus. Ce comportement serait d'ailleurs renforcé dans les situations où le personnel est trop peu nombreux.

Il semble également que, de manière générale, la propension au présentéisme soit plus prononcée dans le centre hospitalier que dans l'établissement bancaire. Bien que la comparaison des deux distributions soit à prendre avec prudence, elle renvoie directement à deux résultats récurrents de la littérature sur le présentéisme. D'une part, les individus exerçant des métiers du secteur médical, ou plus largement du service

à la personne, sont plus présentéistes que les autres (Aronsson & Gustafsson, 2005). Ils savent que des personnes, souvent en situation de vulnérabilité, dépendent de leur présence au travail (Demerouti et al., 2009 ; Elstad & Vabø, 2008) et ont la conviction profonde que leur responsabilité est engagée (McKevitt et al., 1997). D'autre part, dans le même esprit, les salariés du secteur public ont souvent à cœur la continuité du service (Baker-McClearn et al., 2010 ; Dew et al., 2005). C'est l'une des explications possibles d'un présentéisme plus élevé dans le secteur public (Böckerman & Laukkanen, 2009b). Les interviews réalisés par Dew et al. (2005) dans un hôpital public vont d'ailleurs dans le même sens : l'une des dimensions qui ressort est celui du « sacrifice » consenti par les soignants au profit du service, dans un contexte parfois particulièrement tendu.

Dans l'établissement bancaire, même si aucune variable de la régression n'est directement liée au métier des salariés, une information est donnée par la variable BAC+3/+5. Le niveau de diplôme fait classiquement partie des facteurs démographiques : nous l'aborderons d'ailleurs plus complètement dans la partie qui leur est dédiée. Mais, il peut aussi exprimer la qualité de l'emploi dans lequel est inséré l'individu. Les tâches d'encadrement, plus souvent dévolues à des salariés plus diplômés, n'entrent pas en ligne de compte ici puisque cet effet est capté par une autre variable. Le niveau de diplôme peut refléter la nature des tâches professionnelles, voire le degré d'autonomie dont dispose le salarié. Dans la régression, le coefficient de cette variable est positif et significatif (rapport des chances de 1,7) : les salariés détenteurs d'une licence ou d'un master ont un niveau de présentéisme plus élevé. Or, ils occupent également davantage des postes en responsabilité (chargés d'affaires, juristes, chefs de projets, etc.).

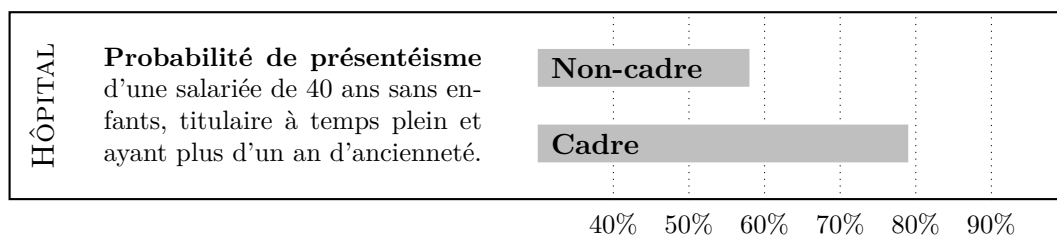
1.2 Responsabilité d'équipe : analyse détaillée

La mise en lumière de la responsabilité d'équipe comme facteur de présentéisme est un élément-clef de cette thèse. Comme l'indique Johns (2010), il reste beaucoup à découvrir sur la façon dont l'organisation est elle-même productrice de présentéisme. D'après nous, les managers de proximité détiennent une place centrale dans ce mécanisme. C'est pourquoi nous voudrions accorder une place particulière à ce point au travers d'une analyse plus détaillée.

1.2.1 Effet propre

Parmi les cadres, certains sont spécifiquement chargés du management d'une équipe de travail. C'est par exemple le cas des directeurs d'agence, au sein de l'établissement bancaire ou des cadres de santé à l'hôpital, qui planifient quotidiennement le travail de leur équipe et prennent les décisions nécessaires au fonctionnement de leur unité. On peut faire l'hypothèse que, à l'instar d'autres catégories d'emploi pour lesquelles l'absence-maladie a de lourdes conséquences (cf. plus haut), la propension au présentéisme est relativement plus élevée pour ces individus en responsabilité d'équipe. Dans les deux entreprises que nous étudions, une variable indiquant si un individu possède une telle charge managériale est incluse dans l'équation d'inflation.

Dans l'établissement bancaire, il s'agit de la variable RESPONSABILITÉ qui indique simplement si le salarié est responsable direct d'une équipe de travail. Il s'agit le plus souvent de directeurs d'agences bancaires. Le coefficient de cette variable est positif et relativement élevé (0,63 c'est-à-dire un rapport des chances de 1,88). Ceci est encore plus vrai pour la variable CADRE, dans la régression du centre hospitalier : son coefficient est le plus élevé de l'ensemble de la régression (1,029 avec un rapport des chances de 2,8). Cette variable est majoritairement composée des cadres de santé, mais comprend également les cadres administratifs et leurs adjoints. Le présentéisme des managers d'équipe semble donc bien être une réalité. L'exemple suivant est frappant :



Dans l'établissement bancaire, ces comportements se développent sans doute plus facilement parmi les salariés qui sont responsables de point de vente (par exemple du réseau bancaire). Étant les seuls cadres de niveau équivalent dans les locaux de l'agence, aucun autre salarié ne peut réellement se substituer à eux. Ils peuvent donc craindre un impact important de leur absence sur le fonctionnement de l'agence, crainte qui les amène à venir travailler malgré la maladie.

Cet argument de l'isolement géographique est moins valable pour le personnel encadrant de l'hôpital, où la plupart des unités sont centralisées en un même lieu. Même si

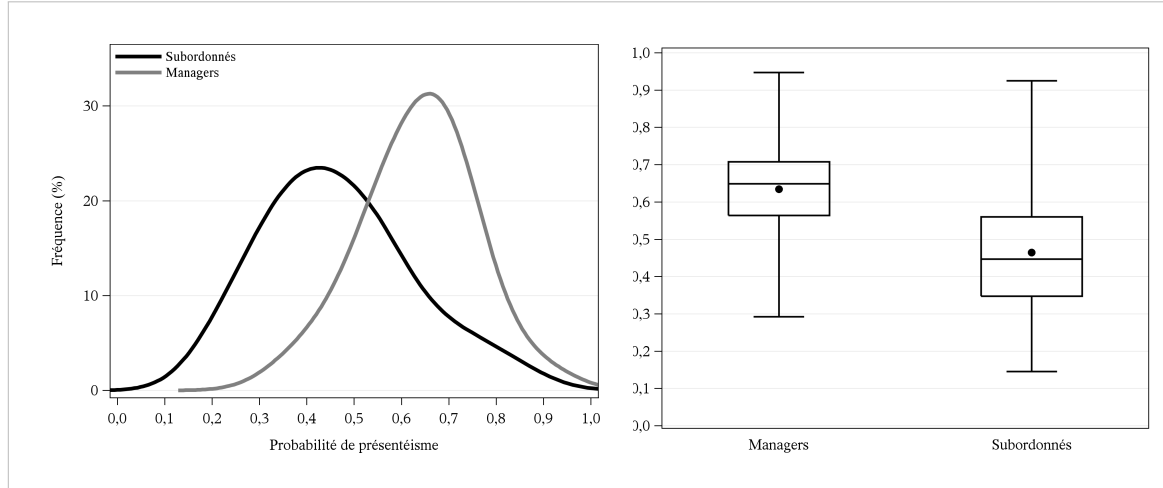
l'isolement « décisionnel » peut être tout aussi important, dans le contexte de l'hôpital, il faut sans doute davantage imputer le présentéisme des cadres aux spécificités intrinsèques de l'emploi qu'ils occupent. Les cadres de santé sont spécifiquement chargés de veiller à la qualité des soins et à leur continuité. En ce sens, leur présence dans l'unité peut donc leur apparaître comme nécessaire, non seulement pour le bon fonctionnement de l'équipe, mais également pour le bien des patients.

Ces résultats concordent largement avec les conclusions d'autres auteurs. Hansen et Andersen (2008) signalent, par exemple, que les comportements présentéistes sont plus fréquents chez les managers. Soit parce qu'ils estiment que leur absence serait trop coûteuse pour l'organisation (Johansen, 2012) : perte ou mécontentement de la clientèle, suspension des tâches quotidiennes en l'attente de décisions, etc. Soit parce qu'ils sont simplement soucieux des subordonnés qu'ils encadrent (Caverley et al., 2007). Nielsen (2008) avance même que le comportement du manager vis-à-vis de l'absence pourrait faire office d'exemple dans son groupe de travail (ce que nous testerons plus loin). Ses décisions à cet égard auraient donc d'autant plus d'importance, étant données les conséquences managériales possibles.

En partant de la régression réalisée précédemment pour l'établissement bancaire, on peut analyser séparément les probabilités de présentéisme des managers et celles des subordonnés. Comme l'atteste le graphique 5.1 (page 195), les différences sont nettes. Dans l'ensemble de la population, 75% des managers ont une probabilité de présentéisme supérieure à 56,2%, tandis qu'environ la même part des subordonnés (75,2%) se situe en-deçà de ce seuil. La probabilité de présentéisme moyenne est de 63,4% pour les managers contre 46,5% pour les subordonnés. En outre, la distribution des probabilités des managers est située très à droite en comparaison à celle des subordonnés.

Les facteurs générant ces différences restent toutefois à identifier. Intuitivement, il semble que les managers et les subordonnés n'aient pas les mêmes raisons d'être présentéistes. On peut légitimement penser que, lorsqu'un manager est malade, le contexte professionnel et la situation de son service (charge de travail, nature de l'équipe, etc.) représentent des éléments décisifs l'incitant plus ou moins à travailler malgré tout. Ces critères ne sont certainement pas prioritaires pour les subordonnés, même s'ils peuvent eux aussi avoir à cœur la continuité du service. Les facteurs explicatifs du présentéisme pourraient en tout état de cause être différents pour ces deux catégories de salariés.

Graphique 5.1 – Managers et subordonnés (Banque en 2008)



L'analyse comparative que nous allons mener dans un premier temps peut répondre à cette interrogation.

Cependant, cette comparaison entre deux catégories de salariés ne prend pas en compte l'intégralité du problème. En effet, une deuxième question émerge autour de la relation entre un manager et ses collaborateurs. Cette relation est constitutive du travail d'équipe : elle permet d'orienter les activités quotidiennes, d'impartir à chacun son rôle, de planifier et d'organiser le travail. D'après nous, il s'agit aussi d'une relation d'influence, même non verbalisée, qui peut se traduire par une forme d'exemplarité du manager. Nous nous intéressons particulièrement ici aux managers de proximité, c'est-à-dire à ceux qui partagent quotidiennement le vécu de leurs équipes. Leur position est très particulière : ils se situent à l'interface entre l'organisation et les individus qu'ils encadrent. À ce titre, leur personnalité, les valeurs dont ils sont porteurs par le discours et le comportement influencent les décisions de leurs collaborateurs. Si Nielsen (2008) a déjà montré que ceci est vrai pour les comportements d'absence, notre objectif est de savoir si ce statut d'exemple joue également s'agissant du présentéisme. Un manager présentéiste, crée-t-il une sorte de norme collective que suivent les subordonnés ? Si sa propension au présentéisme est élevée, les membres de son équipe sont-ils incités à adopter le même comportement ?

1.2.2 Stratégie d'analyse complémentaire

Pour répondre à ces questions, nous reprenons la base de données de l'établissement bancaire. Les informations qu'elle contient permettent de réaliser des appariements entre chaque salarié et son responsable direct, c'est-à-dire, la personne qui l'encadre au quotidien. Cette mise en forme des données se fait au prix de quelques sacrifices sur le nombre d'observations. En effet, les équipes dont le manager a changé en cours d'année ou dont le manager a plus de 45 jours d'absence-maladie en 2008 ont été exclues de l'analyse³. Des 2 574 individus inclus dans la première analyse, demeurent 2 424 salariés, soit 2 067 subordonnés répartis en 357 équipes ayant chacune à sa tête un responsable. La taille moyenne des équipes est donc de 6,8 salariés⁴. Dans la majorité des cas (67%), il s'agit d'agences réparties sur l'ensemble du territoire : par nature elles sont décentralisées et le directeur d'agence y est souvent le seul représentant de la hiérarchie. Le tableau 5.1 donne quelques indications sur les différences structurelles entre la population des managers et celle des subordonnés.

Tableau 5.1 – Managers et subordonnés : statistiques descriptives

	Responsables	Subordonnés	Ensemble
Âge moyen	45,52	39,61	40,48
Nombre moyen d'enfants	1,87	1,23	1,32
Salariés sans enfant	11,48%	37,30%	33,50%
Salariés avec jeune enfant	2,80%	4,69%	4,41%
Salariés titulaires d'une licence	13,17%	13,01%	13,04%
Salariés titulaires d'un master	19,05%	10,55%	11,80%
Ancienneté moyenne	19,99	15,08	15,80
Salariés du réseau commercial	75,91%	70,88%	71,62%
Salariés en CDD		11,18%	9,53%
Salariés sans indemnisation		13,45%	11,47%
Nombre d'individus	357	2067	2424
dont femmes	26,89%	54,67%	50,58%

3. Dès lors, la variable temporelle n'est plus utile pour les managers, puisque ceux-ci travaillent tous à temps plein et sont tous présents sur l'intégralité de l'année 2008.

4. La taille des équipes varie entre 2 à 24 collaborateurs.

La stratégie d'analyse repose sur trois régressions :

- la première régression est réalisée sur l'ensemble de la population. Elle reprend les variables explicatives incluses dans le modèle initial et permet de vérifier que les résultats obtenus restent stables malgré la réduction de la population.
- La deuxième régression porte quant à elle sur la seule population des managers. Elle fournit des informations sur les facteurs de présentéisme propres à cette population. À partir des coefficients de cette régression, on calcule une probabilité de présentéisme pour chacun des 357 managers.
- La troisième régression, effectuée sur la population des subordonnés, intègre ces probabilités de présentéisme des managers sous la forme d'une variable explicative de l'équation d'inflation. Chaque subordonné est apparié à la probabilité de son propre manager, ce qui nous permettra d'évaluer l'existence d'une relation entre leurs comportements respectifs.

1.2.3 *Analyse comparative*

Avant de détailler les enseignements à tirer d'une analyse comparative, nous fournissons quelques informations sur l'ajustement des modèles, notamment celui concernant les managers qui ne porte que sur 357 observations. Malgré ce faible nombre d'individus, nous sommes parvenus à spécifier un modèle satisfaisant : celui qui est présenté est le mieux ajusté aux données⁵. Comme on le constate sur le graphique 5.2 (page 198), celui-ci conduit à un ajustement convenable aux données empiriques, à l'instar du modèle des subordonnés et de celui portant sur l'ensemble de la population. Ce dernier produit par ailleurs des résultats très proches du modèle initial : aucune variation majeure liée au changement de population n'est à signaler.

À partir des sorties économétriques présentées dans le tableau 5.2 (page 198), plusieurs explications peuvent être avancées pour mettre en lumière les contrastes existants entre managers et subordonnés. Comme indiqué plus haut, le simple fait de se sentir responsable d'un ensemble de personnes, dont le travail quotidien dépend de soi, est déjà probablement une raison suffisante pour venir travailler étant malade. Sans doute le manager a-t-il conscience de ce rôle-clef et des corollaires organisationnels et/ou financiers de son absence. Dans l'équation d'inflation du modèle consacré aux managers

5. Malgré une constante d'inflation assez fortement négative, ce qui peut être le signe d'une meilleure adéquation du modèle binomial négatif standard, un test de Vuong significatif (=4,04) montre que le modèle ZINB demeure le mieux ajusté.

Graphique 5.2 – Ajustement des modèles

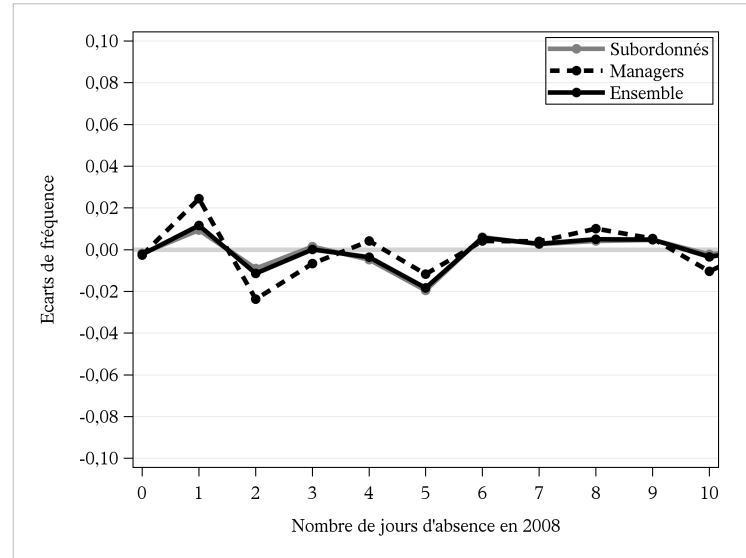


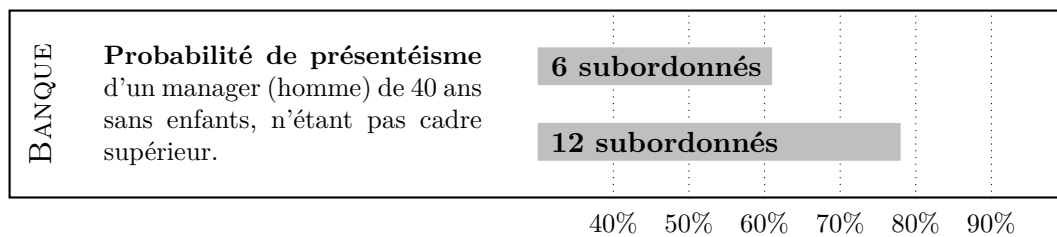
Tableau 5.2 – Estimation des modèles

Paramètre	(1) Managers		(2) Subordonnés		(3) Population	
	Coefficient	Err.-type	Coefficient	Err.-type	Coefficient	Err.-type
CONSTANTE	0,209	0,756	0,952 **	0,302	0,951 ***	0,287
FEMME	0,002	0,244	0,289 ***	0,084	0,265 ***	0,080
AGE	0,048 **	0,017	0,017 ***	0,004	0,018 ***	0,004
NOMBRE D'ENFANTS	-0,385 ***	0,100				
JEUNE ENFANT	1,094 °	0,570				
CDD			0,924 ***	0,162	0,912 ***	0,156
HIÉRARCHIE			-0,388 *	0,195	-0,389 *	0,191
EXPOSITION			-0,112 **	0,041	-0,118 ***	0,032
DISPERSION			0,650 **	0,226	0,658 **	0,227
	1,169 ***	0,340	1,226 ***	0,108	1,245 ***	0,105
CONSTANTE	-2,962 *	1,154	-0,315	0,422	0,067	0,363
FEMME	-0,554 °	0,338	-0,487 ***	0,120	-0,496 ***	0,112
AGE	0,063 **	0,022	0,030 ***	0,007	0,034 ***	0,006
UN SEUL ENFANT			-0,345 *	0,147	-0,333 *	0,133
BAC +3/5			0,602 ***	0,143	0,506 ***	0,130
SANS INDEMNISATION			0,597 **	0,221	0,653 **	0,221
RESPONSABILITÉ					0,666 ***	0,150
MOBILITÉ			-0,625 **	0,222	-0,440 *	0,193
CADRE SUPÉRIEUR	1,412 *	0,550				
TAILLE DE L'ÉQUIPE	0,130 **	0,049	0,012	0,016		
PROB. MANAGER			0,968 *	0,418		
EXPOSITION			-1,646 ***	0,287	-1,663 ***	0,288
Individus	357		2067		2424	

Significativité : <0,1%=*** <1%=** <5%=* <10%=°

(modèle noté 1), les effets de deux variables sont comparables à ceux mis en lumière dans l'ensemble de la population (modèle 3). Le fait d'être un homme contribue à l'augmentation de la probabilité de présentéisme (à un seuil de 10%) et celle-ci progresse également avec l'âge. Le coefficient de la variable d'âge semble avoir plus de poids pour les managers, mais l'erreur-type est plus élevée. Ces deux variables démographiques sont aussi présentes dans l'équation de comptage et ont le signe attendu. Toutefois, les résultats de la régression portant sur les seuls managers apporte deux éléments nouveaux, qui n'apparaissaient pas dans le modèle global.

D'abord, la probabilité de présentéisme des managers progresse avec le nombre de subordonnés dont ils ont la charge. Ce constat est assez intuitif : plus nombreuses sont les personnes qui pâttissent de son absence, plus un manager est susceptible de renoncer à un arrêt-maladie. Les managers des plus grandes unités de travail anticipent en quelque sort un effet de désorganisation plus important que les responsables d'équipes de taille restreinte. À notre connaissance il n'existe pas d'autre article soulignant l'incidence de la taille de l'équipe sur le présentéisme du manager, qui est pourtant fortement marquée dans nos résultats. Le rapport des chances entre un salarié qui encadre 6 personnes et un salarié qui en encadre 12 est de 2,2. Ce qu'illustre l'exemple concret suivant :

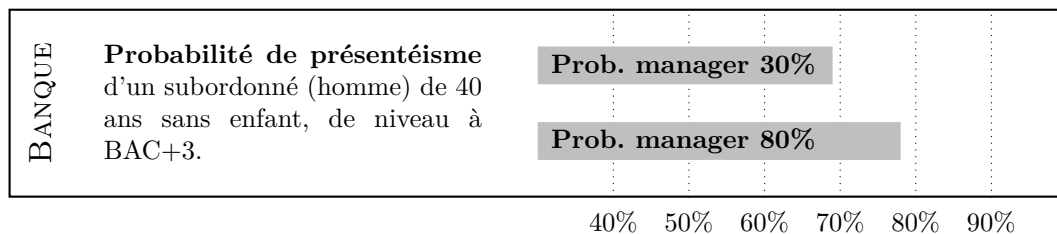


Ensuite, parmi les managers, les cadres supérieurs sont plus présentéistes. Il s'agit principalement ici des membres du comité de direction et des directeurs de groupes (des entités géographiques englobant un ensemble d'agences). Ces personnes (11,2% des managers) occupent des postes stratégiques : leurs décisions sont cruciales pour les orientations de l'entreprise. *A priori*, ils ont conscience de leur importance et ne s'autorisent pas ou très peu d'absences-maladie. De plus, les plannings de travail de ces salariés sont peut-être également moins souples. Par exemple, la visite d'un directeur de groupe dans les agences qu'il dirige est préparée de longue date et tient une place importante dans les orientations commerciales des points de vente. En cas de maladie, l'annulation d'une telle visite peut s'avérer problématique (tout comme une réunion du comité de direction) et mener le cadre supérieur au présentéisme.

1.2.4 *Relation manager-subordonné*

Les estimations suggèrent donc que l'attitude d'un manager vis-à-vis du présentéisme est influencée par sa position dans l'entreprise, la portée de ses décisions, mais également par la taille de son équipe. Cependant, les exemples proposés ci-dessus laissent aussi entendre que le comportement de présentéisme est une réponse aux attentes de l'organisation. Dès lors, cela pourrait signifier que les « injonctions » au présentéisme se transmettent progressivement tout au long de l'échelle hiérarchique et finissent par toucher l'ensemble des salariés. Il est alors intéressant de se demander s'il existe une « diffusion » de la propension au présentéisme entre un manager et ses subordonnés.

C'est pour répondre à cette interrogation que la probabilité de présentéisme des managers a été introduite comme une variable explicative dans l'équation d'inflation des subordonnées (modèle 2), selon la stratégie d'analyse présentée ci-dessus. Le coefficient est significatif et estimé à une valeur de 0,968. La probabilité de présentéisme du manager est la variable ayant le plus de poids dans la régression, plus encore que le coût de l'absence (variable SANS INDEMNISATION). L'attitude d'un manager semble donc avoir une réelle influence sur son équipe, ce qu'illustre l'exemple suivant :



Au premier abord, l'incidence ne semble pas spectaculaire. Mais cet effet rejaille sur l'ensemble d'une équipe, dont le nombre de membres est potentiellement élevé. Il s'ensuit que le manager dispose d'un pouvoir d'influence sur les normes informelles de l'équipe ou, pour le dire avec les mots de Johns (2010), sur la culture de présentéisme de l'équipe. Au travers de son propre comportement, le responsable peut soit positionner les contraintes organisationnelles par-dessus tout, y compris l'attention à sa propre santé, ou au contraire encourager des attitudes équilibrées vis-à-vis de l'absence et du présentéisme. Si l'on considère les effets délétères du présentéisme sur la santé des salariés, cette simple précaution est déjà une première réponse pour éviter des niveaux d'absences très élevés dans les années qui suivent.

2 Facteurs individuels de présentéisme

La décision d'un salarié de se présenter à son travail lorsque se produit un événement de santé pouvant légitimer son absence, intervient certainement au terme d'un jugement dans lequel le contexte de travail tient une place-clef. Néanmoins, ce choix ne se pose pas de façon uniforme selon que l'individu est une femme, un senior ou que sa situation personnelle est marquée par une lourde charge familiale. C'est pourquoi, après avoir envisagé les aspects spécifiquement professionnels, nous abordons les facteurs individuels du présentéisme. À nouveau, une place particulière est accordée à l'une des variables qui nous semble éclairer l'analyse d'une lumière nouvelle. Il s'agit cette fois de l'âge, que nous détaillerons dans un second temps.

2.1 Effets des variables liées à l'individu

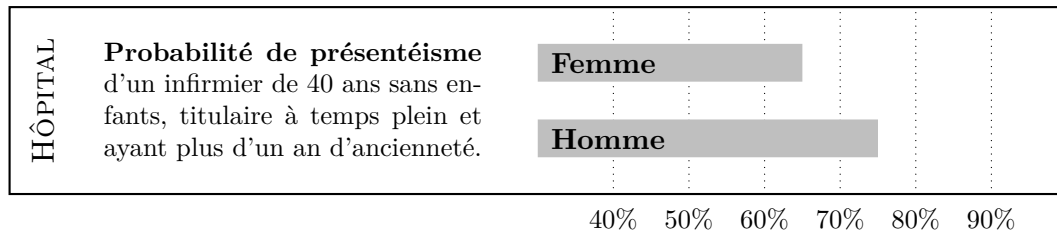
Comparativement aux facteurs liés à l'emploi, la littérature laisse peu de place aux variables démographiques. Celles-ci sont souvent appréhendées comme des variables de contrôle, nécessaires dans l'analyse, mais secondaires pour l'interprétation. À titre d'exemple, nous n'avons relevé aucune étude sur le présentéisme spécifiquement dédiée à une analyse différenciée hommes/femmes. Cet aspect ainsi que l'influence de la situation familiale et du niveau de diplôme des individus sont évoqués dans cette première partie.

2.1.1 *Sexe*

Au premier abord, peu d'éléments permettent de poser une hypothèse sur la relation entre le sexe et le présentéisme. La littérature est elle-même plutôt partagée : certains travaux ne détectent aucune différence entre hommes et femmes (Johansen, 2012 ; Johansson & Lundberg, 2004) et quelques études penchent en faveur d'un comportement plus typiquement féminin (Aronsson et al., 2000 ; Leineweber et al., 2010). Des pistes contradictoires apparaissent également dans les études statistiques nationales : si les femmes déclarent davantage renoncer à se soigner pour des raisons financières (Dourgnon et al., 2012), elles sont cependant globalement plus proches des filières de soins, y engagent plus de dépenses et semblent plus attentives à leur santé (Montaut, 2010). Ceci suggère que le seuil de déclenchement de l'absence-maladie interviendrait plus tôt

chez les femmes, lorsqu'elles en ont les moyens financiers. On pourrait donc supposer que leur propension au présentéisme est également plus faible.

En tout état de cause, nos conclusions, aussi bien dans l'établissement bancaire que dans le centre hospitalier, orientent vers un risque de présentéisme plus faible pour les femmes, l'effet étant de la même ampleur dans les deux cas (le rapport des chances est de 0,64). Les femmes seraient donc moins susceptibles de venir travailler étant malade, ce qui confirmerait le résultat de Demerouti et al. (2009), dans un contexte hospitalier. On peut illustrer cela en calculant les probabilités de présentéisme pour une situation de départ précise :



D'après nous, trois éléments permettraient d'expliquer ce risque de présentéisme plus faible pour les femmes. Premièrement, pour en revenir à l'article de Montaut (2010), on peut avancer que les femmes sont effectivement plus attentives à leur propre santé que les hommes et donc plus promptes à se soigner. Comme nous l'avons dit, leur seuil de déclenchement d'une absence-maladie serait atteint plus tôt. Cette assertion semble cohérente avec les statistiques nationales sur les dépenses de santé (Montaut, 2010) : les femmes consultent davantage leur généraliste que les hommes, en particulier avant 55 ans. Par ailleurs, elles recourent davantage à des actions de dépistage de maladies et sont plus vigilantes à leur équilibre alimentaire. Une autre confirmation provient de l'article de Shmueli (2003). Dans une analyse comparative de plusieurs méthodes d'auto-évaluation de la santé, il indique que les femmes déclarent généralement un état de santé perçu inférieur à leur état de santé réel, tandis que les hommes ont tendance à surestimer leur niveau de santé⁶. Devant une même maladie, les femmes seraient donc potentiellement plus promptes à opter pour l'absence-maladie que les hommes.

Deuxièmement, la relation entre sexe et présentéisme peut être dû au contexte organisationnel. C'est l'argument défendu par Simpson (1998). Dans son article, les candidats à des opportunités professionnelles (accès à des postes de management) cherchent à

6. Cette étude est réalisée sur un échantillon représentatif de la population israélienne comportant 2 030 individus.

donner des gages à leurs supérieurs hiérarchiques en vue d’une éventuelle promotion, par exemple en pratiquant des dépassements d’horaires. D’après Simpson (1998), ces stratégies basées sur le présentéisme seraient davantage le fait de jeunes hommes placés dans un contexte de concurrence.

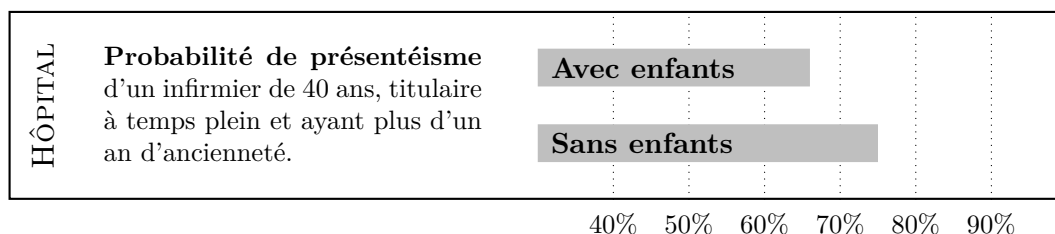
Troisièmement, dans une définition élargie du présentéisme, certaines tâches familiales (notamment liées à la garde des enfants) pourraient empêcher les femmes de pratiquer des dépassements horaires dans la soirée ou d’abandonner un jour de congé pour le travail. L’effet de la charge familiale (cf. ci-après) est en partie prise en compte dans notre régression par le nombre d’enfants. Néanmoins, cela ne signifie pas que la variable SEXE ne puisse capter une part des effets inhérents à l’équilibre entre vie privée et vie professionnelle. En effet, les femmes doivent parfois endosser un double rôle : après la journée de travail professionnel, une autre « journée », de travail familial cette fois, débute au domicile. Ces deux dimensions peuvent à l’occasion entrer en conflit. En cas de maladie, par exemple, délaissier les nécessités de la vie familiale n’est pas envisageable et l’absence-maladie peut constituer un moyen de faire face malgré tout. Ce choix est sans doute moins prégnant pour les hommes.

2.1.2 Charge familiale

Cet effet du sexe est donc directement en lien avec la question de la charge familiale. Les régressions effectuées dans chacun de nos deux contextes intègrent une variable permettant d’apprécier l’importance de cette charge familiale. Au centre hospitalier, par exemple, une variable muette indique si l’individu a des enfants ou pas. Intuitivement, on s’attend à ce que les salariés sans enfants soient plus présentéistes, en estimant par exemple que, pour ces salariés, les conflits entre vie familiale et vie professionnelle sont moins fréquents et qu’ils peuvent davantage se consacrer à leur carrière. Cependant, la littérature sur le présentéisme est dominée par la mise en lumière d’un effet inattendu pour les salariés ayant des enfants. Aronsson et al. (2000), puis Hansen et Andersen (2008) indiquent que le présentéisme augmenterait avec l’importance de la charge familiale. D’après eux, ceci pourrait être dû à une préférence pour des activités professionnelles plus valorisantes que les tâches familiales. En dehors de ces deux travaux, la relation n’est quasiment pas étudiée.

Les résultats de nos estimations ne vont pas dans ce sens et tendent à confirmer une hypothèse plus conventionnelle. Dans le cas du centre hospitalier, les salariés sans

enfants semblent être plus présentéistes que les autres (rapport des chances de 1,5)⁷. Ce qui est illustré par l'exemple suivant :



D'autre part, dans la régression de l'établissement bancaire, le fait d'avoir un enfant implique une diminution de la probabilité de présentéisme⁸ (rapport des chances de 0,66). Néanmoins, cet effet n'est plus significatif pour les salariés ayant plus d'un enfant. Le résultat est donc à considérer avec prudence.

L'interprétation de ces coefficients s'accompagne d'un problème statistique. Dans les deux cas, les variables utilisées sont construites à partir du même type de données : le nombre d'enfants du salarié. Il ne s'agit pas là du nombre d'enfants à charge : aucune indication n'est donnée sur l'âge des enfants. Il est donc difficile d'expliquer clairement les variables incluses puisqu'elles recouvrent potentiellement des situations familiales assez hétérogènes. Au minimum, elle fournissent une information sur l'existence possible de conflits entre vie privée et vie professionnelle.

Il reste cependant peu évident d'expliquer cette relation négative entre le fait d'avoir des enfants et le présentéisme. D'une part, les salariés sans enfants, qui ont une probabilité de présentéisme plus élevée (au centre hospitalier), peuvent sans doute se permettre un investissement plus total dans les tâches professionnelles. Ces individus se situeraient plutôt à une époque de leur vie axée sur l'insertion professionnelle (le « temps de la carrière »). Contrairement aux personnes ayant un contexte familial prenant (se trouvant dans le « temps de la famille »), les personnes sans enfants ont moins à arbitrer entre les deux dimensions évoquées plus haut (familiale et professionnelle).

D'autre part, si l'on appréhende la santé de l'individu de façon globale, on peut penser qu'un contexte familial chargé peut avoir des effets négatifs en termes de stress ou

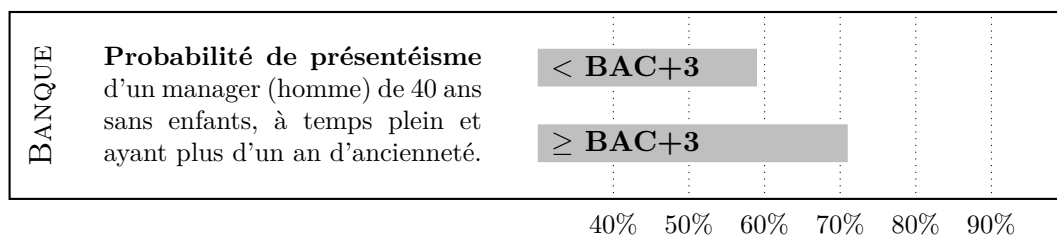
7. Nous avons opté pour cette variable muette différenciant les salariés sans enfants des autres pour sa forte significativité. Toutefois, nous avons également testé l'intégration brute du nombre d'enfants sous la forme d'une variable quantitative. Il existe un effet linéaire : plus le nombre d'enfants est élevé, plus la probabilité d'être présentéiste diminue. Cette variable quantitative est significative à un seuil de 5%.

8. Et ce indifféremment selon le sexe : l'hypothèse d'un effet différencié pour les hommes et les femmes a été testée, mais les différences sont très faibles et non significatives.

de bien-être personnel. À certains moments, cet élément peut avoir du poids dans l'équation et faire pencher la décision d'un individu malade dans le sens de l'absence plutôt que du présentéisme. En cas de maladie, un salarié peut donc être conduit à prendre des jours d'absence, certes pour se soigner, mais également pour pouvoir faire face au retour des enfants, le soir. Dans l'établissement bancaire, la relation entre un enfant unique et une probabilité de présentéisme réduite, s'explique très bien en ce sens. L'effet de désorganisation peut en effet être maximal pour le premier enfant : recherche d'un équilibre familial, sommeil insuffisant, choix d'un mode de garde, etc. Pour une personne malade, ce contexte difficile peut peser dans la décision d'absence ou de présence.

2.1.3 Niveau de diplôme

La littérature est très peu prolixe sur les relations entre niveau de diplôme et présentéisme. Taloyan et al. (2012) indiquent que les salariés détenteurs de diplômes intermédiaires seraient plus à risque que ceux ayant un niveau de diplôme très faible ou très élevé. En ce sens, l'inclusion d'une variable indiquant si le salarié est détenteur d'un diplôme de licence ou d'un master permet d'apporter un élément nouveau sur le présentéisme. C'est ce qui a été fait dans le cadre de l'établissement bancaire pour lequel nous disposons de l'information. Le fait d'avoir un diplôme supérieur ou égal à la licence a un effet significatif sur la probabilité de présentéisme et constitue même un facteur assez conséquent. Le rapport des chances est de 1,7. En termes de probabilité cela peut-être illustré par l'exemple suivant :



La recherche d'éclaircissements à propos de cet effet peut bénéficier d'un rapprochement avec la question du renoncement aux soins médicaux, qui est une thématique importante en économie de la santé. Certes, le présentéisme, c'est-à-dire le renoncement à un arrêt-maladie, n'implique pas systématiquement un renoncement aux soins, puisque dans certains cas le soin peut avoir lieu parallèlement, en dehors des heures de travail. Néanmoins, des mécanismes communs peuvent être décelés, notamment sur le

niveau de diplôme. Dourgnon et al. (2012) signalent que les individus ayant effectué des études supérieures renoncent davantage aux soins médicaux que ceux ayant un faible diplôme (BAC ou inférieur). Ce renoncement concerne les soins dentaires et optiques, mais également les consultations de médecins généralistes ou spécialistes. Le présentéisme pourrait donc partiellement s'expliquer par une barrière qu'opposent certains salariés au soin. Cette barrière n'est *a priori* pas financière, puisque le salaire annuel des plus diplômés est plus élevé, en moyenne⁹. Elle doit donc relever d'une attitude du salarié lui-même. Comme nous l'avons vu précédemment, le niveau de diplôme est également un déterminant de la nature des tâches exercées. Les plus diplômés renonceraient alors davantage aux soins pour privilégier un emploi qui les valorise, une carrière dans laquelle leur niveau d'étude permet d'envisager une progression.

Par ailleurs, le niveau de diplôme reste également une indication pertinente de l'origine sociale des individus. Ainsi, dans un rapport récent du ministère de l'enseignement supérieure¹⁰, 67% des enfants d'enseignants et 64% des enfants de cadres entrés en sixième en 1995 ont obtenu un diplôme de niveau supérieur ou égal à BAC+3. Seuls 19% des enfants d'ouvriers et 12% des enfants d'ouvriers non-qualifiés sont dans ce cas. Le niveau de diplôme renseigne donc d'une certaine façon sur les normes et les valeurs véhiculées dans les classes sociales. Les plus diplômés sont issus de milieux où l'on mise davantage sur la poursuite d'études et où l'on est en mesure d'y investir financièrement, afin de garantir une meilleure insertion professionnelle. Un retour sur investissement est alors attendu au cours de la carrière, qui appelle une plus grande implication dans le travail et un développement du potentiel acquis en cours d'études. Nos résultats montrent que cela se joue parfois au prix de sacrifices dans d'autres domaines, par exemple celui de la santé au travers du présentéisme.

2.2 Âge et présentéisme : analyse détaillée

La variable d'âge est sans doute l'une des plus intéressantes abordées dans cette thèse. L'allongement de la durée de vie au travail, associée à la dégradation naturelle de la santé avec le vieillissement, interroge légitimement les entreprises sur les moyens d'anticiper les effets de telles évolutions et sur les politiques à mettre en œuvre pour leurs

9. 33 694 euros contre 31 810 euros, en incluant le salaire de base et la part variable.

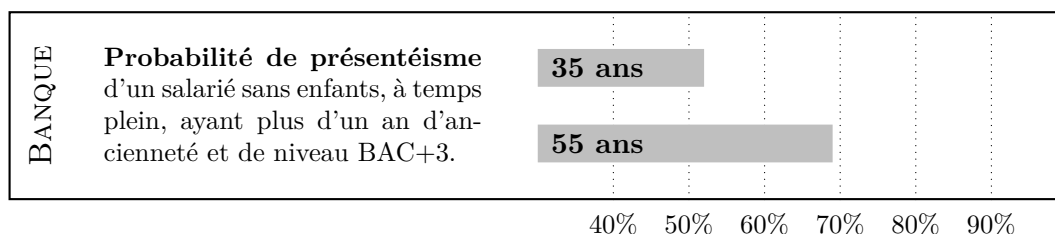
10. « Les parcours dans l'enseignement supérieur : devenir après le baccalauréat des élèves entrés en sixième en 1995 », note d'information n° 12 du 5 juin 2012, disponible en ligne à l'adresse suivante : <http://www.enseignementsup-recherche.gouv.fr>.

salariés seniors. À cet égard, les comportements d'absence des seniors constituent une inconnue majeure : la question de l'âge se situe au carrefour entre absence-maladie et présentéisme. Bien que la prochaine section soit spécifiquement consacrée à la question de l'absence à la lumière du modèle ZINB (section 3), l'analyse détaillée de la variable d'âge y empiète déjà. De nombreuses études soulignent en effet qu'il y a une corrélation positive entre l'âge et le niveau d'absence-maladie. Mais tout n'est plus aussi simple lorsqu'il est mis en évidence que le présentéisme est, lui aussi, plus fréquent à mesure que l'on vieillit.

2.2.1 Effet propre

Au cours du chapitre 2, il a été souligné que, lorsque l'absence est appréhendée selon sa gravité, les résultats de la littérature orientent vers une progression du niveau d'absence avec l'âge (Donders et al., 2012 ; Frick & Malo, 2008 ; Missègue, 2007). Une nouvelle fois, ceci est corroboré d'un point de vue national par les dépenses de santé des seniors (Montaut, 2010). Pas de consensus en revanche concernant le présentéisme, même si certains travaux penchent pour une augmentation de ce comportement avec l'âge (Dew et al., 2005 ; Leineweber et al., 2010).

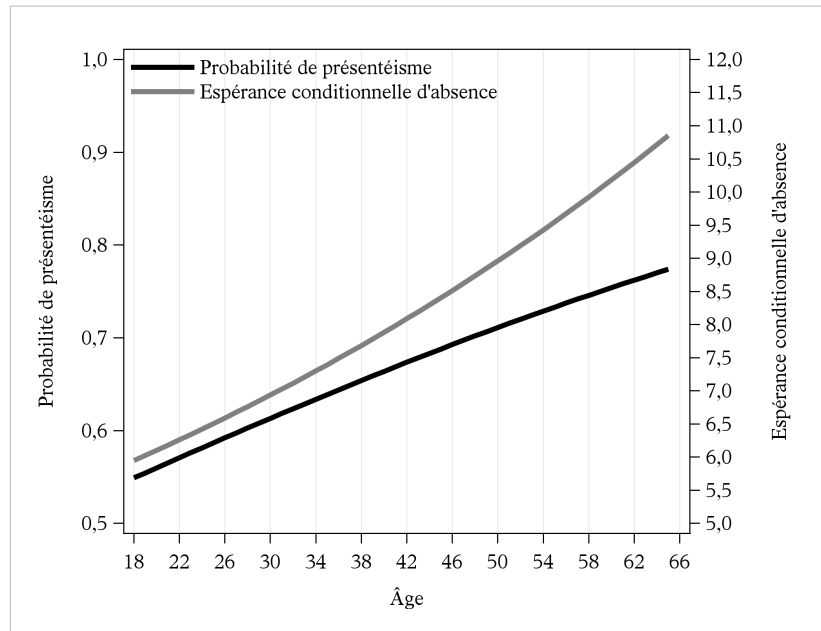
Dans nos deux cas d'entreprises, l'âge apparaît à la fois comme un facteur important dans les comportements présentéistes et comme un facteur aggravant de l'absence : son coefficient positif dans l'équation de comptage indique bien son influence sur l'accroissement de l'espérance d'absence. Au total, les salariés les plus âgés auraient donc des absences plus longues, mais « compenseraient » ce rallongement par le présentéisme à d'autres occasions. Dans l'établissement bancaire, pour un écart de 20 ans, le multiplicateur de rapport des chances est de 2,08 (de 1,55 au centre hospitalier). Voici un exemple en termes de probabilités :



Cette particularité de l'âge d'augmenter à la fois le présentéisme et l'absence est illustrée par le graphique 5.3 (page 208), produit sur la base de la régression effectuée

au centre hospitalier. L'effet marginal de l'âge est positif dans l'équation de comptage comme dans l'équation d'inflation. La figure illustre la progression de l'espérance conditionnelle d'absence et de la probabilité de présentéisme lorsque l'âge augmente, toutes choses égales par ailleurs¹¹. Deux effets jouent donc en sens inverse, ce qui pourrait se manifester par une stagnation apparente du niveau empirique d'absence : la santé se dégrade avec l'âge ce qui rallonge l'absence lorsqu'elle a lieu, mais le présentéisme augmente également, ce qui conduit le salarié à renoncer à certains arrêts-maladie.

Graphique 5.3 – Âge et présentéisme (Hôpital en 2010)



Ce double effet peut se traduire en termes statistiques. Examinons le cas concret de deux salariés, l'un de 30 ans, l'autre de 50 ans mais dont les autres variables sont identiques. L'espérance conditionnelle d'absence du premier est de 6,93 contre 8,96 pour le second. Il s'agit de l'effet naturel de la dégradation de la santé avec l'âge. Parallèlement, le salarié de 30 ans a une probabilité de présentéisme de 61,3% contre 71,1% pour le salarié de 50 ans. En effet, la propension à venir travailler étant malade, augmente également en vieillissant. Au total, il est possible de calculer une espérance d'absence globale, c'est-à-dire l'espérance du modèle ZINB. Elle est de 2,68 pour le

11. Ici, il s'agit du cas d'un infirmier titulaire de 40 ans, travaillant à temps plein et de plus d'un an d'ancienneté.

premier salarié et de 2,59 pour le second¹². En d'autres termes, même si, en apparence, il semble n'y avoir qu'une différence très faible entre les deux salariés, cette similarité est due en réalité à deux effets qui se compensent statistiquement.

Plusieurs explications peuvent être avancées pour justifier que les salariés les plus âgés s'autorisent moins que les autres à être absents en cas de maladie. D'abord, leur employabilité est plus faible : craignant pour leur emploi ou leur réemploi en cas de licenciement, les salariés les plus âgés pourraient souhaiter s'afficher comme pleinement productifs. Présentant toutefois objectivement un risque d'absence maladie plus important que les salariés plus jeunes, les collaborateurs plus âgés pourraient être amenés à vouloir le compenser en venant plus souvent que les autres travailler malgré la maladie. Ce type de raisonnement est d'ailleurs présent dans les déclarations de salariés au cours d'entretiens (Dew et al., 2005).

Une seconde explication est à chercher dans leur attitude morale vis à vis de l'absence, différente des autres collaborateurs, en particulier les plus jeunes. De nombreux travaux font état de conflits entre les classes d'âges sur les valeurs au travail : d'après Grima (2007, p.29), les conflits intergénérationnels procèdent « d'une conception de la réalité du travail différente fondée sur, à la fois une empreinte historique différente, mais aussi des règles institutionnelles distinctes dans la relation au travail ». Il ne peut être exclu que l'on observe ici une expression claire de ces différences. Les salariés les plus âgés se distingueraient des plus jeunes par une propension plus forte à refuser l'absence, mais une fois la maladie réellement déclarée, ils se voient rattrapés par d'autres réalités (un état de santé plus fragile) qui contribuent cette fois à la rallonger.

En troisième lieu, l'incidence de l'âge sur le présentéisme, mais également sur l'absence-maladie, pourrait s'expliquer par une répartition différente des salariés les plus âgés dans la structure de l'entreprise. Autrement dit, il pourrait y avoir des interactions entre l'âge et d'autres variables explicatives introduites dans le modèle. Il est possible d'éclaircir ce point par une analyse complémentaire destinée à repérer ces interactions.

2.2.2 *Stratégie d'analyse complémentaire*

Afin d'en savoir plus, la population du centre hospitalier (strictement identique à celle de la première régression) est divisée en quatre quarts selon les valeurs quartiles de

12. Rappelons que l'espérance globale $\mu = \lambda \times (1 - p)$, avec λ l'espérance conditionnelle d'absence et p la probabilité de présentéisme.

l'âge. Un tel découpage nous permet d'obtenir des catégories équilibrées en nombre d'individus et a, en outre, du sens pour l'interprétation :

- **1^{er} quartile** : salariés âgés de 18 à 29 ans : la borne supérieure correspondant à peu près à l'âge du premier enfant en France, soit 28,1 ans en 2010 selon l'INSEE ;
- **2^e quartile** : salariés âgés de 29 à 39 ans : la borne supérieure correspondant à l'âge limite de la maternité ;
- **3^e quartile** : salariés âgés de 39 à 48 ans ;
- **4^e quartile** : salariés âgés de 48 ans à 65 ans, soit la catégorie des seniors.

Une description des quatre sous-populations est proposée dans le tableau 5.3 (page 211). La stratégie consiste à effectuer une seule régression, sur l'ensemble de la population, en mesurant les interactions entre les variables explicatives et les catégories d'âge. Pour faciliter l'interprétation des résultats, nous avons opté pour un codage cumulatif des catégories d'âge. Ceci permet d'interpréter les coefficients non pas comme un simple effet d'une variable selon la catégorie d'âge d'un salarié, mais d'après la variation d'un coefficient qu'entraîne le passage d'un quartile au quartile supérieur.

Dans la pratique, trois variables muettes M_1 , M_2 et M_3 , exprimant trois passages possibles entre quartiles ont d'abord été codées comme indiqué dans le tableau suivant :

	M_1	M_2	M_3
Quartile 1	0	0	0
Quartile 2	1	0	0
Quartile 3	1	1	0
Quartile 4	1	1	1

Ce type de codage ne peut s'opérer commodément qu'en choisissant pour population de référence la première ou la dernière modalité : nous avons opté pour la catégorie des plus jeunes. Toutes les variables ont donc été intégrées à la régression à la fois additivement (c'est-à-dire séparément) et multiplicativement de chaque muette (pour mesurer les effets d'interaction). Après élimination des éléments non significatifs, on obtient la sortie brute présentée en annexe 6 (page 238).

Cependant, pour une meilleure compréhension des résultats, il est possible de reconstituer les coefficients de chaque catégorie d'âge. Le tableau 5.4 (page 211) présente des coefficients reconstitués tenant compte des variations liées au passage d'un quartile au quartile supérieur. Prenons un exemple : la constante de l'équation de comptage, dont

Tableau 5.3 – Âge et présentéisme : statistiques descriptives

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Nombre moyen d'enfants	0,17	1,33	1,84	1,92
Salariés sans enfant	86,22%	26,05%	12,62%	13,38%
Salariés avec jeune enfant	4,50%	10,48%	1,66%	0,10%
Salariés en couple	21,99%	71,99%	72,70%	69,63%
Ancienneté moyenne	2,97	8,95	16,95	24,92
Personnel administratifs	7,82%	11,66%	10,47%	15,04%
Personnel de soins	72,83%	71,89%	69,37%	61,91%
Salariés titulaires	17,01%	70,23%	85,52%	88,57%
Salariés cadres de santé	1,27%	2,25%	5,28%	7,03%
Salariés en CDD	52,10%	11,56%	3,62%	3,81%
ASH	24,73%	5,88%	9,20%	12,79%
Aide-soignant	14,17%	19,29%	26,52%	21,97%
Infirmier	30,40%	42,41%	26,22%	19,73%
Nombre d'individus	1023	1021	1022	1024
dont femmes	79,96%	80,12%	79,45%	77,15%

Tableau 5.4 – Estimation du modèle (coefficients reconstitués)

Paramètre	Quartile 1	Quartile 2	Quartile 3	Quartile 4
CONSTANTE	1,953	2,144	2,144	2,341
JEUNE ENFANT (< 1 an)	0,591	0,591	0,591	0,591
ASH	0,249	0,249	0,249	0,249
AIDE-SOIGNANT	0,391	0,391	0,391	0,391
EXPOSITION	-0,100	-0,100	-0,100	-0,100
DISPERSION	1,330	1,330	1,330	1,330
CONSTANTE	0,863	1,800	2,132	2,132
FEMME		-0,526	-0,526	-0,526
SANS ENFANTS	0,470	0,470	0,470	0,470
INFIRMIER	0,299	0,299	0,299	0,299
CADRE		1,035	1,035	1,035
CDD	0,642	0,642	0,642	0,642
TITULAIRE		-0,414	-0,414	-0,414
ANCIENNETÉ FAIBLE (< 1 an)	0,394	0,394	0,394	0,394
MOBILITÉ		-0,314	-0,314	-0,314
EXPOSITION	-1,087	-1,087	-1,087	-1,087

les variations sont ici imputables à l'effet propre de l'âge sur l'absence, est de 1,953 au départ. Le passage au 2^e quartile implique une variation de cette constante de 0,191 ($1,953+0,191=2,144$)¹³. Le passage du 2^e au 3^e n'entraîne pas d'évolution significative de la constante (on reste à 2,144), mais le passage du 3^e au 4^e génère une nouvelle variation de 0,197 ($2,144+0,197=2,341$).

2.2.3 Résultats de l'analyse complémentaire

Lecture des résultats. Un autre commentaire s'impose sur la présentation des résultats dans le tableau 5.4 (page 211). Lorsque les variables sont intégrées sous forme simple dans la régression, elles captent uniquement l'effet de la population de référence, c'est-à-dire celle du premier quartile (colonne « Quartile 1 »). Lorsqu'elles sont multipliées par les muettes, elles rapportent la variation du coefficient liée au passage entre :

- le 1^{er} et le 2^e quartile pour M_1 ;
- le 2^e et le 3^e quartile pour M_2 ;
- le 3^e et le 4^e quartile pour M_3 .

Du point de vue de l'ajustement, l'identification des interactions avec l'âge ne présente pas de gain important, puisque les variables intégrées sont approximativement les mêmes : la log-vraisemblance passe de -6362 dans le modèle initial à -6358 dans celui-ci. Toutefois, l'analyse des effets de l'âge sur les comportements de présentéisme et d'absence gagne en finesse. Les autres statistiques d'ajustement présentées pour le modèle initial restent valables ici.

Equation d'inflation. Suite à cette nouvelle régression, passons en revue les principales informations fournies par l'équation d'inflation du modèle. D'une part, le présentéisme évolue moins linéairement avec l'âge. Lorsqu'on passe du premier au deuxième quartile, la rupture est franche : le coefficient passe alors 0,863 à 1,8. Ensuite, au franchissement du troisième quartile, bien que la valeur du coefficient augmente significativement (1,8 à 2,132), le pas est un peu moins marqué. L'augmentation nette du présentéisme autour de l'âge de 30 ans peut être lié à une stabilisation de la vie professionnelle et familiale ou à l'accès à des postes de travail plus satisfaisants.

13. Les variations brutes des coefficients peuvent être lues dans le tableau en annexe 6 page 238.

Néanmoins, pour les femmes, une baisse relative du présentéisme intervient à cette période de la vie. Jusqu'à 30 ans, elles semblent davantage se consacrer à leur insertion professionnelle (temps de la carrière). Elle peuvent s'impliquer davantage dans leur travail et en faire une priorité. Mais, passé ce seuil, s'ouvre une nouvelle époque, notamment celle de la maternité et de l'accroissement de la charge familiale. La rupture marquée dans leur propension au présentéisme le confirme : elles doivent trouver un nouvel équilibre entre leur profession et leur vie familiale.

Sur la thématique du contenu de l'emploi, le métier d'infirmier est durablement marqué par le fait de travailler étant malade : il n'y a pas de rupture significative dans la propension au présentéisme au cours de la carrière. Ceci confirme un comportement lié aux spécificités professionnelles plus qu'à une question d'âge. Les cadres quant à eux sont davantage présentéistes au passage de la première à la deuxième catégorie. Ceci peut naturellement être dû au faible nombre de cadres dans le premier quartile. En plus, la variable intègre à la fois les cadres (administratifs et de santé) et leurs adjoints. Or les adjoints constituent la quasi-totalité de la première catégorie : 92% des cadres du premier quartile sont en réalité des adjoints. L'effet obtenu peut donc aussi être en rapport avec un accès à un poste en responsabilité directe (passage d'un poste d'adjoint à un poste de cadre).

Le même type de remarque vaut pour les salariés titulaires. Le passage au deuxième quartile est marqué par une très forte augmentation du nombre de titulaires (17,01% dans le premier, contre 70,23% dans le deuxième quartile.). Il est donc *a priori* normal que l'effet ne soit perceptible qu'à partir de ce seuil. Néanmoins, les titulaires du premier quartile sont principalement des infirmiers ou assimilés¹⁴ (à 67%), dont la part relative est plus faible dans les autres classes d'âge. Puisque cette profession est également connue pour son présentéisme, on peut penser que l'effet pur (et négatif) du fait d'être titulaire est compensé par la nature des tâches exercées.

Par ailleurs, quoique l'effet de la variable CDD ne se démente pas avec l'âge, il n'y a pas d'augmentation du présentéisme pour les salariés en CDD les plus âgés. On aurait pu s'attendre à ce que les salariés seniors en CDD s'adonnent plus au présentéisme que les autres, puisqu'ils ont davantage à craindre pour leur employabilité. Comme les autres, les seniors en CDD se soucient de l'avenir de leur emploi et sont incités au

14. Y compris les sages-femmes et les puéricultrices.

présentéisme, mais sans doute les limites de leur santé et l'incidence de maladies plus graves les empêchent-elles de « sortir du lot ».

Équation de comptage. Pour terminer cette analyse complémentaire, nous anticipons sur la partie à venir, consacrée à l'absence et à l'absentéisme. D'une part, l'effet propre de l'âge conforte l'idée d'une augmentation du nombre de jours d'absence à mesure que les salariés vieillissent. L'effet de l'âge s'applique de manière relativement constante. On peut situer un premier seuil à l'âge de 30 ans : comparativement aux plus jeunes, les personnes d'âge intermédiaire (de 29 à 48 ans) ont une espérance d'absence légèrement plus élevée. Puis, un palier supplémentaire est franchi lorsque l'on passe dans la catégorie des séniors : il y a alors une autre augmentation significative du niveau d'absence. L'importance de l'effet peut paraître moindre, mais rappelons que nous n'avons retenu ici que les absences de courte durée (moins de 45 jours).

Les variables de l'équation de comptage portant sur les métiers exercés demeurent stables selon l'âge. Concernant les métiers d'ASH et d'aide-soignant, l'effet demeure identique quelle que soit la classe d'âge étudiée. *A priori*, ce résultat paraît surprenant. On s'attendrait à ce que la pénibilité de ces emplois conduise les salariés à s'absenter plus à mesure que leur carrière progresse. Or, bien que le niveau d'absence soit plus élevé pour l'ensemble du personnel exerçant ces métiers, l'espérance d'absence ne progresse pas avec l'âge. L'une des interprétations possibles serait que les salariés les plus âgés aient progressivement acquis une expérience leur permettant d'exécuter les gestes professionnels en limitant les conséquences sur leur propre santé. Mais il est également envisageable que, chez les salariés les plus jeunes, un phénomène parallèle contribue à l'augmentation de l'espérance d'absence, donnant une impression de stabilité avec l'âge. Ici, l'absentéisme (absences injustifiées) pourrait être un élément explicatif.

La variable JEUNE ENFANT ne manifeste aucune interaction avec l'âge. Dans l'absolu, il est raisonnable de penser que la maternité et l'arrivée d'un jeune enfant est susceptible d'impliquer un niveau d'absence plus élevé quel que soit l'âge. Néanmoins, on s'attendrait à ne plus pouvoir mesurer la relation à partir du troisième quartile. Or, au centre hospitalier, cette mesure est permise, puisque 18 femmes de plus de 40 ans (borne inférieure du troisième quartile) ont eu un enfant dans l'année.

3 Facteurs d'absence et d'absentéisme

Jusqu'ici, nous avons passé en revue les facteurs de présentéisme, dont la mise en lumière est l'un des deux objectifs de cette thèse. Hormis sur la question de l'âge des salariés, les facteurs d'absence n'ont guère été évoqués. Or, l'un des intérêts du modèle ZINB est justement de mesurer simultanément le présentéisme et l'absence : il permet de tirer des enseignements sur l'un de ces comportements à la lumière de l'autre. Cette section est donc consacrée à un regard sur les facteurs d'absence (et d'absentéisme), regard renouvelé par la prise en compte conjointe du présentéisme.

Notons que la caractérisation de l'absentéisme, vu ici comme le comportement d'un salarié qui s'absente alors même qu'il n'est pas malade, ne peut être réalisée directement. En effet, les données dont nous disposons n'indiquent en rien le caractère injustifié de l'absence et ne permettent donc pas de séparer directement l'absence et l'absentéisme. Mais ce travail peut être réalisé *ex post* par l'analyse des facteurs d'absence et leur classification en deux catégories. Une première, où les facteurs d'absence paraissent légitimes parce qu'en rapport avec le motif qui officialise l'absence, à savoir la maladie. Une seconde, où les facteurs paraissent cette fois non légitimes, et correspondent à un comportement d'absentéisme. Bien entendu, il convient de rester prudent : un facteur donné peut difficilement être attribué exclusivement à l'absentéisme.

3.1 Facteurs d'absence

Âge. L'influence de l'âge des salariés sur leur niveau d'absence a déjà été largement décrite au cours de ce travail (chapitre 2 et section précédente). Comme dans de nombreux autres travaux (Donders et al., 2012 ; Frick & Malo, 2008 ; Missègue, 2007), nous trouvons une relation positive entre âge et absence, s'expliquant par la dégradation naturelle de l'état de santé. Ce constat s'appuie sur la régression effectuée au centre hospitalier, mais demeure vrai pour l'établissement bancaire : l'espérance d'absence d'un salarié de 30 ans y est de 5,35 contre 7,34, toutes choses égales par ailleurs. À la fin de la section précédente, nous avons cependant souligné que si l'espérance d'absence augmente avec l'âge, il en va de même pour la probabilité de présentéisme. Statistiquement, ces deux effets conjoints peuvent s'annuler en apparence. D'où l'enjeu d'une analyse de l'absence à la lumière du présentéisme.

Sexe. Le niveau d'absence des femmes est également plus élevé que celui des hommes, dans le contexte de l'établissement bancaire. Ce résultat est conforme à une bonne partie de la littérature sur l'absence (Dionne & Dostie, 2007 ; Jensen & MacIntosh, 2007 ; Kristensen et al., 2006). Il paraît difficile d'attribuer cette différence aux rôles sociaux plus spécifiquement dévolus aux femmes, et ce pour deux raisons. D'une part, les salariés de l'établissement bancaire étudié bénéficient de journées autorisées d'absence pour enfants malades, ce qui a pour conséquence statistique de les sortir de la comptabilisation des absences-maladie. D'autre part, l'équation de comptage intègre également une variable indiquant si la personne a connu un congé maternité au cours de l'année : les effets liés à l'accueil d'un enfant et aux pathologies de la grossesse sont donc mesurés.

Ce niveau d'absence féminin élevé pourrait, en premier lieu, s'expliquer par une plus grande attention à la santé (Montaut, 2010), élément déjà évoqué pour le présentéisme. En plus de déclencher l'absence-maladie plus précocement que chez les hommes (moins de présentéisme), cette vigilance pourrait également s'accompagner d'un allongement de la durée des absences. Les femmes ayant tendance à sous-évaluer leur état de santé réel (Shmueli, 2003), il ne serait pas surprenant que le seuil auquel elles décident de retourner au travail intervienne plus tardivement que celui des hommes. En second lieu, une espérance d'absence féminine plus élevée pourrait aussi être due à des conditions de travail ou de vie plus difficiles que pour les hommes. Notons par exemple que dans cet établissement bancaire, 44,6% des femmes, contre 21,9% des hommes occupent des emplois peu qualifiés. Ceci est en partie confirmé par la régression complémentaire effectuée sur les seuls managers (26,9% de femmes). La variable SEXE n'y est plus significative, ce qui orienterait effectivement vers une relation de l'absence féminine avec la nature des tâches exercées.

Charge familiale. Le contexte familial des salariés intervient non seulement dans la décision de se rendre à son travail étant malade, mais conduit également à une augmentation de l'espérance d'absence : dans l'équation de comptage, le fait d'avoir eu un enfant dans l'année en cours a un poids très significatif, et ce dans les deux entreprises. L'augmentation de la durée d'absence-maladie est bien entendu liée à toutes les pathologies et complications suivant ou précédant l'accouchement. C'est justement l'intérêt de la variable JEUNE ENFANT que d'isoler l'influence des grossesses pathologiques. Mais cet effet englobe également l'accueil d'un nouveau-né, qui implique de lourdes modifications dans l'équilibre familial : le fait d'avoir un enfant en bas âge est donc également

générateur d'absence. La présence d'un enfant au foyer est aussi l'un des critères légitimant certaines absences au sein des équipes de travail. On accepte ainsi probablement plus facilement l'absence de salariés ayant des enfants : il est convenu que la présence d'enfants au foyer, *a fortiori* s'ils sont très jeunes, s'assortit d'un aléa supplémentaire pesant sur l'organisation du salarié. Les maladies, la scolarité, la dépendance à l'égard d'un mode de garde peuvent amener un salarié à s'absenter pour faire face aux imprévus. Dans certains cas, les salariés peuvent même franchir le pas de l'absentéisme, entendu comme une absence non liée à la santé du salarié lui-même. Malgré la possibilité de poser un congé enfant-malade dans les deux contextes¹⁵, l'absence-maladie peut aussi être vue comme une solution d'ajustement entre divers types de contraintes.

Pénibilité du travail. L'étude de l'absence au centre hospitalier met en lumière un mécanisme particulièrement intéressant. Plusieurs variables d'appartenance à des emplois spécifiques (qui ont été décrit au début du chapitre 4) se sont avérées significatives dans l'équation de comptage. On constate en effet que l'espérance d'absence est plus élevée pour les personnes occupant certains postes : les aides-soignants et les agents des services hospitaliers (ASH) ont respectivement une espérance d'absence de 46% ($e^{0,380} = 1,46$) et 25% ($e^{0,225} = 1,25$) plus élevée, comparativement aux autres salariés. La différence essentielle entre les métiers hospitaliers tient sans doute à la nature des tâches exécutées. Les métiers d'aide-soignant et d'ASH se caractérisent ainsi par une pénibilité du travail plus marquée (toilette et manipulation des personnes, entretien des locaux, etc...), ce qui expliquerait un niveau d'absence plus élevé. Mais comme nous le verrons dans la partie suivante, l'absentéisme, véhiculé par une faible qualité du travail, est également une interprétation possible.

Aucune information directe sur l'emploi occupé n'apparaît dans la régression de l'établissement bancaire. En revanche, les salariés sont répartis en trois grands niveaux hiérarchiques (techniciens, agents de maîtrise et cadres) eux-mêmes subdivisés en plusieurs échelons : la variable HIÉRARCHIE indique l'échelon (de 1 à 10) auquel se situe le salarié. L'espérance d'absence augmente à mesure que l'on descend dans l'échelle hiérarchique. Or, les échelons les plus faibles sont principalement composés de deux corps de métiers : de l'intégralité des conseillers commerciaux et de la majorité des assistants de gestion bancaire (63,9%). Ces deux métiers, qui représentent 86,4% de la

15. Notons toutefois que ce congé enfant-malade n'est pas un droit, mais reste soumis à l'autorisation du cadre. En outre, il est d'ampleur limitée (5 jours par salarié et par an).

catégorie des techniciens, sont également reconnus pour leur pénibilité. Les conseillers commerciaux sont quotidiennement en première ligne vis-à-vis de la clientèle et subissent un stress important, tandis que les assistants de gestion bancaire exercent des tâches administratives de routine. De façon générale, au fur et à mesure que l'on évolue dans les échelons hiérarchiques, les métiers sont davantage orientés vers l'encadrement d'équipes, la gestion de projets ou la prise de décision. Les métiers de faible échelon se caractérisent par une faible autonomie et, comme on l'a vu, par des tâches stressantes ou routinières. Ces postes moins gratifiants et moins qualifiés peuvent expliquer des répercussions négatives sur la santé (notamment psychologique), mais également une motivation plus faible, source d'absentéisme.

3.2 Facteurs d'absentéisme

Sécurité de l'emploi. Au centre hospitalier, les salariés en CDD se distinguaient par un présentéisme plus élevé, que nous imputons notamment à un effort d'implication en vue d'une éventuelle pérennisation de l'emploi. Un niveau d'absence élevé constitue en effet un obstacle quasiment rédhibitoire à l'accès au CDI. Les salariés en CDD sont donc incités au présentéisme. Cependant, dans l'établissement bancaire, la question se pose différemment. Les salariés en CDD ont en moyenne une espérance d'absence inférieure de 36% ($=1 - e^{-0,446}$) à celle des autres salariés. Certes, leur attitude pourrait consister à tenter de restreindre la longueur de leurs absences pour donner des gages à leur hiérarchie. Mais, d'après nous, ce coefficient a une signification différente. Ichino et Riphahn (2005) ont très clairement montré que lorsque la sécurité de l'emploi est acquise (par exemple après une période probatoire), l'absentéisme augmente significativement. L'accès à un poste pérenne constitue un seuil, au-delà duquel le salarié s'autorise à nouveau des absences-maladies justifiées, mais prend également davantage de libertés concernant des absences non justifiées. Au regard de nos résultats, il semble donc que la sécurité de l'emploi qu'offre le CDI est propre à générer de l'absentéisme.

Qualité de l'emploi. Notre travail nous conduit évidemment à nous interroger fondamentalement sur les sources de l'absentéisme. Pourquoi un salarié mobilise-t-il volontairement un motif d'absence à contre-emploi ? Indéniablement, cela tient à la motivation au travail, et donc à son contenu. Certains travaux soulignent en effet l'influence de la variété des tâches réalisées, de la latitude décisionnelle, de l'ampleur des responsa-

bilités exercées sur la motivation et l'absentéisme des salariés (Hausknecht et al., 2008 ; Kristensen et al., 2006 ; Punnett et al., 2007). Dans notre cas, ces éléments peuvent être approchés à partir du niveau hiérarchique, compris entre 1 et 10 et caractérisant le degré de qualification d'un poste : il s'agit de la variable HIÉRARCHIE intégrée dans l'équation de comptage. Elle traduit également le poids du poste dans l'organisation et dépend de la nature des tâches confiées et de l'autonomie décisionnelle de l'individu qui l'occupe.

La situation vis-à-vis de l'absence-maladie est très différente suivant le niveau hiérarchique des salariés. La régression met en lumière une relation linéaire négative entre la position hiérarchique d'un salarié et son niveau d'absence : plus l'échelon du salarié est élevé, plus l'espérance d'absence est faible. À chaque échelon supplémentaire, l'espérance d'absence est multipliée par 0,89. Ainsi, lorsqu'un conseiller commercial (niveau 3) a par exemple une espérance d'absence de 6, un directeur d'agence (niveau 6) aura une espérance de 4,2, toutes choses égales par ailleurs. Un absentéisme plus élevé dans les postes procurant le moins de satisfaction aux salariés n'est pas à exclure ici. Les postes les moins gratifiants, les moins motivants, se situent au bas de l'échelle hiérarchique et pourraient conduire les salariés à plus d'absences, dont le motif n'est pas toujours la maladie. Cet exemple peut être transféré au cas du centre hospitalier : le niveau d'absence des ASH et des aides-soignants pourrait être interprété en ce sens.

3.3 Implications pour le management de l'absence et de l'absentéisme

En considérant que les salariés peuvent être présentéistes et délibérément renoncer à une absence-maladie justifiée, et au regard des effets potentiellement néfastes du présentéisme sur la santé et les absences futures, l'employeur ne peut plus se contenter d'agir de manière unilatérale dans le sens d'une réduction de l'absence. Atteindre le plus bas niveau possible d'absence ne peut constituer la seule orientation des politiques RH, et ce malgré le risque d'absentéisme pour certaines catégories de salariés : des évolutions dans les méthodes de gestion de l'absence semblent nécessaires.

Au travers de la littérature (Baker-McClearn et al., 2010 ; Böckerman & Laukkanen, 2010 ; Grinyer & Singleton, 2000 ; Rocheleau & Renaud, 2003) on peut recenser de nombreux dispositifs traditionnellement mis en place dans la « lutte contre l'absentéisme ». L'ajustement de la politique salariale, le contrôle et la pénalisation des absences ou la flexibilité du temps de travail sont des exemples parmi d'autres. À la lumière de notre

analyse, ces solutions ne peuvent plus être envisagées de la même manière : elles doivent parvenir à un équilibre situé entre l'objectif légitime de réduction de l'absentéisme et la vigilance à une progression du présentéisme.

Dans de nombreuses organisations, des dispositifs incitant financièrement les salariés à être présents sur leur lieu de travail sont en place. Du calcul de l'intéressement sur la base du temps de présence au délai de carence, en passant par les primes d'assiduité, l'arsenal des moyens est conséquent. L'identification du présentéisme comme étant potentiellement néfaste à la santé de la firme, ne suppose peut-être pas une suppression pure et simple des incitations financières à la présence, mais doit alerter l'entreprise sur la nécessité d'y fixer des seuils, en considérant que certaines absences lui sont profitables (Brown & Sessions, 2004).

Le contrôle et la pénalisation des absences est à tempérer dans le même sens. Punnett et al. (2007) mettent en garde l'entreprise contre une systématisation des sanctions. Certains salariés risquent de venir travailler uniquement en raison de la politique organisationnelle, et ce quel que soit leur état de santé (Grinyer & Singleton, 2000). Ceci renvoie directement aux nombreuses répercussions du présentéisme sur la santé des salariés (Bergström et al., 2009a ; Kivimäki et al., 2005) mises en avant dans le premier chapitre et qui peuvent aller jusqu'au burn-out (Demerouti et al., 2009).

Enfin, une plus grande flexibilité du temps de travail a parfois été présentée comme une solution possible vis-à-vis des problèmes d'absence (Afsa & Givord, 2009 ; Askenazy, 2004 ; Dionne & Dostie, 2007). L'aménagement des temps de présence est alors partiellement confié à la discrétion du salarié, rendant celui-ci plus maître de ses choix. Ce faisant, l'attention de l'employeur ne se porte plus prioritairement sur les heures de travail du salarié, mais bien sur l'accomplissement final des tâches qui lui incombent. Or, nos résultats montrent que les salariés occupant les plus hauts échelons hiérarchiques et disposant d'une plus grande autonomie, sont certes moins absents, mais également plus exposés au risque de présentéisme. Ainsi, même les solutions favorisant une plus grande autonomie dans l'organisation des horaires de travail doivent être appréhendées avec vigilance par les entreprises.

Conclusion

Bien que l'apport essentiel de cette thèse soit de nature méthodologique, ce chapitre montre à quel point deux simples applications du modèle ZINB à des cas empiriques permettent d'éclairer certains mécanismes comportementaux et aident à dresser le profil statistique d'une population à risque. Il faut souligner que les conclusions détaillées tout au long de ce dernier chapitre sont en totale cohérence avec ce que d'autres auteurs ont pu dire du présentéisme. Ceci constitue un argument supplémentaire confirmant la validité de notre démarche. Si la plupart des enseignements récurrents sur le présentéisme sont reproduits, nos observations appuient parfois des pans plus minoritaires de la littérature ou apportent un éclairage nouveau sur certains points. Nous reprenons nos résultats dans cette perspective.

En premier lieu, parmi les éléments les plus congruents avec les résultats antérieurs, citons ceux qui concernent le niveau d'absence-maladie. Si ce chapitre se concentrait d'abord sur la thématique du présentéisme, l'incidence des variables explicatives sur l'espérance d'absence a également été évoquée en dernière partie. En résumé, les salariés dont l'espérance d'absence est la plus forte sont les femmes, même lorsque l'effet de la maternité est pris en compte, les seniors et les individus occupant les postes les moins qualifiés et les plus pénibles. Nous avons notamment montré que les aide-soignants ou les ASH au centre hospitalier et les conseillers commerciaux dans l'établissement bancaire ont un niveau d'absence plus élevé.

Du côté du présentéisme, l'effet de la sécurité dans l'emploi et de la nature des tâches sont également confirmés. Les salariés qui craignent le plus de perdre leur emploi, en particulier parce qu'ils sont embauchés à durée déterminée ou parce qu'ils n'ont pas encore été titularisés, sont incités à venir travailler étant malades. Ce faisant, ils pensent donner à l'organisation et à leurs supérieurs un gage de loyauté et une preuve d'investissement dans le projet de l'entreprise. Dans certains métiers, la loyauté des salariés s'oriente davantage vers les personnes dont ils ont la charge. Il s'agit d'un constat très répandu : les salariés exerçant des métiers en lien avec des personnes fragiles ou dépendantes (malades, enfants, etc.) sont plus susceptibles de travailler étant malades. À cet égard, le niveau élevé des probabilités de présentéisme au centre hospitalier est une bonne illustration. Les infirmières, en relation quotidienne avec les patients, sont particulièrement à risque. De même, l'hypothèse d'un présentéisme plus élevé pour les salariés du secteur public est aussi confortée.

En deuxième lieu, quelques unes de nos conclusions ne sont convergentes qu'avec une partie minoritaire de la littérature. C'est surtout le cas de la dimension masculine du présentéisme. Celle-ci avait par exemple été pointée par Simpson (1998), puis plus récemment par Demerouti et al. (2009) dans un contexte hospitalier. Dans nos deux études de cas, le risque de présentéisme est plus marqué pour les hommes.

En troisième lieu, nous apportons quelques éléments nouveaux sur le comportement de présentéisme. C'est notamment le cas pour les effets de l'âge, de la responsabilité d'équipe et du lien d'exemplarité qui unit les managers et leurs subordonnés. Les développements sur la question de l'âge mettent en relief le véritable enjeu de l'analyse conjointe de l'absence et du présentéisme. Sans cette double approche, la véritable évolution des comportements d'absence avec l'âge peut demeurer tout à fait invisible. En effet, si l'absence-maladie augmente avec l'âge, comme l'ont montré d'autres travaux, le présentéisme progresse également en ce sens. Ces deux phénomènes peuvent donc se compenser en apparence et conduire à une analyse erronée.

Par ailleurs, un présentéisme plus marqué est mis à jour dans la population des responsables d'équipe. Les causes de ce comportement sont sans doute à chercher dans leur souci d'épauler coûte que coûte leurs équipes de travail, mais également dans la satisfaction liée à un poste en responsabilité. Ils sont conscients de leur position-clef dans l'organisation, et leur présence leur apparaît comme une nécessité. Cependant, leurs subordonnés sont particulièrement attentifs à leurs décisions : plus la propension au présentéisme des managers est marquée, plus importante sera celle des salariés qu'ils encadrent. On entrevoit aisément jusqu'où peut aller l'effet boule-de-neige de cette exemplarité, et à quel point le présentéisme doit maintenant intégrer les politiques RH des entreprises.

Conclusion générale

Depuis déjà plusieurs décennies, la question de l'absence de la main d'œuvre figure parmi les thématiques classiques de l'économie du travail et de la gestion des ressources humaines. Les approches de cette question sont variées et les raisons d'y répondre sont nombreuses. Dans certains cas, il s'agit de résoudre un dysfonctionnement organisationnel, nuisible à la santé de l'entreprise, en proposant des éléments incitatifs qui conduiraient les salariés à s'absenter moins. Dans d'autres cas, l'objectif est simplement d'identifier les facteurs qui, en agissant sur la santé des salariés, produisent de l'absence-maladie. Mais que le niveau des absences soit envisagé comme une source de coûts, un problème de santé publique, un indicateur du climat social ou comme un révélateur de la motivation des salariés, l'objectif ultime est toujours, peu ou prou, sa réduction. *A fortiori* lorsqu'on soupçonne qu'une part de ce niveau d'absence est imputable à l'absentéisme, c'est-à-dire à des absences injustifiées.

Cependant, derrière les préoccupations, légitimes, de certaines entreprises vis-à-vis du coût économique des absences, derrière les « taux d'absentéisme » et autres indicateurs mis en avant dans les baromètres nationaux, se dissimule une réalité moins bien connue. Dans la plupart des organisations, la part des salariés n'ayant jamais été absents au cours d'une année donnée avoisine souvent les deux tiers de la main d'œuvre, voire davantage. Les études consacrées à l'absence, qu'elles proviennent de la recherche académique ou de sources organisationnelles, considèrent cette catégorie d'employés jamais

absents comme le cas normal, la situation de référence. Or, ne pas avoir été absent ne signifie aucunement ne pas avoir été touché par la maladie. Au sein de cette population certains individus choisissent, pour des raisons diverses, de poursuivre l'exercice de leur travail malgré un état de santé dégradé.

Un tel constat ne vise nullement à remettre en cause l'intérêt des analyses produites sur l'absence au travail, mais à les reconsidérer sous un jour nouveau. Elles gagneraient en effet à tenir compte de ce comportement que la littérature appelle *présentéisme* et qui constitue le sujet de ce travail de thèse. Néanmoins, la simple identification d'un phénomène, si intéressant soit-il, ne justifie pas à elle seule que l'on s'y arrête, ni surtout que l'on y consacre un travail aussi conséquent qu'une thèse. Deux conditions semblent devoir être réunies pour légitimer une analyse approfondie du présentéisme. Il faudrait d'abord que ce comportement constitue réellement un problème pour l'organisation et ses protagonistes. Certes, des individus ne s'absentent pas alors qu'ils le devraient, mais la présence d'un salarié n'est-elle pas toujours préférable à son absence ? Par ailleurs, ce phénomène est-il simplement mesurable ? Il est difficilement envisageable pour une entreprise d'obtenir des informations sur la santé des salariés présents, et les moyens d'investigation réclamant l'interrogation des salariés ne peuvent y être mis en place ni de façon évidente ni de façon récurrente. En plus de poser concrètement un diagnostic dans le cas de deux organisations, notre travail de thèse répond à ces deux objections.

D'une part, il ne fait guère de doute que le présentéisme constitue véritablement un problème pour l'entreprise et, qui plus est, un problème sournois puisque statistiquement invisible. Ses implications sont en effet de plusieurs ordres. En premier lieu, la productivité d'un salarié présentéiste n'est pas optimale : la quantité et/ou la qualité du travail accompli lorsqu'il est malade s'avèrent plus faibles qu'à l'accoutumée. En deuxième lieu, le présentéisme est un vecteur de contagion, le salarié présentéiste pouvant propager à ses collègues une maladie généralement suivie d'absences. Le risque d'une telle propagation est d'ailleurs bien souvent inconsciemment amplifié par le manager, s'il encourage ses collaborateurs au présentéisme. Enfin, le présentéisme est générateur d'absences ultérieures, résultant d'une dégradation de l'état de santé de l'individu.

D'autre part, cette thèse répond amplement à la question de la mesure. Si les entreprises ne sont pas souvent en capacité de mener de lourdes enquêtes auprès de leur personnel, le fait est qu'elles disposent d'une information précieuse qu'il est possible de mobiliser pour une estimation du présentéisme : les enregistrements individuels des

arrêts-maladie. On peut en effet penser que l'absence-maladie n'est pas le seul phénomène capté dans ces données d'absence : la situation normale de présence au travail (se rendre au travail en bonne santé), le présentéisme (se rendre à son travail étant malade) et l'absentéisme (rester chez soi étant en bonne santé) sont trois autres comportements « emmêlés » au sein des micro-données d'absence. L'enjeu est alors de parvenir à faire le tri entre ces divers comportements imbriqués dans une seule et même variable. Nous montrons qu'un modèle binomial négatif à inflation de zéros (ZINB) permet d'estimer conjointement pour chaque individu, une espérance d'absence et une probabilité de présentéisme, et ce même lorsque la richesse des données disponibles est limitée. Cet outil permet d'obtenir une estimation en première intention et de poser un diagnostic préliminaire.

Ainsi, dans le cas des deux organisations auxquelles cette technique a été appliquée, il a été possible de caractériser avec plus d'acuité les profils des individus occupant un rôle-clef dans le processus causal du présentéisme. Plusieurs exemples de populations typiques peuvent être proposés. Premièrement, les managers, à l'interface entre les salariés et l'employeur, ont une influence remarquable sur le comportement de présentéisme de leurs subordonnés. Cette influence ne passe pas seulement par une « transmission » des injonctions organisationnelles (charge de travail, organisation du travail, politique d'absence), mais également par leur propre comportement. Deuxièmement, certains salariés interprètent le présentéisme comme une preuve d'investissement dans le travail, comme un témoignage concret de leur motivation. Ceux qui prétendent à une promotion ou qui souhaitent accéder à des postes pérennes sont alors incités à venir travailler étant malade. Troisièmement, les salariés pour qui l'absence est coûteuse parce que leur indemnisation en cas de maladie n'est que partielle, sont susceptibles de privilégier la présence au travail alors qu'un arrêt-maladie aurait été justifié. Dans ces trois exemples, l'organisation a des moyens d'agir concrètement en sensibilisant ses cadres ou en réorientant ses politiques. Encore doit-elle être capable de poser un tel diagnostic.

Bien entendu, nous ne prétendons pas à l'exhaustivité de ce travail de recherche. Le modèle que nous proposons est un outil d'estimation. Il permet d'évaluer l'importance du présentéisme dans une organisation et, plus largement, de décrire les comportements de présence et d'absence de ses salariés. En tant que tel, il renseigne sur le niveau d'absence que l'on peut attendre d'un salarié et sur sa propension à venir travailler étant malade. Son but premier n'est pas le dénombrement des occurrences de présentéisme. Mais il peut être utilisé en complémentarité avec d'autres outils, par exemple de type

qualitatif : rien n'empêche en effet l'entreprise de se concentrer par la suite sur l'analyse de certaines sous-populations plus particulièrement touchées. Notre outil permet une première orientation de ces moyens d'approfondissement.

De plus, ce travail n'offre pas un panorama exhaustif des facteurs de présentéisme. L'exploration se limite aux variables communément disponibles dans les services de ressources humaines. Mais, l'éventail des causalités et des implications du présentéisme pourrait être grandement élargi : de nombreuses autres pistes attendent en effet de pouvoir être sondées. Ainsi, si nous avons mis en évidence que la non-indemnisation de l'absence est sans doute un facteur de présentéisme, il reste à découvrir si les salariés en besoin financier, dont les revenus du ménage sont faibles par ailleurs, sont davantage présentéistes. Si les résurgences du présentéisme sur la productivité ont déjà été soulignées, bien peu de choses ont été dites sur ses rapports avec des mesures consistantes de la performance des salariés. De même, bien qu'avec d'autres auteurs nous ayons pointé l'importance du collectif de travail dans les comportements de présentéisme, ses répercussions sur le climat social de l'organisation et sur les dynamiques interpersonnelles demeurent largement méconnues.

En cela, notre travail n'est qu'un premier pas dans la perspective d'une meilleure compréhension du phénomène : bien des chantiers restent encore à explorer. Notre démarche prendrait tout son sens si d'autres chercheurs pouvaient s'emparer de la méthodologie proposée. D'après nous, la multiplication des études en situation est très certainement la voie royale pour approfondir la connaissance du présentéisme. C'est notamment le foisonnement d'études de cas qui a permis de mieux caractériser le phénomène d'absence de la main d'œuvre et ses facteurs. En ce sens, l'instrument que nous proposons dispose d'un atout de taille, puisqu'il autorise deux niveaux d'analyse. En premier lieu, il permet d'obtenir une moyenne des probabilités de présentéisme sur la seule base de la distribution des absences des salariés. Les données disponibles pour certaines entreprises peuvent en effet être très pauvres, voire se limiter aux enregistrements des absences individuelles. À partir de ce matériau restreint, le modèle est capable de mesurer un degré de présentéisme dans l'organisation. Cette information faciliterait par exemple la comparaison entre firmes, secteurs d'activité, branches professionnelles, etc. Les approches de ce type font largement défaut dans la littérature actuelle. En second lieu, le prolongement logique de la mesure est le diagnostic. L'inclusion de variables explicatives, lorsqu'elles sont disponibles, permet alors de progresser dans la découverte ou l'approfondissement des causalités du présentéisme.

Si notre travail ouvre de nombreuses voies sur la mesure « brute » du phénomène et sur la caractérisation de ses facteurs, il pourrait également permettre de mieux répondre à une autre question : celle des implications du présentéisme. Bien que celles-ci ne figuraient pas parmi les objectifs de cette thèse, elles y ont été largement abordées. Il s'agit sans doute d'une priorité dans l'agenda de recherche, puisque c'est la mise en relief des conséquences du présentéisme qui convaincra les entreprises de la nécessité d'y remédier. Cet aspect du problème bénéficiera largement d'un développement des analyses en panels des données d'absence. Au cours de ce travail, nous avons posé les premières bases d'une exploitation longitudinale des données d'absence. Il s'agit sans doute du meilleur moyen d'enrichir les informations déjà précieuses apportées par des approches en coupe transversale à propos du présentéisme. Une approche diachronique permettrait en effet de repérer des spécificités comportementales individuelles et surtout d'étudier des trajectoires particulières reflétant l'évolution des comportements d'absence au cours de la carrière. Elles autoriseraient par exemple une meilleure description des conséquences du présentéisme sur les absences ultérieures, en particulier sur les absences de longue durée.

Une meilleure prise en compte des effets néfastes du présentéisme contribuerait d'ailleurs également au renouvellement de la réflexion autour de l'absence-maladie et de la « lutte contre l'absentéisme ». Les entreprises misent encore beaucoup sur des dispositifs contraignants destinés à réduire le niveau d'absence. Le fait est que le présentéisme n'est encore guère connu sur le terrain, et l'absence de méthodologies de mesure accessibles n'est pas étrangère à cette méconnaissance. Une plus grande publicité de ce comportement constitue naturellement la condition nécessaire à sa prise en compte. L'émergence de la question du présentéisme s'inscrit dans une dynamique déjà largement amorcée dans le monde du travail : celle de l'attention grandissante aux risques psycho-sociaux encourus par les salariés, celle de la recherche du bien-être au travail. Même si ces évolutions sont progressives, elles sont de bon augure pour les évolutions à venir de la relation de travail.

Annexes

1 Du modèle de Poisson au modèle binomial négatif

Partons d'une variable aléatoire Y suivant une loi de Poisson d'espérance λ :

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}$$

Si l'on intègre multiplicativement de λ un paramètre $u \sim \Gamma(\theta, \theta)$ destiné à capter la surdispersion dans le modèle de Poisson, on obtient :

$$P(Y = y) = \frac{e^{-\lambda u} (\lambda u)^y}{y!}$$

On intègre à présent cette distribution Gamma au modèle de Poisson :

$$\begin{aligned} &= \int_0^\infty \frac{e^{-\lambda u} (\lambda u)^y}{y!} g(u) du \\ &= \int_0^\infty \frac{e^{-\lambda u} (\lambda u)^y}{y!} \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta u} u^{\theta-1} du \end{aligned}$$

On sort de l'intégrale tout ce qui ne dépend pas de u :

$$= \frac{\lambda^y}{y!} \frac{\theta^\theta}{\Gamma(\theta)} \int_0^\infty e^{-\lambda u} u^y e^{-\theta u} u^{\theta-1} du$$

En sachant que $y! = \Gamma(y+1)$ et que $e^x e^y = e^{(x+y)}$:

$$= \frac{\lambda^y}{\Gamma(y+1)} \frac{\theta^\theta}{\Gamma(\theta)} \int_0^\infty e^{-(\lambda+\theta)u} u^{(y+\theta)-1} du$$

Afin de retrouver la fonction de densité de Gamma, on multiplie et on divise le contenu de l'intégrale par le même terme :

$$\begin{aligned} &= \frac{\lambda^y}{\Gamma(y+1)} \frac{\theta^\theta}{\Gamma(\theta)} \int_0^\infty e^{-(\lambda+\theta)u} u^{(y+\theta)-1} \frac{(\theta+\lambda)^{\theta+y} \Gamma(\theta+y)}{(\theta+\lambda)^{\theta+y} \Gamma(\theta+y)} du \\ &= \frac{\lambda^y}{\Gamma(y+1)} \frac{\theta^\theta}{\Gamma(\theta)} \frac{\Gamma(\theta+y)}{(\theta+\lambda)^{\theta+y}} \int_0^\infty \frac{e^{-(\lambda+\theta)u} u^{(y+\theta)-1} (\theta+\lambda)^{\theta+y}}{\Gamma(\theta+y)} du \end{aligned}$$

On fait un changement de variable dans l'intégrale avec $\lambda + \theta = \delta$ et $\theta + y = \beta$:

$$= \frac{\lambda^y}{\Gamma(y+1)} \frac{\theta^\theta}{\Gamma(\theta)} \frac{\Gamma(\theta+y)}{(\theta+\lambda)^{\theta+y}} \int_0^\infty \frac{\delta^\beta}{\Gamma(\beta)} u^{\beta-1} e^{-\delta u} du$$

L'intégrale correspond bien à la fonction de densité de Gamma (cf. équation 2.3), qui est égale à 1. Il reste :

$$= \frac{\lambda^y}{\Gamma(y+1)} \frac{\theta^\theta}{\Gamma(\theta)} \frac{\Gamma(\theta+y)}{(\theta+\lambda)^{\theta+y}}$$

Le développement parvient ensuite à la fonction de distribution du modèle binomial négatif, qui peut s'écrire :

$$P(Y = y) = \frac{\Gamma(\theta+y)}{\Gamma(y+1)\Gamma(\theta)} \left(\frac{\theta}{\theta+\lambda} \right)^\theta \left(\frac{\lambda}{\theta+\lambda} \right)^y$$

2 Estimation par maximisation de la vraisemblance

2.1 Principe général de la maximisation de la vraisemblance

L'ensemble des modèles présentés dans cette thèse sont estimés par maximisation de la vraisemblance. Le principe général de cette méthode d'estimation est de chercher la valeur des coefficients ayant le plus « vraisemblablement » conduit à la production des données observées (Atkins & Gallop, 2007). Dans la pratique, il s'agit des coefficients qui maximisent la fonction de vraisemblance (ou de log-vraisemblance), c'est-à-dire qui annulent la dérivée de cette fonction.

Ce travail doit, dans la plupart des cas, se faire par itération grâce à un algorithme d'optimisation. L'algorithme utilisé par le logiciel SAS 9.3, pour lequel nous avons opté dans cette thèse¹⁶, est basé sur une variante de l'algorithme de Newton-Raphson¹⁷. Des valeurs de départ sont fournies à l'algorithme (valeurs 0 ou 1, voire dans le cas de la procédure GENMOD, les coefficients des MCO) qui les réévalue jusqu'à convergence, c'est-à-dire jusqu'à ce que l'évolution de la valeur des coefficients obtenus dans la dernière itération de l'algorithme soit inférieure à un certain seuil (0,0001 pour la procédure GENMOD).

Nous indiquons ci-après les fonctions de log-vraisemblance ($\mathcal{L}$) pour les quatre modèles exploités dans cette thèse.

2.2 Log-vraisemblance du modèle de Poisson :

$$\mathcal{L} = \sum_{i=1}^n [-\lambda_i + y_i \log(\lambda_i) - \log(y_i!)]$$

2.3 Log-vraisemblance du modèle binomial négatif :

$$\begin{aligned} \mathcal{L} = \sum_{i=1}^n & [\log \Gamma(\theta + y_i) - \log \Gamma(y_i + 1) - \log \Gamma(\theta) + \theta \log(\theta) \\ & + y_i \log(\lambda_i) - (y_i + \theta) \log(\theta + \lambda_i)] \end{aligned}$$

16. Avec la procédure GENMOD pour les modèles en coupe transversale et la procédure NLMIXED pour les modèles en panels.

17. Ridge-stabilized Newton-Raphson algorithm.

2.4 Log-vraisemblance du modèle *Zero-Inflated Poisson* (ZIP) :

$$\mathcal{L} = \begin{cases} \sum_{i=1}^n \log[p_i + (1 - p_i)e^{-\lambda_i}] & \text{si } y_i = 0 \\ \sum_{i=1}^n \log(1 - p_i) + y_i \log(\lambda_i) - \log(y_i!) - \lambda_i & \text{si } y_i > 0 \end{cases}$$

2.5 Log-vraisemblance du modèle *Zero-Inflated Negative Binomial* (ZINB) :

$$\mathcal{L} = \begin{cases} \sum_{i=1}^n \log \left[p_i + (1 - p_i) \left(\frac{\theta}{\theta + \lambda_i} \right)^\theta \right] & \text{si } y_i = 0 \\ \sum_{i=1}^n \log(1 - p_i) + \log \Gamma(y_i + \theta) + \log \Gamma(y_i + 1) - \log \Gamma(\theta) \\ \quad + \theta \log(\theta) + y_i \log(\lambda_i) - (\theta + y_i) \log(\theta + \lambda_i) & \text{si } y_i > 0 \end{cases}$$

3 Régressions sur la fréquence d'absence

Dans le corps de la thèse, la variable dépendante des régressions était le nombre de jours d'absence au cours d'une année donnée. L'absence était donc caractérisée par sa durée, sa gravité. Ce choix s'est rapidement avéré nécessaire parce qu'il permet de sauvegarder l'information de la longueur de la maladie, expression de la gravité de la pathologie dont souffre l'individu. Toutefois, il serait intéressant de savoir si les résultats obtenus auraient été similaires en adoptant pour variable dépendante le nombre d'occurrences d'absence au cours d'une année, c'est-à-dire la fréquence d'absence.

Pour vérifier ceci, nous reprenons le cas du centre hospitalier en 2010. Les caractéristiques de la variable de fréquence d'absence sont présentées en détail dans le chapitre 2 (page 64). À des fins de comparaison, nous conservons les mêmes variables explicatives et la même population de référence (4 090 individus) que précédemment, lorsque l'absence était exprimée par sa durée (cf. chapitre 4). Les résultats des régressions *zero-inflated* sont présentés dans le tableau suivant :

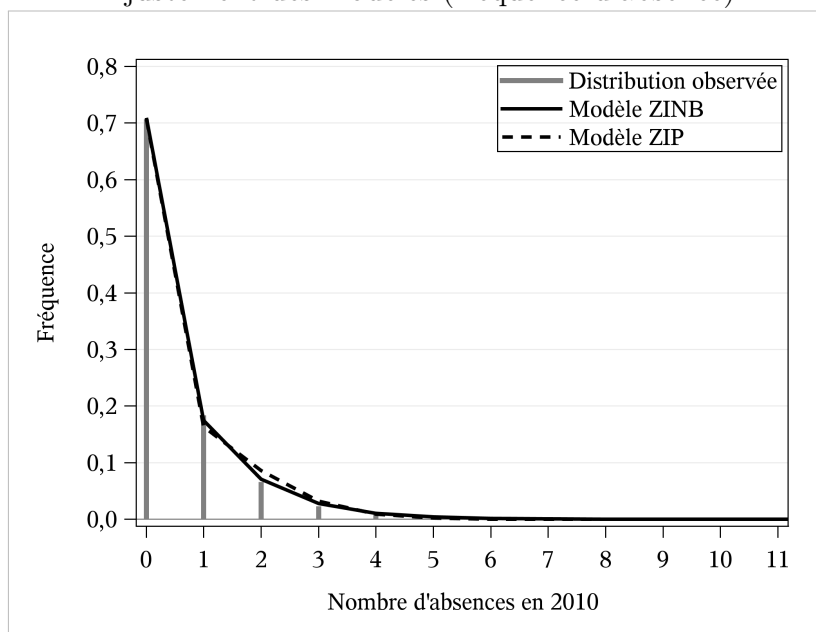
Estimation des modèles *zero-inflated* (fréquence d'absence)

Paramètre	ZIP		ZINB	
	Coefficient	Err.-type	Coefficient	Err.-type
<i>Équation de comptage :</i>				
CONSTANTE	-0,541 *	0,228	-0,914 **	0,288
AGE	0,007 *	0,003	0,006	0,004
JEUNE ENFANT (< 1 an)	0,267 *	0,107	0,347 **	0,126
ASH	0,375 ***	0,086	0,420 ***	0,103
AIDE-SOIGNANT	0,378 ***	0,064	0,384 ***	0,073
EXPOSITION	0,171	0,203	0,177	0,257
DISPERSION			0,756 ***	0,136
<i>Équation d'inflation :</i>				
CONSTANTE	0,035	0,414	-1,454 °	0,813
FEMME	-0,454 ***	0,121	-0,578 **	0,207
AGE	0,029 ***	0,007	0,049 ***	0,013
SANS ENFANTS	0,453 ***	0,126	0,755 **	0,234
INFIRMIER	0,205 °	0,122	0,321	0,236
CADRE	1,031 ***	0,262	1,424 ***	0,393
CDD	0,748 ***	0,180	1,122 ***	0,286
TITULAIRE	-0,323 *	0,153	-0,563 °	0,287
ANCIENNETÉ FAIBLE (< 1 an)	0,461 **	0,143	0,716 **	0,227
EXPOSITION	-1,031 ***	0,284	-1,447 **	0,466
Log-vraisemblance (AIC)	-2975 (7435)		-3658 (7351)	
Significativité : <0.1%=*** <1%=** <5%=* <10%=°				

À l'exception de la variable INFIRMIER, les résultats de l'équation d'inflation sont comparables à ceux obtenus avec les modèles dont la variable expliquée était le nombre de jours d'absence (cf. page 154). Dans l'équation de comptage, bien que les conclusions restent cohérentes, les coefficients présentent quelques variations : ceci confirme que les effets captés par les deux types de données d'absence ne sont pas similaires (cf. chapitre 2). À titre d'exemple, la variable d'âge a moins d'effet sur l'espérance d'absence lorsqu'on utilise la fréquence d'absence. Ce constat semble indiquer que le nombre de jours d'absence (gravité) est effectivement davantage lié aux problématiques de santé et à l'incidence des maladies.

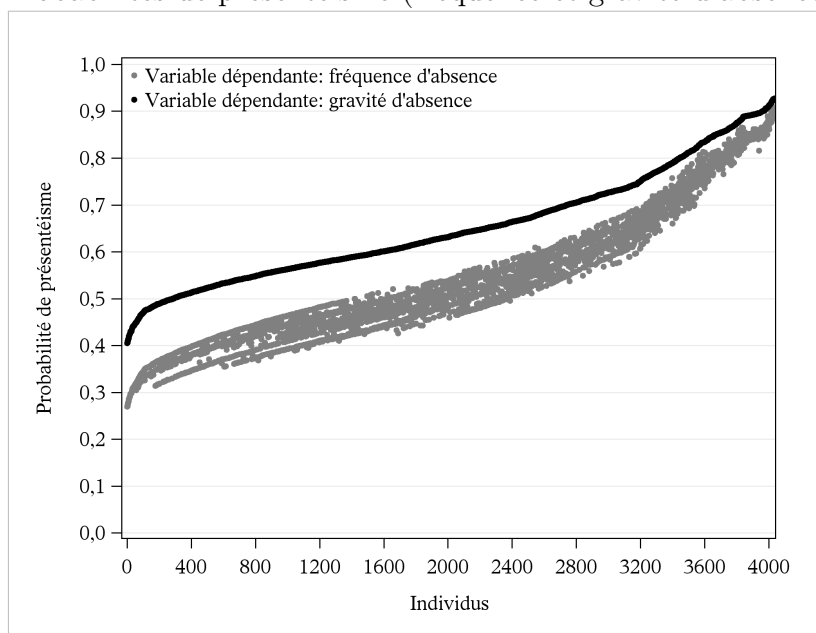
Un autre enseignement peut être tiré de ces résultats. Lorsque la variable dépendante est la fréquence d'absence, le choix du modèle ZINB au détriment du modèle ZIP est moins catégorique. La log-vraisemblance du modèle ZIP est en effet plus proche de zéro que celle du modèle ZINB, ce qui présage d'un meilleur ajustement. Empiriquement, ceci peut s'expliquer par une plus grande proximité de la variance et de la moyenne dans la variable de fréquence d'absence, autrement dit par une faible surdispersion. Or un modèle ZIP peut être suffisant dans cette situation. Le paramètre d'hétérogénéité inclut dans le modèle ZINB est d'ailleurs relativement faible, tout en restant néanmoins significativement différent de zéro. La représentation graphique de l'ajustement des deux modèles par rapport aux données empiriques ne permet de trancher dans un sens ou dans l'autre :

Ajustement des modèles (fréquence d'absence)



Enfin, si l'on choisissait le modèle ZIP en vertu de sa log-vraisemblance et de la faible surdispersion des données, il serait intéressant de confronter l'ordonnement des probabilités de présentéisme selon qu'on utilise pour variable dépendante la gravité ou la fréquence d'absence. Globalement, on s'aperçoit que même si les probabilités estimées se situent à un autre palier lorsqu'on utilise la fréquence d'absence, l'ordonnement général est néanmoins respecté :

Probabilités de présentéisme (fréquence et gravité d'absence)

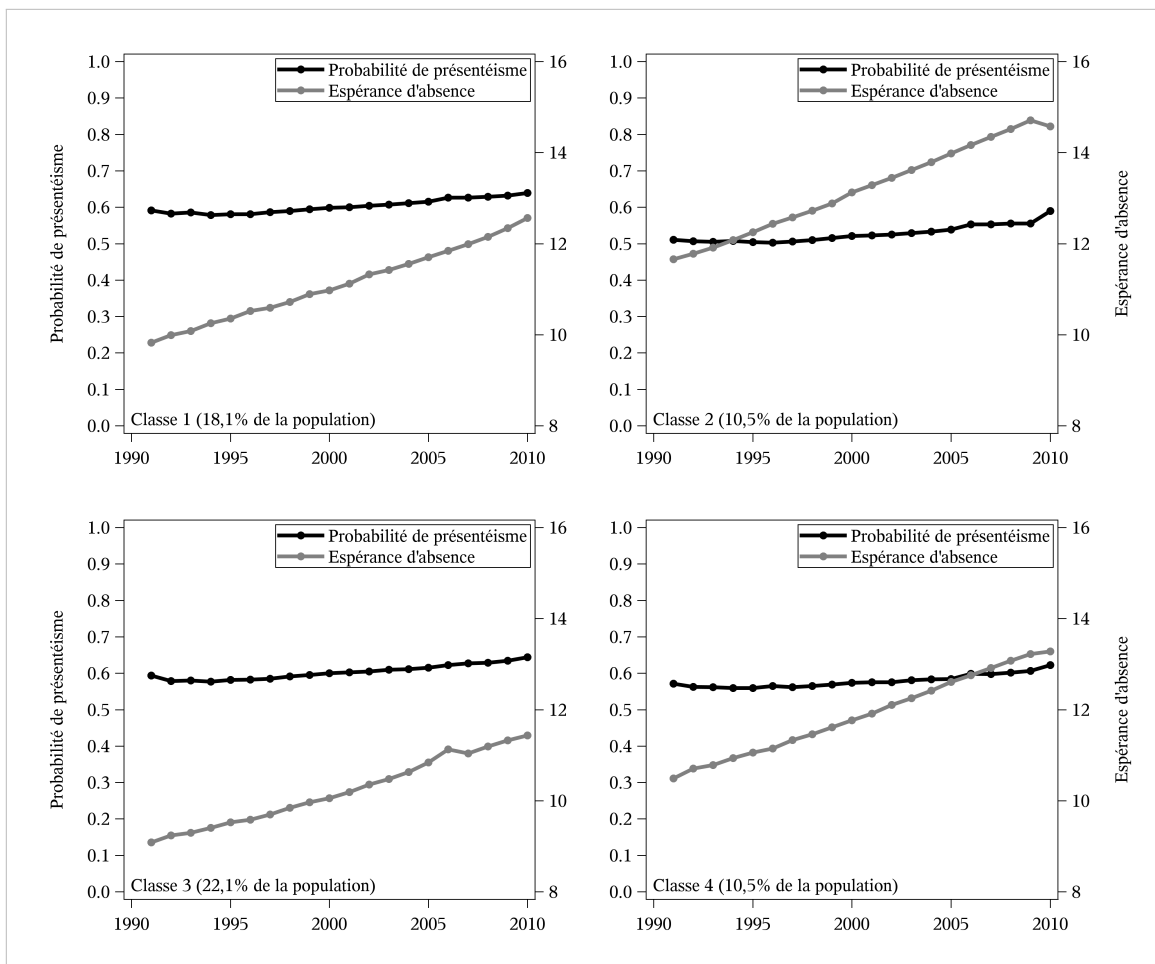


4 Résultats des régressions annuelles

Paramètre	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
CONSTANTE	1,435 ***	1,303 ***	1,428 ***	1,515 ***	1,478 ***	1,436 ***	1,482 ***	1,370 ***	1,467 ***	1,583 ***
AGE	0,015 ***	0,014 ***	0,015 ***	0,013 ***	0,015 ***	0,015 ***	0,021 ***	0,019 ***	0,016 ***	0,013 ***
JEUNE ENFANT	0,439 ***	0,532 ***	0,565 ***	0,654 ***	0,493 ***	0,674 ***	0,703 ***	0,494 ***	0,658 ***	0,713 ***
ASH	0,156 *	0,269 ***	0,258 ***	0,167 *	0,249 **	0,116	0,135	0,153	0,239 **	0,172 *
AIDE-SOIGNANT	0,173 *	0,246 ***	0,093	0,096	0,121	0,118	0,116	0,133	0,203 **	0,207 ***
VARIABLE TEMPORELLE	0,497 ***	0,584 ***	0,439 **	0,392 *	0,305	0,376 *	0,098	0,258	0,263	0,270
DISPERSION	0,738 ***	0,736 ***	0,744 ***	0,886 ***	0,900 ***	0,858 ***	0,837 ***	0,941 ***	0,961 ***	0,916 ***
CONSTANTE	1,669 ***	1,900 ***	2,333 ***	2,466 ***	1,656 ***	1,958 ***	2,125 ***	2,053 ***	1,709 ***	1,956 ***
FEMME	-0,698 ***	-0,603 ***	-0,707 ***	-0,504 ***	-0,498 ***	-0,579 ***	-0,819 ***	-0,658 ***	-0,418 ***	-0,503 ***
AGE	0,012 *	0,005	0,014 **	0,004	0,019 ***	0,014 **	0,014 **	0,013 *	0,013 **	0,015 **
SANS ENFANTS	0,303 **	0,412 ***	0,427 ***	0,346 ***	0,340 ***	0,329 ***	0,374 ***	0,328 ***	0,247 *	0,392 ***
INFIRMIER	0,382 ***	0,300 **	0,310 **	0,457 ***	0,322 ***	0,482 ***	0,351 ***	0,252 **	0,459 ***	0,375 ***
CADRE DE SANTÉ	0,652 **	0,492 *	0,463 *	0,755 ***	0,776 ***	0,902 ***	0,753 ***	0,691 ***	1,119 ***	0,944 ***
CDD	1,010 ***	0,813 ***	0,224	0,248	0,960 ***	0,651 **	0,794 ***	0,674 **	0,837 ***	0,338
TITULAIRE	0,230	0,271	-0,050	0,006	-0,282 *	-0,471 ***	-0,027	-0,309 *	-0,197	-0,345 **
ANCIENNETÉ FAIBLE	-0,026	0,051	0,110	0,118	-0,053	0,216	0,356 *	0,208	-0,074	-0,014
VARIABLE TEMPORELLE	-1,691 ***	-1,792 ***	-2,238 ***	-2,132 ***	-1,793 ***	-1,618 ***	-2,018 ***	-1,755 ***	-1,821 ***	-2,139 ***
Population	3321	3324	3509	3851	3715	3696	3799	3846	4030	4073

Paramètre	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
CONSTANTE	1,539 ***	1,777 ***	1,615 ***	1,118 ***	1,412 ***	1,643 ***	1,733 ***	1,594 ***	1,587 ***	1,658 ***
AGE	0,014 ***	0,010 ***	0,010 ***	0,017 ***	0,015 ***	0,009 **	0,010 ***	0,011 ***	0,013 ***	0,013 ***
JEUNE ENFANT	0,766 ***	0,622 ***	0,568 ***	0,833 ***	0,781 ***	0,757 ***	0,710 ***	0,766 ***	0,707 ***	0,628 ***
ASH	0,236 **	0,269 **	0,267 **	0,286 **	0,166	0,126	0,113	0,259 **	0,163	0,225 *
AIDE-SOIGNANT	0,153 *	0,177 **	0,196 **	0,252 ***	0,067	0,209 **	0,204 **	0,140 *	0,189 **	0,380 ***
VARIABLE TEMPORELLE	0,277 *	0,142	0,298 *	0,445 **	0,312	0,325 *	0,105	0,209	0,093	-0,106
DISPERSION	0,871 ***	0,913 ***	1,012 ***	1,133 ***	1,074 ***	1,031 ***	1,066 ***	1,063 ***	1,125 ***	1,331 ***
CONSTANTE	1,460 ***	1,484 ***	0,733 **	1,148 ***	1,227 ***	1,705 ***	0,928 ***	0,777 **	0,941 ***	0,837 **
FEMME	-0,546 ***	-0,393 ***	-0,459 ***	-0,630 ***	-0,523 ***	-0,695 ***	-0,492 ***	-0,386 ***	-0,579 ***	-0,447 ***
AGE	0,020 ***	0,024 ***	0,033 ***	0,030 ***	0,022 ***	0,026 ***	0,025 ***	0,019 ***	0,024 ***	0,022 ***
SANS ENFANTS	0,421 ***	0,472 ***	0,545 ***	0,607 ***	0,401 ***	0,351 ***	0,521 ***	0,382 ***	0,498 ***	0,401 ***
INFIRMIER	0,378 ***	0,168	0,254 **	0,215 *	0,461 ***	0,311 **	0,282 **	0,325 ***	0,234 **	0,319 ***
CADRE DE SANTÉ	0,717 ***	0,941 ***	0,963 ***	1,178 ***	0,883 ***	0,988 ***	1,215 ***	1,127 ***	0,777 ***	1,029 ***
CDD	0,277	0,230	0,328 *	0,243	0,262	0,131	0,244	0,615 ***	0,105	0,585 ***
TITULAIRE	-0,469 ***	-0,237	-0,261 *	-0,178	-0,205	-0,323 *	-0,235 *	-0,067	-0,422 **	-0,317 *
ANCIENNETÉ FAIBLE	0,134	0,180	0,311 **	0,070	0,203	-0,283 **	0,007	0,105	0,292 *	0,388 **
VARIABLE TEMPORELLE	-1,597 ***	-1,973 ***	-1,633 ***	-1,756 ***	-1,728 ***	-1,784 ***	-1,486 ***	-1,368 ***	-1,115 ***	-1,039 ***
Population	4146	4122	4200	4039	3940	3996	4135	4083	3966	4090

5 Classification des individus : profils moyens



6 Recherche d'interactions avec l'âge (hôpital)

Paramètre	Coefficient	Err.-type
<i>Équation de comptage :</i>		
CONSTANTE	1,953 ***	0,188
M_1	0,191 *	0,091
M_3	0,197 *	0,081
JEUNE ENFANT (< 1 an)	0,591 ***	0,138
ASH	0,249 *	0,109
AIDE-SOIGNANT	0,391 ***	0,080
EXPOSITION	-0,100	0,047
DISPERSION	1,330 ***	0,100
<i>Équation d'inflation :</i>		
CONSTANTE	0,863 ***	0,221
M_1	0,938 ***	0,185
M_2	0,332 ***	0,098
FEMME $\times M_1$	-0,526 ***	0,112
SANS ENFANTS	0,470 ***	0,109
INFIRMIER	0,299 ***	0,090
CADRE $\times M_1$	1,035 ***	0,238
CDD	0,642 ***	0,148
TITULAIRE $\times M_1$	-0,414 **	0,136
ANCIENNETÉ FAIBLE (< 1 an)	0,394 **	0,122
MOBILITÉ $\times M_1$	-0,314 *	0,151
EXPOSITION	-1,087 ***	0,186
Log-vraisemblance (AIC)	-6358 (12756)	
Significativité : <0,1%=*** <1%=** <5%=*		

Bibliographie

- Afsa, C. & Givord, P. (2009). Le rôle des conditions de travail dans les absences pour maladie : le cas des horaires irréguliers. *Économie et Prévision*, 187(1), 83-103.
- Agudelo-Suárez, A. A., Benavides, F. G., Felt, E., Ronda-Pérez, E., Vives-Cases, C. & García, A. M. (2010). Sickness presenteeism in Spanish-born and immigrant workers in Spain. *BMC Public Health*, 10, 1-7.
- Allain, E. & Brennac, T. (2001). Modèles linéaires généralisés appliqués à l'étude des nombres d'accidents sur des sites routiers : le modèle de Poisson et ses extensions. *Recherche Transports Sécurité*, 72, 3-18.
- Allen, S. G. (1981). An empirical model of work attendance. *Review of Economics and Statistics*, 63, 77-87.
- Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks : Sage Publications Inc.
- Aronsson, G. & Gustafsson, K. (2005). Sickness presenteeism : prevalence, attendance pressure factors, and an outline of a model for research. *Journal of Occupational and Environmental Medicine*, 47(9), 958-966.
- Aronsson, G., Gustafsson, K. & Dallner, M. (2000). Sick but yet at work : an empirical study of sickness presenteeism. *Journal of Epidemiology and Community Health*, 54, 502-509.
- Askenazy, P. (2004). Shorter work time, hours flexibility and labor intensification. *Eastern Economic Journal*, 30(4), 603-614.
- Askildsen, J. E., Bratberg, E. & Øivind Anti Nilsen. (2005). Unemployment, labour force composition and sickness absence : a panel data study. *Health Economics*, 14(11), 1087-1101. Document de travail, Institute for the Study of Labor.
- Atkins, D. C. & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes : a tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 21, 726-735.

- Baker-McCleary, D., Greasley, K., Dale, J. & Griffith, F. (2010). Absence management and presenteeism : the pressures on employees to attend work and the impact of attendance on performance. *Human Resource Management Journal*, 20(3), 311-328.
- Barmby, T. (2002). Worker absenteeism : a discrete hazard model with bivariate heterogeneity. *Labour Economics*, 9, 469-476.
- Barmby, T., Nolan, M. & Winkelmann, R. (2001). Contracted workdays and absence. *The Manchester School*, 69(3), 269-275.
- Barmby, T., Sessions, J. & Treble, J. (1994). Absenteeism : efficiency wages and shirking. *Scandinavian Journal of Economics*, 96(4), 561-566.
- Barmby, T. & Sibly, S. (2004). Analysing absence behaviour using event history models. *Managerial and Decision Economics*, 25, 141-145.
- Bergström, G., Bodin, L., Hagberg, J., Aronsson, G. & Josephson, M. (2009a). Does sickness presenteeism have an impact on future general health? *International Archives of Occupational and Environmental Health*, 82(10), 1179-1190.
- Bergström, G., Bodin, L., Hagberg, J., Aronsson, G. & Josephson, M. (2009b). Sickness presenteeism today, sickness absenteeism tomorrow? A prospective study on sickness presenteeism and future sickness absenteeism. *Journal of Occupational and Environmental Medicine*, 51(6), 629-638.
- Böckerman, P. & Laukkanen, E. (2009a). Presenteeism in Finland : determinants by gender and the sector of economy. *Ege Academic Review*, 9(3), 1007-1016.
- Böckerman, P. & Laukkanen, E. (2009b). What makes you work while you are sick? Evidence from a survey of workers. *European Journal of Public Health*, 20(1), 43-46.
- Böckerman, P. & Laukkanen, E. (2010). Predictors of sickness absence and presenteeism : does the pattern differ by a respondent's health? *Journal of Occupational and Environmental Medicine*, 52(3), 332-335.
- Boot, C. R. L., Koppes, L. L. J., Van den Bossche, S. N. J., Anema, J. R. & Van der Beek, A. J. (2011). Relation between perceived health and sick leave in employees with a chronic illness. *Journal of Occupational Rehabilitation*, 21, 211-219.
- Boucher, J.-P., Denuit, M. & Guillen, M. (2009). Number of accidents or number of claims? An approach with zero-inflated Poisson models for panel data. *The Journal of Risk and Insurance*, 76(4), 821-846.
- Brown, S. & Sessions, J. (2004). Absenteeism, presenteeism and shirking. *Economic Issues*, 9(1), 15-21.

- Cameron, A. C. & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge : Cambridge University Press.
- Campolieti, M. (2002). The recurrence of occupational injuries : estimates from a zero-inflated count model. *Applied Economics Letters*, 9, 565-600.
- Canfield, G. W. & Soash, D. G. (1955). Presenteeism : a constructive view. *Industrial Medicine and Surgery*, 24(9), 417-418.
- Carayol, N. (2007). Academic incentives, research organization and patenting at a large French university. *Economics of Innovation and New Technology*, 16(2), 119-138.
- Caverley, N., Cunningham, J. B. & MacGregor, J. N. (2007). Sickness presenteeism, sickness absenteeism, and health following restructuring in a public service organization. *Journal of Management Studies*, 44(2), 304-319.
- Chadwick-Jones, J. K., Brown, C. A., Nicholson, N. & Sheppard, C. (1971). Absence measures : their reliability and stability in an industrial setting. *Personnel Psychology*, 24, 463-470.
- Chatterji, M. & Tilley, C. J. (2002). Sickiness, absenteeism, presenteeism, and sick pay. *Oxford Economic Papers*, 54, 669-687.
- Chaupain-Guillot, S. & Guillot, O. (2007). Les absences au travail : une analyse à partir des données françaises du panel européen des ménages. *Economie et Statistique*, 408-409, 45-80.
- Chaupain-Guillot, S. & Guillot, O. (2011). Les déterminants individuels de l'absentéisme au travail : une comparaison européenne. *Revue Economique*, 62, 419-427.
- Claes, R. (2011). Employee correlates of sickness presence : a study across four European countries. *Work and Stress*, 25(3), 224-242.
- Cocker, F., Martin, A., Scott, J., Venn, A., Otahal, P. & Sanderson, K. (2011). Factors associated with presenteeism among employed Australian adults reporting lifetime major depression with 12-month symptoms. *Journal of Affective Disorders*, 135, 231-240.
- Coles, M. & Treble, J. (1993). The price of worker reliability. *Economics Letters*, 44, 149-155.
- Culpin, M. & Smith, M. (1928). Sickiness absenteeism : a preliminary study. *British Medical Journal*, 2(3524), 95-96.
- D'Abate, C. P. & Eddy, E. R. (2007). Engaging in personal business on the job : extending the presenteeism construct. *Human Resource Development Quarterly*, 18(3), 361-383.

- DARES. (2013). Les absences au travail des salariés pour raisons de santé : un rôle important des conditions de travail. *Analyses*, 9, 1-10.
- Delgado, M. A. & Kniesner, T. J. (1997). Count data models with variance of unknown form : an application to a hedonic model of worker absenteeism. *Review of Economics and Statistics*, 79(1), 41-49.
- Delsart, V., Rys, A. & Vaneecloo, N. (2009). *Méthodes statistiques de l'économie et de la gestion - Tome 3 - Économétrie théorie et application sous SAS*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.
- Demerouti, E., Le Blanc, P., Bakker, A. B., Schaufeli, W. B. & Hox, J. (2009). Present but sick : a three-wave study on job demands, presenteeism and burnout. *Career Development International*, 14(1), 50-68.
- Dew, K., Keefe, V. & Small, K. (2005). 'Choosing' to work when sick : workplace presenteeism. *Social Science and Medicine*, 60(10), 2273-2282.
- Dionne, G. & Dostie, B. (2007). New evidence on the determinants of absenteeism using linked employer-employee data. *Industrial and Labor Relations Review*, 61(1), 108-120.
- Démier, F. (1982). Les ouvriers de Rouen parlent à un économiste en juillet 1848. *Le Mouvement Social*, 119, 3-31.
- Donders, N., Bos, J. T., Van der Velden, K. & Van der Gulden, J. W. J. (2012). Age differences in the associations between sick leave and aspects of health, psychosocial workload and family life : a cross-sectional study. *BMJ Open*, 2, e000960.
- Dourgnon, P., Jusot, F. & Fantin, R. (2012). Payer nuit gravement à la santé : une étude de l'impact du renoncement financier aux soins sur l'état de santé. *Economie Publique*, 28-29, 123-147.
- Dunn, L. F. & Youngblood, S. A. (1986). Absenteeism as a mechanism for approaching an optimal labor market equilibrium. *Review of Economics and Statistics*, 68, 668-674.
- Elstad, J. I. & Vabø, M. (2008). Job stress, sickness absence and sickness presenteeism in Nordic elderly care. *Scandinavian Journal of Public Health*, 36(5), 467-474.
- Erdman, D., Jackson, L. & Sinko, A. (2008). *Zero-inflated Poisson and zero-inflated negative binomial using the COUNTREG procedure*. The SAS Institute Paper 322-2008.

- Ferrie, J. E., Kivimäki, M., Head, J., Shipley, M. J., Vahteraand, J. & Marmot, M. G. (2005). A comparison of self-reported sickness absence with absences recorded in employers' registers : evidence from the Whitehall II study. *Occupational and Environmental Medicine*, 62, 74-79.
- Firns, I., Travaglione, A. & O'Neill, G. (2006). Absenteeism in times of rapid organizational change. *Strategic Change*, 15, 113-128.
- Frick, B. & Malo, M. A. (2008). Labor market institutions and individual absenteeism in the European Union : the relative importance of sickness benefit systems and employment protection legislation. *Industrial Relations*, 47(4), 505-529.
- Gallois, P. (2005). *L'absentéisme : comprendre et agir*. Paris : Liaisons, coll. Entreprise et Carrières.
- Gatrell, C. J. (2011). 'I'm a bad mum' : pregnant presenteeism and poor health at work. *Social Science and Medicine*, 72(4), 478-485.
- Goetzel, R. Z., Long, S. R., Ozminkowski, R. J., Hawkins, K., Wang, S. & Lynch, W. (2004). Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers. *Journal of Occupational and Environmental Medicine*, 46(4), 398-411.
- Gosselin, E. & Lauzier, M. (2011). Le présentéisme : lorsque la présence n'est pas garante de la performance. *Revue Française de Gestion*, 37(211), 15-27.
- Gosselin, E., Lemyre, L. & Corneil, W. (2013). Presenteeism and absenteeism : differentiated understanding of related phenomena. *Journal of Occupational Health Psychology*, 18(1), 75-86.
- Greene, W. H. (2005). *Économétrie - Cinquième Edition*. Paris : Pearson Education.
- Grima, F. (2007). Impact du conflit intergénérationnel sur la relation à l'entreprise et au travail : proposition d'un modèle. *Management & Avenir*, 3-2007(13), 27-41.
- Grinyer, A. & Singleton, V. (2000). Sickness absence as risk-taking behaviour : a study of organisational and cultural factors in the public sector. *Health, Risk and Society*, 2(1), 7-21.
- Hammer, T. H. & Landau, J. (1981). Methodological issues in the use of absence data. *Journal of Applied Psychology*, 66(5), 574-581.
- Hansen, C. D. & Andersen, J. H. (2008). Going ill to work : what personal circumstances, attitudes and work-related factors are associated with sickness presenteeism? *Social Science and Medicine*, 67, 956-964.

- Hansen, C. D. & Andersen, J. H. (2009). Sick at work : a risk factor for long-term sickness absence at a later date ? *Journal of Epidemiology and Community Health*, 63(5), 397-402.
- Harrison, D. A. & Martocchio, J. J. (1998). Time for absenteeism : a 20-year review of origins, offshoots and outcomes. *Journal of Management*, 24(3), 305-350.
- Hausknecht, J. P., Hiller, N. J. & Vance, R. J. (2008). Work-unit absenteeism : effects of satisfaction, commitment, labor market conditions and time. *Academy of Management Journal*, 51(6), 1223-1245.
- Head, J., Ferrie, J. E., Alexanderson, K., Westerlund, H., Vahtera, J. & Kivimäki, M. (2008). Diagnosis-specific sickness absence as a predictor of mortality : the Whitehall II prospective cohort study. *British Medical Journal*, 337, a1469.
- Hilbe, J. M. (2011). *Negative binomial regression - Second Edition*. New York : Cambridge University Press.
- Hur, K., Hedecker, D., Henderson, W., Khuri, S. & Daley, J. (2002). Modeling clustered count data with excess zeros in health care outcomes research. *Health Services and Outcomes Research Methodology*, 3, 5-20.
- Ichino, A. & Riphahn, R. T. (2005). The effect of employment protection on worker effort : absenteeism during and after probation. *Journal of the European Economic Association*, 3(1), 120-143.
- Jensen, S. & MacIntosh, J. (2007). Absenteeism in the workplace : results from Danish sample survey data. *Empirical Economics*, 32, 125-139.
- Johansen, V. (2012). Sickness presenteeism in Norway and Sweden. *Nordic Journal of Social Research*, 3, 1-15.
- Johansson, G. & Lundberg, I. (2004). Adjustment latitude and attendance requirements as determinants of sickness absence or attendance. Empirical tests of the illness flexibility model. *Social Science and Medicine*, 58(10), 1857-1868.
- Johns, G. (2003). How methodological diversity has improved our understanding of absenteeism from work. *Human Resource Management Review*, 13, 157-184.
- Johns, G. (2010). Presenteeism in the workplace : a review and research agenda. *Journal of Organizational Behavior*, 31(4), 519-542.
- Johns, G. (2011). Attendance dynamics at work : the antecedents and correlates of presenteeism, absenteeism, and productivity loss. *Journal of Occupational Health Psychology*, 16(4), 483-500.

- Kivimäki, M., Head, J., Ferrie, J. E., Hemingway, H., Shipley, M. J., Vahtera, J. & Marmot, M. G. (2005). Working while ill as a risk factor for serious coronary events : the Whitehall II study. *American Journal of Public Health*, 95, 98-102.
- Kivimäki, M., Head, J., Ferrie, J. E., Shipley, M. J., Vahtera, J. & Marmot, M. G. (2003). Sickness absence as a global measure of health : evidence from mortality in the Whitehall II prospective cohort study. *British Medical Journal*, 327(7411), 1-6.
- Kivimäki, M., Vahtera, J., Thomson, L., Griffiths, A., Cox, T. & Pentti, J. (1997). Psychosocial factors predicting employee sickness absence during economic decline. *Journal of Applied Psychology*, 82(6), 858-872.
- Koopman, C., Pelletier, K. R., Murray, J. F., Sharda, C. E., Berger, M. L., Turpin, R. S., ... Bendel, T. (2002). Stanford presenteeism scale : health status and employee productivity. *Journal of Occupational and Environmental Medicine*, 44(1), 14-20.
- Kristensen, K., Juhl, H. J., Eskildsen, J., Nielsen, J., Frederiksen, N. & Bisgaard, C. (2006). Determinants of absenteeism in a large Danish bank. *International Journal of Human Resource Management*, 17(9), 1645-1658.
- Krohne, K. & Magnussen, L. H. (2011). Go to work or report sick? A focus group study on decisions of sickness presence among offshore catering section workers. *BMC Research Notes*, 4, 1-7.
- Kumar, S., Grefenstette, J. J., Galloway, D., Albert, S. M. & Burke, D. S. (2013). Policies to reduce influenza in the workplace : impact assessments using an agent-based model. *American Journal of Public Health*, 103(8), 1406-1411.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.
- Lawal, B. H. (2012). Zero-inflated count regression models with applications to some examples. *Quality and Quantity*, 46, 19-38.
- Leineweber, C., Westerlund, H., Hagberg, J., Svedberg, P., Luokkala, M. & Alexanderson, K. (2010). Sickness presenteeism among Swedish police officers. *Journal of Occupational Rehabilitation*, 21, 17-22.
- Lerner, D., Amick, B. C., Rogers, W. H., Malpeis, S., Bungay, K. & Cynn, D. (2001). The work limitations questionnaire. *Medical Care*, 39(1), 72-85.
- Letvak, S. A., Ruhm, C. J. & Gupta, S. N. (2012). Nurses' presenteeism and its effects on self-reported quality of care and costs. *American Journal of Nursing*, 112(2), 30-38.

- Loeys, T., Moerkerke, B., Smet, O. D. & Buysse, A. (2012). The analysis of zero-inflated count data : beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology*, 65, 163-180.
- Lofland, J. H., Pizzi, L. & Frick, K. D. (2004). A review of health-related workplace productivity loss instruments. *Pharmacoeconomics*, 22(3), 165-184.
- Lu, X., Leino-Arjas, P., Piha, K., Aittomäki, A., Saastamoinen, P., Rahkonen, O. & Lahelma, E. (2010). On methodologies for analysing sickness absence data : an insight into a new method. *International Journal of Biological and Medical Sciences*, 5(1), 32-37.
- MacGregor, J. N., Cunningham, J. B. & Caverley, N. (2008). Factors in absenteeism and presenteeism : life events and health events. *Management Research News*, 31(8), 607-615.
- Marmot, M., Feeney, A., Shipley, M. & Syme, F. N. S. L. (1995). Sickness absence as a measure of health status and functioning : from the UK Whitehall II study. *Journal of Epidemiology and Community Health*, 49, 124-130.
- McKevitt, C., Morgan, M., Dundas, R. & Holland, W. W. (1997). Sickness absence and 'working through' illness : a comparison of two professional groups. *Journal of Public Health*, 19(3), 295-300.
- Min, Y. & Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, 5, 1-19.
- Missègue, N. (2007). Les arrêts de travail des séniors en emploi. *Dossiers Solidarité et Santé*, 2, 1-23.
- Montaut, A. (2010). Santé et recours aux soins des femmes et des hommes. *Études et Résultats*, 717, 1-23.
- Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., . . . Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3), 381-391.
- Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M. & Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics*, 15, 39-57.
- Nicholson, N. & Johns, G. (1985). The absence culture and the psychological contract : who's in control of absence ? *Academy of Management Review*, 10(3), 397-407.

- Nielsen, A.-K. L. (2008). Determinants of absenteeism in public organizations : a unit-level analysis of work absence in a large Danish municipality. *International Journal of Human Resource Management*, 19(7), 1330-1348.
- Nyberg, A., Westerlund, H., Magnusson-Hanson, L. L. & Theorell, T. (2008). Managerial leadership is associated with self-reported sickness absence and sickness presenteeism. *Scandinavian Journal of Public Health*, 36, 803-811.
- O'Hara, R. B. & Kotze, J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, 1, 118-122.
- Ose, S. O. (2005). Working conditions, compensation and absenteeism. *Journal of Health Economics*, 24, 161-188.
- Pauly, M. V., Nicholson, S., Polsky, D., Berger, M. L. & Sharda, C. (2008). Valuing reductions in on-the-job illness : 'presenteeism' from managerial and economic perspectives. *Health Economics*, 17(4), 469-485.
- Pepin, M. & Moisdon, J.-C. (2010). Les impacts de la T2A sur les modes d'organisation et de fonctionnement des établissements de santé. *Dossiers Solidarité et Santé*, 16, 1-24.
- Punnett, B. J., Greenidge, D. & Ramsey, J. (2007). Job attitudes and absenteeism : a study in the English speaking Caribbean. *Journal of World Business*, 42, 214-227.
- Rentsch, J. R. & Steel, R. P. (2003). What does unit-level absence mean ? Issues for future unit-level absence research. *Human Resource Management Review*, 13, 185-202.
- Rocheleau, I. & Renaud, S. (2003). Les politiques de gestion de l'absence des entreprises et leurs impacts sur l'absentéisme au travail. *Revue Canadienne des Sciences de l'Administration*, 20(2), 149-165.
- Schultz, A. B., Chen, C.-Y. & Edington, D. W. (2009). The cost and impact of health conditions on presenteeism to employers. *Pharmacoeconomics*, 27(5), 365-378.
- Schultz, A. B. & Edington, D. W. (2007). Employee health and presenteeism : a systematic review. *Journal of Occupational Rehabilitation*, 17(3), 547-579.
- Shapira-Lishchinsky, O. & Rosenblatt, Z. (2009). Perceptions of organizational ethics as predictors of work absence : a test of alternative absence measures. *Journal of Business Ethics*, 88, 717-734.
- Shapiro, C. & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74, 433-444.

- Shmueli, A. (2003). Socio-economic and demographic variation in health and in heterogeneity : the issue of reporting heterogeneity. *Social Science and Medicine*, 57, 125-134.
- Simpson, R. (1998). Presenteeism, power and organizational change : long hours as a career barrier and the impact on the working lives of women managers. *British Journal of Management*, 9, 37-50.
- Stankiewicz, F. & Geuze, F. (2007). *Manager RH - Des concepts pour agir*. Bruxelles : De Boeck.
- Staub, K. E. & Winkelmann, R. (2012). Consistent estimation of zero-inflated count models. *Health Economics, Online*, 1-13.
- Steel, R. P. (2003). Methodological and operational issues in the construction of absence variables. *Human Resource Management Review*, 13, 243-251.
- Steel, R. P. & Rentsch, J. R. (1995). Influence of cumulation strategies on the long-range prediction of absenteeism. *Academy of Management Journal*, 38(6), 1616-1634.
- Steers, R. M. & Rhodes, S. R. (1978). Major influences on employee attendance : a process model. *Journal of Applied Psychology*, 63(4), 391-407.
- Taimela, S., Läärä, E., Malmivaara, A., Tiekso, J., Sintonen, H., Justén, S. & Aro, T. (2007). Self-reported health problems and sickness absence in different age groups predominantly engaged in physical work. *Occupational and Environmental Medicine*, 64, 739-746.
- Taloyan, M., Aronsson, G., Leineweber, C., Hanson, L. M., Alexanderson, K. & Westerlund, H. (2012). Sickness presenteeism predicts suboptimal self-rated health and sickness absence : a nationally representative study of the Swedish working population. *PlosOne*, 7(9), e44721.
- Treble, J. & Barmby, T. (2011). *Worker absenteeism and sick pay*. New York : Cambridge University Press.
- Trempé, R. (1971). *Les mineurs de Carmaux 1848-1914*. Paris : Les Editions Ouvrières.
- Villermé, L. (1840). *Tableau de l'état physique et moral des ouvriers dans les manufactures de coton, de laine et de soie (Tome 1)*. Paris : J. Renouard et Cie.
- Vingård, E., Alexanderson, K. & Norlund, A. (2004). Sickness presence. *Scandinavian Journal of Public Health*, 32(Suppl 63), 216-221.
- Virtanen, M., Stansfeld, S. A., Fuhrer, R., Ferrie, J. E. & Kivimäki, M. (2012). Over-time work as a predictor of major depressive episode : a 5-year follow-up of the Whitehall II study. *PlosOne*, 7(1), e30719.

- Weiss, A. (1985). Absenteeism and wages. *Economic Letters*, 19, 277-279.
- Whitston, C. & Edwards, P. K. (1990). Managing absence in an NHS hospital. *Industrial Relations Journal*, 21(4), 287-297.
- Winkelmann, R. (1999). Wages, firm size and absenteeism. *Applied Economics Letters*, 6(6), 337-341.
- Winkelmann, R. (2008). *Econometric Analysis of Count Data - Fifth Edition*. Berlin Heidelberg : Springer-Verlag.
- Yamamoto, S., Loerbroks, A. & Terris, D. D. (2009). Measuring the effect of workplace health promotion interventions on presenteeism : a potential role for biomarkers. *Preventive Medicine*, 48(5), 471-472.
- Yau, K. K., Lee, A. H. & Carrivick, P. J. W. (2004). Modeling zero-inflated count series with application to occupational health. *Computer Methods and Programs in Biomedicine*, 74, 47-52.
- Zhang, W., Bansback, N. & Anis, A. H. (2011). Measuring and valuing productivity loss due to poor health : a critical review. *Social Science and Medicine*, 72(2), 185-192.
- Zhang, W., Bansback, N., Boonen, A., Severens, J. L. & Anis, A. H. (2012). Development of a composite questionnaire, the valuation of lost productivity, to value productivity losses : application in rheumatoid arthritis. *Value in Health*, 15(1), 46-54.

Liste des tableaux

1.1	Prévalence du présentéisme selon différents articles	28
2.1	Description statistique des données d'absence	66
2.2	Absences et santé perçue (EWCS en 2010)	72
3.1	Coefficients de l'équation de comptage (sim. n° 2 et n° 3)	115
3.2	Coefficients estimés (simulation n° 4)	117
3.3	Coefficients estimés (simulation n° 5 et n° 6)	119
3.4	Coefficients estimés (simulation n° 7 & n° 8)	122
3.5	Coefficients estimés (simulation n° 9)	125
3.6	Coefficients estimés (simulation n° 10)	127
4.1	Variables démographiques (Banque en 2008)	139
4.2	Variables organisationnelles (Banque en 2008)	141
4.3	Variables démographiques (Hôpital)	146
4.4	Variables organisationnelles (Hôpital)	149
4.5	Estimation des modèles standards (Banque)	153
4.6	Estimation des modèles <i>zero-inflated</i> (Banque)	153
4.7	Estimation des modèles standards (Hôpital)	154
4.8	Estimation des modèles <i>zero-inflated</i> (Hôpital)	154
4.9	Probabilité de présentéisme et jours d'absence (Banque)	167
4.10	Probabilité de présentéisme et jours d'absence (Hôpital)	167
4.11	Estimation du modèle ZINB sans variables explicatives	169
4.12	Probabilité de présentéisme et moyenne d'absence sur 10 ans (Hôpital)	172

4.13	Probabilité de présentéisme et moyenne d'absence sur 20 ans (Hôpital)	172
4.14	Statistiques sur les coefficients des modèles annuels (Hôpital)	175
4.15	Estimation des modèles en panels (Hôpital)	178
5.1	Managers et subordonnées : statistiques descriptives	196
5.2	Estimation des modèles	198
5.3	Âge et présentéisme : statistiques descriptives	211
5.4	Estimation du modèle (coefficients reconstitués)	211

Table des figures

1.1	Production scientifique sur l'absence au travail (1970-2011)	13
1.2	Production scientifique sur le présentéisme (1970-2011)	13
1.3	Synthèse des modèles théoriques	56
2.1	Nombre de jours et d'épisodes d'absence (CHR en 2010)	65
2.2	Nombre de jours d'absence déclarés (EWCS en 2010)	65
2.3	Concentration du nombre de jours d'absence (CHR en 2010)	67
2.4	Absence, santé perçue et incidents de santé (EWCS en 2010)	71
2.5	Absences au CHR et épidémies saisonnières (2003-2012)	73
2.6	Âge, santé perçue et absences (EWCS en 2010)	75
2.7	Âge, santé perçue et absences de moins de 90 jours (EWCS en 2010) . .	75
2.8	Distributions de Poisson selon la moyenne λ	79
2.9	Distributions de la loi Gamma selon son paramètre α	83
2.10	Distributions binomiales négatives selon λ et α	85
2.11	Distributions de la loi binomiale négative (avec $\lambda = 10$)	88
3.1	Distributions ZIP selon λ et la probabilité d'inflation de zéro p	97
3.2	Distributions ZINB selon λ et la probabilité d'inflation de zéro p . . .	99
3.3	Estimation de deux variables : modèles sans inflation de zéro	111
3.4	Probabilités de présentéisme théoriques et estimées (simulation n° 1) .	112
3.5	Estimation du coefficient de la variable AGE (simulation n° 1)	112
3.6	Probabilités de présentéisme théoriques et estimées (sim. n° 2 & n° 3) .	114
3.7	Probabilités de présentéisme théoriques et estimées (simulation n° 4) .	116
3.8	Probabilités de présentéisme théoriques et estimées (simulation n° 5) .	118

3.9	Probabilités de présentéisme théoriques et estimées (simulation n° 6)	118
3.10	Distributions des variables corrélées (simulation n° 7)	120
3.11	Distributions de l'hétérogénéité (simulation n° 8)	121
3.12	Probabilités de présentéisme théoriques et estimées (simulation n° 7)	122
3.13	Probabilités de présentéisme théoriques et estimées (simulation n° 8)	123
3.14	Probabilités de présentéisme théoriques et estimées (simulation n° 9)	124
3.15	Probabilités de présentéisme théoriques et estimées (simulation n° 10)	126
3.16	Probabilités de présentéisme théoriques et estimées (simulation n° 11)	129
4.1	Évolution des absences selon leur type (Banque de 2006 à 2008)	136
4.2	Distribution et concentration des jours d'absence (Banque en 2008)	138
4.3	Pyramide des âges (Banque en 2008)	140
4.4	Évolution des absences-maladie (Hôpital de 1987 à 2012)	144
4.5	Pyramide des âges (Hôpital en 2010)	147
4.6	Évolution de la structure des emplois (Hôpital de 1987 à 2012)	149
4.7	Distribution du paramètre d'hétérogénéité (Hôpital)	159
4.8	Ajustement des modèles	164
4.9	Distribution de la probabilité de présentéisme	166
4.10	Ajustement des modèles ZINB sans variables explicatives	170
4.11	Comparaison avec les données longitudinales	173
4.12	Stabilité des modèles annuels (Hôpital)	175
4.13	Ajustement des modèles en panels	178
5.1	Managers et subordonnés (Banque en 2008)	195
5.2	Ajustement des modèles	198
5.3	Âge et présentéisme (Hôpital en 2010)	208

Table des matières

Remerciements	1
Introduction générale	2
1. État des lieux sur le présentéisme	8
Introduction	8
1 Émergence du concept de présentéisme	10
1.1 Approche bibliométrique	10
1.2 Extension de la théorie économique de l'absence	15
2 Mesure du présentéisme	19
2.1 Questionnement direct	20
2.2 Échelles psychométriques	23
2.3 Limites des méthodes utilisées	25
3 Études empiriques sur les causes du présentéisme	29
3.1 Facteurs individuels	30
3.2 Facteurs liés à l'emploi	36
4 Conséquences du présentéisme	46
4.1 Conséquences sur la santé des salariés	46
4.2 Conséquences sur le niveau d'absence ultérieur	49
4.3 Conséquences sur la productivité au travail	50
4.4 Autres conséquences organisationnelles	53
Conclusion	54
2. Regard sur les données d'absence	57

Introduction	57
1 Choix de la variable dépendante	60
1.1 Gravité et fréquence d'absence	60
1.2 Période d'analyse	62
1.3 Absences individuelles et absences agrégées	63
1.4 Distribution statistique	64
2 Relation entre santé et absence au travail	68
2.1 Absence et indicateurs de santé	69
2.2 Absence, santé et âge	73
3 Modéliser les données d'absence	76
3.1 Outils classiques	76
3.2 Loi de Poisson	77
3.3 Traitement de l'hétérogénéité inter-individuelle	80
Conclusion	89
3. Modèles à inflation de zéros	91
Introduction	91
1 Traitement de l'hétérogénéité comportementale	93
1.1 Idée maîtresse	93
1.2 Fonctionnement des modèles à inflation de zéros	94
1.3 Interprétation pour les données d'absence	102
2 Simulations de Monte-Carlo	107
2.1 Préliminaires	107
2.2 Modèles basiques	110
2.3 Relâchement des hypothèses préliminaires	115
2.4 Cas particulier des données en panels	127
Conclusion	129
4. Application empirique : deux études de cas	132
Introduction	132
1 Présentation des bases de données	134
1.1 Établissement bancaire	135

1.2	Centre hospitalier	142
2	Régressions en coupe transversale	150
2.1	Spécification des modèles	151
2.2	Résultats des régressions	152
2.3	Éléments d'interprétation	155
3	Analyse des résultats	158
3.1	Hétérogénéité non observée	158
3.2	Ajustement général des modèles	160
3.3	Probabilités de présentéisme	166
3.4	Modèle sans variables explicatives	169
4	Exploitation de la dimension longitudinale	170
4.1	Probabilités de présentéisme et absences antérieures	171
4.2	Stabilité des régressions annuelles	174
4.3	Modèles en panels	176
	Conclusion	179
5.	Des facteurs explicatifs du présentéisme	181
	Introduction	181
1	Facteurs organisationnels de présentéisme	185
1.1	Effets des variables liées à l'emploi	185
1.2	Responsabilité d'équipe : analyse détaillée	192
2	Facteurs individuels de présentéisme	201
2.1	Effets des variables liées à l'individu	201
2.2	Âge et présentéisme : analyse détaillée	206
3	Facteurs d'absence et d'absentéisme	215
3.1	Facteurs d'absence	215
3.2	Facteurs d'absentéisme	218
3.3	Implications pour le management de l'absence et de l'absentéisme	219
	Conclusion	221
	Conclusion générale	223
	Annexes	228

1	Du modèle de Poisson au modèle binomial négatif	229
2	Estimation par maximisation de la vraisemblance	231
2.1	Principe général de la maximisation de la vraisemblance	231
2.2	Log-vraisemblance du modèle de Poisson :	231
2.3	Log-vraisemblance du modèle binomial négatif :	231
2.4	Log-vraisemblance du modèle <i>Zero-Inflated Poisson</i> (ZIP) : . . .	232
2.5	Log-vraisemblance du modèle <i>Zero-Inflated Negative Binomial</i> (ZINB) :	232
3	Régressions sur la fréquence d'absence	233
4	Résultats des régressions annuelles	236
5	Classification des individus : profils moyens	237
6	Recherche d'interactions avec l'âge (hôpital)	238
Bibliographie		239
Liste des tableaux		250
Table des figures		252
Table des matières		254