



Département d'économie

Microéconométrie

ECN-7120, Hiver 2015

**L'EFFET DES CONDITIONS DE TRAVAIL SUR LA SANTÉ DES
TRAVAILLEURS : UNE ÉTUDE ÉCONOMÉTRIQUE**

Nicolas CORNEAU-TREMBLAY ¹

Supervision : Prof. Carlos Ordás Criado

Mai 2015

1. Étudiant à la maîtrise, Département d'économie, Université Laval

Résumé

Dans ce travail, nous désirons évaluer l'impact des conditions psychologiques du milieu de travail sur la santé des travailleurs. Pour ce faire, nous comparons deux modèles explicatifs proposés dans la littérature afin de déterminer lequel des deux semble le mieux s'adapter à nos données. Ces deux modèles sont le modèle Demande-contrôle et le modèle Iso-contraint. Pour ce faire, nous appliquons un modèle en panel *pooled*. Puisque nos variables dépendantes sont la présence ou non d'un problème de santé, nous estimons un modèle probit. Nous trouvons que le modèle Iso-contraint ne fait pas mieux que le modèle Demande-contrôle.

Table des matières

1	Introduction	4
2	Revue de la littérature	4
3	Méthodologie	6
3.1	Probit	6
3.2	Panel	9
3.3	Tests de spécification	11
4	Données	11
5	Résultats	13
5.1	Problème de coeur	13
5.2	Problème de pression sanguine	14
6	Limites	15
7	Conclusion	16
8	Bibliographie	17
9	Annexe 1	18
9.1	Coeur	18
9.2	Pression	20

Liste des tableaux

1	Contraintes du modèle Demande-contrôle	5
2	Coeur : Modèle Demande-contrôle	18
3	Coeur : Modèle Iso-contraint	19
4	Pression : Modèle Demande-contrôle	20
5	Pression : Modèle Iso-contraint	21

1 Introduction

De plus en plus, l'optimisation des conditions de travail devient un enjeu important pour les entreprises. La concurrence féroce présente sur la plupart des marchés contemporains incite les employeurs à toujours demander le maximum à leurs employés. Si les conditions de travail peuvent avoir un effet sur la productivité des travailleurs, il est également de plus en plus démontré qu'elles peuvent également avoir un impact sur la santé de ceux-ci.

Dans ce travail, nous tentons de comparer deux modèles expliquant le lien entre les conditions de travail et la santé des travailleurs afin de voir lequel des deux semble le mieux supporté. Comme indicateurs de santé, nous nous intéresserons aux problèmes de coeur ainsi qu'aux problèmes de pression sanguine. Également, puisque l'expression "conditions de travail" peut tout autant comprendre l'ambiance entre les travailleurs que l'ergonomie des postes de travail, nous restreignons dans ce travail le sens de cette dernière. Ainsi, lorsque nous parlons dans les pages qui suivent des conditions de travail, nous nous intéressons plus particulièrement aux composantes psychologiques du travail, soit par exemple la perception d'un travailleur par rapport au support qu'il obtient à son travail.

Les modèles explicatifs que nous présentons dans la prochaine section portent donc explicitement sur ce type de conditions de travail. La section trois présente la méthodologie que nous utilisons pour cette comparaison. La présentation des données sur lesquelles nous appliquerons ces modèles et nos résultats suivent ensuite. Enfin, nous terminons ce travail avec une conclusion.

2 Revue de la littérature

La littérature portant sur l'effet des conditions de travail sur la santé des travailleurs s'est beaucoup développée au cours des dernières décennies. Longtemps, le seul facteur psychologique considéré comme pouvant affecter la santé des travailleurs était la charge de travail que ce dernier devait subir. Encore de nos jours, cette idée demeure répandue dans l'esprit collectif. Cependant, l'économie du travail ainsi que l'épidémiologie occupationnelle ont apporté récemment de nouveaux modèles explicatifs plus complets.² Nous nous intéressons dans ce travail à deux modèles en particulier, soit le modèle Demande-contrôle et le modèle Iso-contraint.³

Avant de présenter les modèles, il est important de spécifier que l'ensemble des recherches soulevées ci-dessous s'est intéressé aux maladies cardio-vasculaires chez les travailleurs. La raison est que le stress psychologique se manifeste de façon physiologique, notamment dans le système cardio-vasculaire (Sapolsky, 2004). Ainsi, si un travailleur subit un stress prolongé, ce système sera particulièrement demandé, ce qui risque, à long terme, d'y causer des problèmes divers. Cette raison est également pourquoi nos deux indicateurs de santé sont la présence ou non d'un problème

2. Pour une présentation récente de ces modèles : Berkman, Kawachi et Theorell, 2014.

3. À noter que ce type de comparaison a été effectuée dans le passé, notamment par Bosma *et al*, 1998

cardiaque ou de pression sanguine.

Le premier modèle qu'y nous intéresse est le modèle Demande-contrôle qui a été proposé par R. A. Karasek à la fin des années 70s (Karasek, 1979). Ce modèle propose de ne plus analyser les conditions de travail de façon linéaire, c'est-à-dire "plus de travail égale plus de problèmes de santé", mais plutôt dans une matrice de contraintes, présentée dans le tableau 1. Cette matrice prend en considération la charge de travail, ici appelée la demande, ainsi que le contrôle qu'un travailleur a sur la façon de faire son travail.

TABLE 1 – Contraintes du modèle Demande-contrôle

Conditions de travail	Demande	
	Faible/Faible	Faible/Forte
	Fort/Faible	Fort/Forte

Selon Karasek, les conditions les plus nocives pour un travailleur sont celles dans le cadran supérieur droit, où un travailleur est soumis à une lourde charge de travail, mais possède peu de contrôle sur son travail. Un exemple de ce type de travail est un poste sur une chaîne de montage. Ces lignes fonctionnent à un certain rythme, souvent élevé, et ralentissent rarement. Également, à un poste est assignée une tâche bien précise qui doit généralement être effectuée de manière répétitive. Le travailleur ne possède ainsi aucune latitude sur comment faire son travail. Les risques de santé les plus élevés sont donc issus d'une interaction entre une forte demande et un faible niveau de contrôle combinés. Également, de façon individuelle, une forte demande et un faible contrôle, indépendamment l'un de l'autre, semblent aussi augmenter les risques de santé.

Ce modèle existe depuis maintenant plus de trente ans et a été vérifié chez les Américains (Karasek, 1979), les Asiatiques (Kang *et al.*, 2005) et les Européens (Kivimäki *et al.*, 2012). Ce dernier exemple est particulièrement parlant, puisque les auteurs ont pris les résultats de trente études de cohorte effectuées entre 1985-2006 afin de vérifier la validité de ce modèle. Il semble qu'il tienne bien la route.

L'auteur original du second modèle, le modèle Iso-contraint, est quant à lui plus difficile à identifier. Ce modèle, s'il semble s'être développé en parallèle à celui de Karasek, en est en fait, dans sa forme la plus achevée, une extension. Le modèle stipule que si l'on se réfère aux études en psychologie et en biologie portant sur le stress, il est vrai qu'un manque de contrôle sur sa vie de même qu'un état d'alerte constant, les deux effets psychologiques du manque de contrôle et de la charge de travail lourde, augmentent le niveau de stress d'un individu. Cependant, les recherches dans ces champs d'études ont également trouvé qu'un bon support social permettait de réduire de façon significative le niveau de stress. L'intuition derrière ce constat est que lorsque quelque chose de préoccupant ou de dangereux survient, la gestion du problème est plus simple lorsque l'on n'est pas seul. C'est

ainsi que, dans les mêmes années que l'apparition du modèle Demande-contrôle, d'autres recherches portaient sur de potentiels médiateurs du stress au travail, l'un de ces médiateurs proposés fut le support en milieu de travail (LaRocco *et al*, 1980).

Depuis ce temps, le modèle a mené à des résultats mitigés. Notamment, Johnson *et al* (1996), n'ont trouvé un effet pour le support que lorsqu'il est croisé avec le contrôle. Ainsi, selon eux, les gens ayant peu de contrôle sur leur travail et peu de support ont un risque de problème cardiaque plus élevé. Une méta-analyse arrive également à la même conclusion, soit que le rôle du support comme médiateur pour le niveau de stress et ses conséquences sur la santé sont faibles (Viswesvaran *et al*, 1999).

3 Méthodologie

Pour comparer les deux modèles présentés à la section précédente, nous estimons un modèle à choix discret sur des données panel. Nous présentons dans cette section tout d'abord le modèle par maximum de vraisemblance probit. Nous abordons ensuite la théorie des données panel, et comment notre modèle est estimé. Enfin, nous présentons 2 tests de spécification que nous appliquons afin de comparer les deux modèles explicatifs abordés dans la section précédente.

3.1 Probit

Le modèle de maximum de vraisemblance probit repose sur un modèle à variable latente. Supposons un modèle linéaire tel que

$$y_i^* = X_i\beta + u_i$$

où y_i^* est une variable continue, X_i une matrice de variables explicatives, et u_i un terme d'erreur distribué selon une loi $N(0, \sigma^2)$. Cependant, nous n'observons pas y_i^* . Ce que nous observons plutôt, c'est

$$y_i = \begin{cases} 1 \\ 0 \end{cases} \quad (1)$$

Dans notre cas, y_i est si l'individu souffre ou non d'un problème de santé à une période donnée. On suppose

$$Pr(y_i = 1) = Pr(y_i^* > 0) = Pr(X_i\beta + u_i > 0) = Pr(u_i > -X_i\beta)$$

c'est-à-dire que y_i prend la valeur 1 lorsque la variable latente y_i^* est supérieure à zéro, et la valeur 0 autrement. Ainsi, on obtient

$$Pr(y_i = 1) = Pr(u_i > -X_i\beta)$$

$$Pr(y_i = 0) = Pr(u_i \leq -X_i\beta)$$

Précédemment, nous avons défini $u_i \sim N(0, \sigma^2)$. La distribution normale du terme d'erreur est ce qui définit le modèle probit. Ainsi

$$Pr(u_i > -X_i\beta) = \int_{-X_i\beta}^{\infty} \phi(u_i) du$$

où ϕ est la distribution normale définie comme

$$\phi(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}u_i^2\right)$$

D'après ce que nous venons de voir, on peut définir $Pr(y_i = 1)$ et $Pr(y_i = 0)$ de la façon suivante

$$Pr(y_i = 1) = \int_{-X_i\beta}^{\infty} \phi(u_i) du = 1 - \Phi(-X_i\beta)$$

$$Pr(y_i = 0) = \int_{-\infty}^{-X_i\beta} \phi(u_i) du = \Phi(-X_i\beta)$$

Selon la théorie des probabilités, la probabilité que deux événements indépendants surviennent au même moment est le produit des probabilités de chacun des événements. Selon l'hypothèse que nos observations sont indépendantes et identiquement distribuées, on peut définir la vraisemblance comme la probabilité d'observer simultanément notre échantillon en entier. En supposant que nous connaissons la distribution de nos observations, on peut définir la vraisemblance comme

$$L(y_i|X_i, \beta) = \prod_{i=1}^N Pr(y_i)$$

Soit

$$\Omega_0 = \{y_i = 0\}$$

$$\Omega_1 = \{y_i = 1\}$$

Notre vraisemblance peut également s'écrire

$$L(y_i|X_i, \beta) = \prod_{\Omega_0} (\Phi(-X_i\beta)) \prod_{\Omega_1} (1 - \Phi(-X_i\beta))$$

L'objectif de la méthode du maximum de vraisemblance est de trouver les β qui maximisent la vraisemblance que nous venons de définir. Dit autrement, nous cherchons les β qui rendent la probabilité d'obtenir simultanément nos observations la plus élevée possible. Ce problème se pose comme un problème d'optimisation standard dans lequel nous cherchons le maximum d'une fonction. Cependant, puisque la fonction est un produit de Φ , la fonction cumulative de la loi normale, dériver notre vraisemblance pourrait être laborieux. C'est pourquoi, de façon générale, nous travaillons plutôt avec le *log* de la fonction de vraisemblance. Cette transformation nous donne

$$\ln L(y_i|X_i, \beta) = l(y_i|X_i, \beta) = \sum_{\Omega_0} \ln \Phi(-X_i\beta) + \sum_{\Omega_1} \ln (1 - \Phi(-X_i\beta))$$

Ce type de problème ne possède pas de solutions évidentes ni faciles à dériver, c'est pourquoi son estimation est faite par ordinateur via des systèmes itératifs qui cherchent la valeur des β la plus

probable.

Évidemment, ce type d'estimation repose sur des hypothèses. Une hypothèse fondamentale, commune aux moindres carrés ordinaires, est l'exogénéité du terme d'erreur. À cela s'ajoutent deux hypothèses importantes qui ont été abordées précédemment, mais sur lesquelles il est tout de même pertinent de revenir clairement. Tout d'abord, puisque le modèle de maximum de vraisemblance utilise la propriété que la probabilité jointe de deux événements indépendants est le produit des probabilités de chaque événement, l'hypothèse d'indépendance des observations est fondamentale. Si ce n'est pas le cas, la vraisemblance ne peut être spécifiée. Cette hypothèse repose donc sur la collecte même des données et est souvent considérée comme valide sans vérification.

Également, nous avons fait l'hypothèse que toutes les observations sont identiquement distribuées, donc suivant la même loi de probabilité, et que dans notre cas, cette loi est la loi normale. C'est l'hypothèse que nos données suivent une loi normale qui fait en sorte que nous estimons un modèle probit. Si l'on supposait plutôt que nos données suivent une loi logistique, nous devrions estimer un modèle logit. Il est important de spécifier que l'estimateur par maximum de vraisemblance est valide si la densité de probabilité est bien spécifiée.

Pour ce qui est des coefficients estimés, ils ne peuvent être comparés les un aux autres. En effet, dans un modèle linéaire standard, si l'on veut connaître l'effet marginal d'un x sur y , on obtient simplement

$$\frac{\partial E(y|X)}{\partial x_j} = \beta_j$$

Cependant, dans un modèle non linéaire comme le modèle probit, si l'on dérive l'espérance de y par rapport à un x , on obtient

$$\frac{\partial E(y|X)}{\partial x_j} = \phi(-X_i\beta)\beta_j$$

On voit que l'effet dépend de la matrice X_i . Ainsi, on peut uniquement, à la vue des coefficients, considérer leur signe, car celui-ci ne changera pas (puisque $\phi(X_i\beta) > 0$), ainsi que leur niveau de significativité. Pour obtenir un coefficient qui puisse être comparé à un coefficient estimé par exemple par moindres carrés ordinaires, on peut par exemple estimer le coefficient lorsque toutes les variables explicatives dans la matrice X_i sont à leur moyenne. Cet effet est appelé l'effet partiel aux moyennes des variables.

En revenant au modèle probit, on voit que cet estimateur possède des avantages certains. Il permet notamment des prédictions qui sont bornées entre 0 et 1, ce que ne permet pas, par exemple, l'estimateur en probabilité linéaire. Il serait bien difficile d'expliquer comment on peut prédire pour un individu un risque de subir un événement avec une probabilité supérieure à un.

3.2 Panel

Des données panel sont des données où nous suivons plusieurs individus dans le temps. Nous avons donc des informations pour les mêmes individus sur plusieurs périodes. Cette information supplémentaire que nous n'avons pas dans les analyses transversales⁴ permet de faire des estimations plus sophistiquées. Supposons pour le moment un modèle linéaire standard en analyse transversale. Soit

$$y_i = X_i\beta + u_i$$

Ce que l'analyse en panel nous permet de faire, c'est d'ajouter un terme inobservable propre à chaque individu. Ainsi, notre modèle devient

$$y_{i,t} = X_{i,t}\beta + c_i + u_{i,t}$$

où à présent c_i est un facteur inobservable individuel. L'analyse en panel consiste à faire des hypothèses sur ce terme inobservable afin d'estimer des modèles. À remarquer également que désormais, les variables sont indicées à la fois à l'individu i et à la période t auxquelles elles correspondent.

L'un des modèles d'estimation en panel est le modèle à effet aléatoire. Ce modèle fait l'hypothèse que le terme inobservable n'est pas corrélé aux variables explicatives contenues dans X_{it} , tel que

$$\text{corr}(c_i, X_{i,t}) = 0$$

Ainsi, puisque le terme inobservable n'est pas corrélé avec les variables explicatives, son omission ne biaisera pas l'estimation des coefficients. Le modèle à effet aléatoire est une estimation par moindres carrés généralisés, via une transformation des variables. Cependant, le modèle à effet aléatoire fait l'hypothèse d'exogénéité stricte. Cette hypothèse spécifie que le terme d'erreur est de moyenne zéro conditionnellement aux valeurs passées, présentes et futures des variables explicatives, tel que

$$E(u_{i,t} | c_i, x_{i,1}, \dots, x_{i,T}) = 0$$

Cette hypothèse est non seulement forte, mais dans notre cas, elle est très certainement violée. En effet, le fait de souffrir d'un problème de santé à une période donnée est certainement fonction de si l'on souffre ou non de ce problème à la période précédente. Ainsi, nous introduisons dans nos modèles la variable dépendante retardée d'une période comme variable explicative. Cette pratique n'est cependant pas permise par le modèle à effet aléatoire, puisqu'elle viole l'hypothèse d'exogénéité stricte. Heureusement, cette hypothèse n'est pas nécessaire dans le modèle que nous estimons. Le modèle que nous adoptons est un modèle *pooled*, c'est-à-dire que chaque période de chaque individu est considérée comme un individu indépendant⁵. Nous avons en quelque sorte $(N \times T)$ individus⁶. Ce type de modèle ne nécessite pas l'hypothèse de stricte exogénéité, mais

4. En anglais : cross-sectional

5. Inspiré de (Wooldridge, 2002, pp.404-405, et p.483 et suivantes).

6. Où N est le nombre d'individus et T le nombre de périodes.

s'estime selon les hypothèses standards d'analyse transversale. L'hypothèse d'exogénéité du terme d'erreur est donc plutôt

$$E(u_{i,t}|X_{i,t}) = 0$$

Cette hypothèse permet donc à $u_{i,t}$ et $X_{i,t}$ d'être corrélés entre une période s et une période t lorsque $s \neq t$. Évidemment, d'autres hypothèses doivent être faites, tout comme des corrections doivent être apportées à la matrice de variance-covariance pour contrôler pour l'autocorrélation entre les périodes d'un même individu.

Nous présentons maintenant notre modèle complet, c'est-à-dire notre modèle *pooled* probit. L'une des hypothèses propres au modèle panel que nous devons abandonner est la présence d'un terme inobservable c_i . Cette hypothèse nous permettait d'obtenir de l'hétérogénéité entre les individus dans notre modèle, ce qui est moins contraignant que le modèle que nous estimons. Cependant, en abandonnant les modèles panel plus standards, nous pouvons aussi relâcher, comme nous l'avons dit plus haut, l'hypothèse de stricte exogénéité. Cela nous permet d'introduire un aspect dynamique à notre modèle. Notre modèle est donc un probit sur l'ensemble des périodes des individus. Cependant, nous devons corriger pour l'autocorrélation dans le terme d'erreur.

Nous abordons ci-dessous la théorie concernant la prise en compte de l'autocorrélation par une matrice de variance-covariance corrigée. Cette matrice de variance-covariance s'applique, selon les auteurs qui la présentent, aux modèles non linéaires. Nous savons que l'estimation d'un modèle non linéaire avec des problèmes dans l'estimation de la variance peut mener à des coefficients biaisés. Nous prenons compte de ce problème dans nos limites, mais l'ignorons pour l'estimation de nos modèles. Nous présentons donc la matrice de correction, car c'est ce que l'option *cluster* pour les estimations probit dans Stata semble effectuer. Il est à souligner que la documentation de Stata sur la procédure exacte de cette application est particulièrement obscure.

En définissant ⁷

$$h(y_{i,t}, X_{i,t}, \beta) = \frac{\partial l(y_{i,t}|X_{i,t}, \beta)}{\partial \beta}$$

Notre problème de maximisation de vraisemblance doit résoudre

$$\sum_{t=1}^T \sum_{i=1}^N h(y_{i,t}, X_{i,t}, \beta) = 0$$

où nous croyons que les erreurs sont corrélées entre les périodes pour un même individu, de sorte que

$$\text{corr}(h_{i,t}, h_{i,k}) \neq 0$$

On peut obtenir la matrice de variance-covariance *cluster-robust* selon la formule suivante

$$\hat{V}(\hat{\beta}) = \left(\sum_{i=1}^N \sum_{t=1}^T \frac{\partial h'_{i,t}}{\partial \beta} \Big|_{\hat{\beta}} \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^T h_{i,t}(\hat{\beta}) h_{i,k}(\hat{\beta})' \left(\sum_{i=1}^N \sum_{t=1}^T \frac{\partial h_{i,t}}{\partial \beta} \Big|_{\hat{\beta}} \right)^{-1}$$

7. Théorie tirée de (Cameron et Trivedi, 2005, p.842).

En somme, notre modèle est un modèle *pooled* probit avec une matrice de variance-covariance corrigée pour l'autocorrélation dans les termes d'erreur. Le reste de cette section présente les tests de spécification que nous effectuons pour comparer nos modèles explicatifs.

3.3 Tests de spécification

Pour comparer les deux modèles explicatifs, le modèle Contrôle-demande et le modèle Iso-contraint, nous utilisons deux tests de spécification.

Le premier test est un test de performance prédictive. La prédiction d'un modèle probit à choix binaire comme celui que nous estimons est la probabilité, conditionnellement à certains X_i , que $y_i = 1$. On peut ainsi écrire

$$\widehat{Pr}_i = \widehat{Pr}(y_i = 1|X_i)$$

Nous pouvons décider d'un seuil au-delà nous considérons que la prédiction est suffisamment élevée pour que y_i aurait pris la valeur 1. Nous utiliserons un seuil de 50%, c'est-à-dire

$$\widehat{y}_i = 1 \quad si \quad \widehat{Pr}_i > 0.5$$

Suite à cette estimation, on peut calculer la proportion de bonne prédiction (Correct Classification Ratio) de la façon suivante

$$CCR = \frac{1}{N} \sum \mathbf{1}(y_i = \widehat{y}_i)$$

Un autre test de spécification qui permet de tester une restriction dans un modèle est celui du ratio des vraisemblances. Soit r le nombre de restriction

$$R\beta = q$$

On peut alors comparer les vraisemblances selon la statistique suivante

$$-2 \left(\frac{L_R}{L_{NR}} \right) = -2(l_{NR} - l_R) \sim \chi_r^2$$

où L_R et L_{NR} sont respectivement les vraisemblances des modèles restreint et non-restreint et l_R et l_{NR} les log-vraisemblances. Cette statistique suit une loi khi-deux avec r , le nombre de restrictions, degrés de liberté. L'hypothèse nulle pour ce test est que les deux modèles, soient celui avec et celui sans restriction, sont le même modèle.

4 Données

Nos données proviennent du Longitudinal Internet Studies for the Social sciences (LISS), un panel néerlandais mis en place par l'Université de Tilburg. Ce panel existe depuis 2007 et contient 5000 ménages et 8000 individus. Nous utilisons pour notre analyse plusieurs enquêtes du LISS.

Certaines des enquêtes sont mensuelles et d'autres annuelles. Notre base de données contient chacun de ces types d'enquêtes.

Tout d'abord, nos variables dépendantes portant sur la santé proviennent du questionnaire *Health*, une enquête ayant lieu chaque mois de novembre. Nos données vont de 2008 à 2013, mais pour l'introduction de la variable dépendante retardée, nous utilisons également l'enquête ayant eu lieu en 2007. Cette base de données contient également des informations sur les habitudes de vie. Pour nos variables dépendantes, elles sont la réponse à la question suivante : "Au cours de la dernière année, un médecin vous a-t-il dit que vous aviez un problème de..."⁸. Comme y_i , nous avons conservé des problèmes de coeur et des problèmes de pression sanguine. Comme spécifié précédemment, ces deux problématiques de santé ont été associées précédemment à de hauts niveaux de stress (Sapolsky, 2004).

Nous utilisons également l'enquête *Work and Schooling* qui a lieu chaque mois d'avril. Cette enquête porte notamment sur le revenu et la scolarité. C'est dans cette enquête que nous obtenons nos variables indépendantes d'intérêt, à savoir les conditions de travail. Cette fois, nos variables d'intérêt sont une échelle de 1 à 4, 1 étant complètement en désaccord avec l'affirmation, et 4 étant complètement en accord avec l'affirmation. Notre première variable capte l'effet de la charge de travail et est l'évaluation, de 1 à 4, de l'affirmation suivante : "à cause d'une lourde charge de travail, je suis continuellement sous pression pour des échéanciers de temps."⁹ Notre seconde variable capte le niveau de contrôle que le travailleur considère avoir sur son travail. C'est l'appréciation de 1 à 4 de l'affirmation : "J'ai peu de liberté sur comment je fais mon travail."¹⁰ Ces deux variables sont celles du modèle Demande-contrôle présenté à la section précédente. Nous introduisons également dans nos modèles une interaction entre ces deux variables afin de bien capter la matrice de Karasek présentée plus haut. La prochaine variable, la variable captant le support, est l'ajout fait par le modèle Iso-contraint. Il s'agit de l'évaluation de l'affirmation : "Je reçois suffisamment de support dans les situations difficiles."¹¹ Encore une fois, cette évaluation est faite de 1 à 4.

Pour obtenir des informations plus personnelles sur les individus, comme leur sexe et leur état matrimonial, nous utilisons ensuite l'enquête mensuelle *Background Variables*. Pour obtenir de l'information qui puisse bien expliquer nos variables dépendantes provenant de l'enquête *Health*, nous avons pris les résultats de l'enquête *Background Variables* ayant été récoltés durant les mois de novembre de chaque année.

Nous avons réuni dans une seule base de données tous les individus présents à toutes les périodes. Nous avons ensuite conservé uniquement les individus qui ont travaillé durant les 6 périodes. Puisque nous voulons investiguer l'impact des conditions de travail sur la santé, nous ne pouvons

8. Traduction libre de l'anglais : Has a physician told you this last year that you suffer from one of the following diseases/problems ?

9. Traduction libre de l'anglais : Because of a heavy work burden, I am continually under time pressure.

10. Traduction libre de l'anglais : There is very little freedom for me to determine how to do my work.

11. Traduction libre de l'anglais : I get sufficient support in difficult situations.

utiliser des étudiants, des retraités ou des chômeurs. L'hypothèse que nous faisons ici est que si un individu a répondu avoir un emploi à chaque période, soit il a conservé le même emploi durant l'ensemble des périodes, soit il a changé d'emploi, mais l'impact de la période de transition sur sa santé fut négligeable.

Une fois l'assemblage de nos données complété, nous avons dans notre échantillon 805 individus. 52% de ces individus sont des hommes, 31,6% ont comme dernier diplôme un secondaire 5, 53,5% ont un diplôme professionnel et 10,5% ont un diplôme universitaire. L'âge moyen de notre échantillon est de 47,3 ans.

Pour ce qui est de nos variables dépendantes, nous avons 17 personnes (2% de notre échantillon) dans notre échantillon qui disent avoir eu ont moins une fois au cours des 6 ans un problème de coeur. Pour les problèmes de pression, c'est plutôt 137 personnes (17%).

5 Résultats

Pour chaque problème de santé, nous estimons d'abord le modèle économétrique comprenant les variables *Charge* et *Liberte* afin d'estimer le modèle explicatif Demande-contrôle. Nous estimons ensuite le même modèle, mais cette fois en ajoutant la variable *Support*. Ce deuxième modèle correspond au modèle explicatif Iso-contraint. Les tables complètes des résultats sont en Annexe 1. Avant de présenter les résultats principaux, nous présentons ici les variables retenues pour nos régressions : la variable *ActPhysique* capte le nombre de jours durant lesquels l'individu à fait de l'activité physique dans la dernière semaine ; *Age* et *Age2* captent respectivement l'âge et l'âge au carré des individus en années ; *AlcoolE* indique si l'individu consomme de l'alcool trois fois par semaine ou plus ; *AlcoolM* indique si l'individu consomme de l'alcool de une fois par mois à une fois par semaine ; *FruitE* indique si l'individu consomme des fruits cinq fois par semaine ou plus ; *FruitM* indique si l'individu consomme des fruits de une à quatre fois par semaine ; les variables *ViandeE* et *ViandeM* suivent la même répartition que celle de la consommation des fruits ; *Homme* indique si l'individu est un homme ; *Marie* indique si l'individu est marié ; *Secondaire*, *Professionnel* et *Université* indique le plus haut niveau de diplomation d'un individu ; *RevenuNet* indique le revenu après impôts et transferts ; *Urbain* indique si l'individu a déclaré qu'il résidait dans un environnement très ou extrêmement urbain ; enfin, les variables *CoeurLag* et *PresLag* indiquent si l'individu avait un problème de coeur ou de pression à la période précédente.

5.1 Problème de coeur

Dans le tableau 2 sont reportés les résultats pour le modèle Demande-Contrôle. On voit que le fait d'être exposé à une lourde charge de travail ne semble pas avoir d'effet sur la probabilité d'avoir des problèmes cardiaques. Également, plus un individu indique ne pas avoir de liberté dans

son travail, plus la probabilité d'avoir un problème de santé diminue. Ces deux résultats peuvent a priori sembler contraires à la théorie présentée précédemment. Cependant, lorsque l'on regarde l'interaction entre les deux variables, on voit qu'avoir simultanément une lourde charge de travail et peu de liberté sur son travail augmente les risques associés à des problèmes de coeur. Cette interaction représente la case supérieure droite de notre tableau 1. En testant conjointement les trois variables d'intérêt, nous obtenons une statistique χ^2 avec 3 degrés de liberté de 7,08 (p-value de 0,07). Nous trouvons donc des résultats au moins partiellement en accord avec la théorie.

Pour ce qui est des variables de contrôle, on capte tout d'abord un effet significatif et non linéaire dans l'âge. Faire de l'activité physique prémunie contre les problèmes de coeur. Enfin, avoir eu un problème de coeur diagnostiqué à la période précédente est un excellent prédicteur d'un problème de coeur à la période présente.

Le tableau 3 reporte les résultats du modèle Iso-contraint, c'est-à-dire avec l'ajout du support comme variable médiatrice. Ce qui est intéressant avec ces résultats, c'est tout d'abord qu'ils démontrent bien la robustesse du modèle Demande-contrôle. L'introduction du support n'a pratiquement rien changé au coefficient de même qu'au niveau de significativité des variables explicatives. Pour ce qui est du support lui-même, il n'est pas significatif et ne semble donc pas avoir d'effet médiateur.

Les tests de spécification semblent aller dans le sens des résultats précédents. Si nous prenons le test de performance prédictive, nous trouvons que pour un seuil de 50%, le modèle Demande-contrôle prédit correctement 99,45% des y_i . Le modèle Iso-contraint obtient exactement la même performance, soit lui aussi 99,45%. Sur la base de ce test, on ne peut donc pas rejeter l'hypothèse que l'ajout du support comme variable explicative n'apporte rien au modèle.

Pour ce qui est du test de ratio des vraisemblances, nous obtenons une statistique de 1,134781. Pour une χ^2 avec un degré de liberté, le seuil de rejet à 5% est 3,8415. On ne peut donc pas rejeter l'hypothèse nulle que les deux modèles sont identiques.

5.2 Problème de pression sanguine

Du côté du modèle portant sur les problèmes de pression sanguine, les résultats sont présentés dans les tableaux 4 et 5. Nous semblons trouver sensiblement les mêmes résultats que pour les modèles portant sur les problèmes de coeur dans les deux modèles estimés. Pour le modèle Demande-contrôle, nous trouvons encore une fois que la charge de travail ne semble pas avoir d'effet sur la probabilité d'avoir des problèmes de pression. Également, encore une fois à contre-courant de la théorie, nous trouvons que de considérer avoir peu de liberté pour faire son travail diminue les risques de problèmes de pression. Enfin, nous trouvons un effet positif et significatif pour le coefficient de la variable d'interaction entre la charge de travail et la liberté. Ainsi, un individu considérant avoir une lourde charge de travail simultanément à avoir peu de liberté sur la façon de faire son travail est plus à risque d'avoir des problèmes de pression sanguine. Lorsque l'on test les

trois variables conjointement, nous obtenons une statistique χ^2 à trois degrés de liberté de 13,52 (p-value<0,01). Les variables semblent donc bien pertinentes et encore une fois partiellement en accord avec la théorie.

En ce qui concerne le modèle Iso-contraint, une fois de plus l'ajout du support comme variable explicative n'affecte pas les résultats. Les coefficients demeurent passablement les mêmes, tout comme les niveaux de significativité. Le support ne sort pas comme une variable significative, ce qui est concordant avec nos modèles portant sur les problèmes de coeur. Afin de formaliser ces résultats, on peut se tourner du côté des tests de spécification.

Pour ce qui est du test de performance prédictive, le modèle Contrôle-demande prédit correctement 96,59% des y_i . Ce taux de réussite est exactement le même que pour le modèle Iso-contraint, ce qui laisse croire que les deux modèles sont indifférenciables. Pour ce qui est du test de ratio des vraisemblances, nous obtenons une statistique de test de 0,1123047. Si l'on compare ce résultat à la valeur d'une χ^2 avec un degré de liberté à le seuil de rejet à 5%, soit 3,8415, on ne peut rejeter l'hypothèse nulle que l'inclusion du support n'apporte rien au modèle.

Pour s'assurer de la robustesse de nos résultats, nous avons introduit des interactions entre la variable *Support* et les variables *Charge* et *Liberte* dans les modèles iso-contraint pour chacun des indicateurs de santé. Dans aucune des spécifications explorées, l'une de ces interactions n'est apparue significative. Également, nos résultats sont restés inchangés suite à ces ajouts. Cela semble confirmer la validité de nos estimations.

Il est intéressant de constater que nos résultats, malgré les différences méthodologiques par rapport aux études précédentes, sont cohérents avec ceux obtenus jusqu'à présent dans la littérature. L'interaction entre une lourde charge de travail et un faible contrôle sur son travail est associée à des risques de maladies cardio-vasculaires plus importants. Nous avons également trouver peu de place pour le support comme médiateur dans cette relation, ce qui semble également être le cas dans la littérature.

6 Limites

Notre analyse souffre évidemment de certaines limites. Nous en soulevons quelques une dans cette section.

La première, qui a déjà été mentionnée précédemment, est que nous avons un possible problème avec l'estimation de notre matrice de variance-covariance. En effet, si la matrice à estimer dans nos modèles n'est pas homoscédastique et sans autocorrélation, non seulement sont estimation est invalide, mais les coefficients du modèle son possiblement biaisés. Il semble que ce problème soit souvent ignoré, mais nous tenions tout de même à le souligner. Une des méthodes proposées dans la littérature est d'utiliser une matrice de variance-covariance corrigée pour les écarts-types, mais cela ne règle cependant pas le problème de biais. C'est néanmoins cette procédure que nous avons

utilisée, puisque nos variances sont très certainement autocorrélée.

Une seconde limite à notre modèle est la brève période de temps couverte. Dans la plupart des études présentées dans ce travail, les cohortes étaient suivies durant de nombreuses années. Nous n'avons ici de l'information que sur une durée de 6 ans, ce qui peut être considéré comme court lorsque l'on cherche à capter des effets négatifs qui peuvent se manifester après de longues périodes. Les résultats que nous présentons peuvent toutefois demeurer valides si les individus observés changent peu d'emploi et si les conditions de travail demeurent stables dans le temps.

Une autre limite de notre analyse est que l'évaluation des conditions de travail est subjective. La science économique a traditionnellement préféré des valeurs plus objectives. Cependant, il est à noter que cette notre analyse ne s'écarte pas trop de la littérature actuelle, puisque la plupart des études présentées utilisent des index combinant plusieurs réponses à des appréciations subjectives. De plus, puisque le stress possède une composante psychologique importante, l'aspect subjectif de nos variables peut s'avérer être un atout. Également, nos réponses sont une évaluation d'énoncé sur une échelle allant de 1, "Complètement en désaccord", à 4, "Complètement en accord". Plusieurs auteurs se sont inquiétés que le passage de 1 à 2 n'ait pas la même significativité que de 3 à 4, par exemple. Nous sommes conscients de cette critique, mais désirions conserver cette échelle, malgré ses limites, afin d'obtenir une variable plus parlante que si nous avions transformé nos réponses en variable dichotomique.

Enfin, une limite évidente de notre modèle portant sur les problèmes de coeur est le faible nombre d'événements observés. Nos résultats semblent toutefois robustes, puisque le modèle portant sur les problèmes de pression sanguine obtient sensiblement les mêmes résultats avec un nombre d'événements plus importants.

7 Conclusion

Dans ce travail, nous tentions de comparer deux modèles explicatifs portant sur l'incidence des conditions de travail sur la santé des travailleurs. Le plus populaire de ces deux modèles, le modèle Demande-contrôle, a été largement confirmé dans la littérature, tandis que le second, le modèle Iso-contraint, a obtenu des résultats plus mitigés. Grâce à un modèle en panel de type probit, nous avons pu vérifier ces résultats à l'aide de données néerlandaises. Nos résultats confirment l'efficacité du modèle Demande-contrôle par rapport au modèle Iso-contraint. À présent, de nouvelles versions du modèle Iso-contraint, incluant le support familial, ont été proposées pour pallier au succès incertain du modèle initial. Il serait intéressant à l'avenir d'observer comment ces nouvelles versions se défendent vis-à-vis du modèle Demande-contrôle.

8 Bibliographie

Berkman, L. F., Kwachi, I. et T. Theorell (2014) 'Working conditions and health', in Berkman, L. F., Kawachi, I. et M. M. Glymour (ed.) *Social epidemiology*, New-York : Oxford University Press.

Bosma, H., Peter, R., Siegrist, J. et M. Marmot (1998) 'Two alternative job stress models and the risk of coronary heart disease', *American Journal of Public Health*, vol. 88, no. 1, Janvier, pp. 68-77.

Cameron, A. C. et Trivedi, P. K. (2005) *Microeconometrics : Methods and applications*, Cambridge : Cambridge University Press.

Johnson, J. V., Stewart, W., Hall, E. M., Fredlund, P. et T. Theorell (1996) 'Long-term psychosocial work environment and cardiovascular mortality among Swedish men', *American Journal of Public Health*, vol. 86, no. 3, Mars, pp. 324-331.

Kang, M. G., Sang, B. K., Cha, B. S., Park, J. K., Baik, S. K. et S. J. Chang (2005) 'Job stress and cardiovascular risk factors in male workers', *Preventive Medicine*, vol. 40, no. 5, pp. 583-588.

Karasek, R. A. Jr (1979) 'Job demands, job decision latitude, and mental strain : Implications for job redesign', *Administrative Science Quarterly*, vol. 24, no. 2, Juin, pp. 285-308.

Kivimäki, M., Nyberg, S. T., Batty, G. D., Fransson, E. I., Heikkilä, K. *et al* (2012) 'Job strain as a risk factor for coronary heart disease : A collaborative meta-analysis of individual participant data', *The Lancet*, vol. 380, Octobre, pp.1491-1497.

LaRocco, J. M., House, J. S. et J. R. P. French Jr (1980) 'Social support, occupational stress, and health', *Journal of Health and Social Behavior*, vol. 21, no. 3, Septembre, pp.202-218.

Sapolsky, R. (2004) *Why zebras don't ulcers : The acclaimed guide to stress, stress-related diseases, and coping*, New-York : St. Martin's Griffin.

Viswesvaran, C., Sanchez, J. I. et J. Fisher (1999) 'The role of social support in the process of work stress : A meta-analysis', *Journal of Vocational Behavior*, vol. 54, pp. 314-334.

Wooldridge, J. M. (2002) *Econometric analysis of cross section and panel data*, Cambridge : The MIT Press.

9 Annexe 1

9.1 Coeur

TABLE 2 – Coeur : Modèle Demande-contrôle

Variable	Coefficient	(Std. Err.)
ActPhysique	-0.098**	(0.048)
Age	0.459**	(0.213)
Age2	-0.004**	(0.002)
AlcoolE	-0.279	(0.189)
AlcoolM	-0.500***	(0.167)
FruitE	-0.151	(0.251)
FruitM	-0.025	(0.259)
ViandeE	-0.143	(0.308)
ViandeM	-0.301	(0.318)
Homme	-0.029	(0.149)
Marie	0.092	(0.173)
Secondaire	-0.130	(0.411)
Professionnel	0.003	(0.369)
Université	-0.120	(0.413)
RevenuNet	0.000	(0.000)
Urbain	-0.189	(0.166)
CoeurLag	3.396***	(0.199)
Charge	-0.335	(0.232)
Liberte	-0.848**	(0.355)
CharLib	0.241**	(0.121)
Intercept	-13.275**	(5.416)
Log-likelihood	-104.305	
$\chi^2_{(20)}$	1351.386	

Niveau de significativité : * : 10% ** : 5% *** : 1%

TABLE 3 – Coeur : Modèle Iso-contraint

Variable	Coefficient	(Std. Err.)
ActPhysique	-0.097**	(0.048)
Age	0.463**	(0.214)
Age2	-0.004**	(0.002)
AlcoolE	-0.254	(0.189)
AlcoolM	-0.473***	(0.165)
FruitE	-0.184	(0.242)
FruitM	-0.061	(0.249)
ViandeE	-0.173	(0.308)
ViandeM	-0.321	(0.326)
Homme	-0.017	(0.147)
Marie	0.084	(0.172)
Secondaire	-0.108	(0.409)
Professionnel	0.019	(0.368)
Université	-0.094	(0.412)
RevenuNet	0.000	(0.000)
Urbain	-0.190	(0.166)
CoeurLag	3.413***	(0.204)
Charge	-0.336	(0.237)
Liberte	-0.851**	(0.352)
CharLib	0.250**	(0.119)
Support	0.164	(0.105)
Intercept	-13.888**	(5.459)
Log-likelihood	-103.738	
$\chi^2_{(21)}$	1442.839	
Niveau de significativité : * : 10% ** : 5% *** : 1%		

9.2 Pression

TABLE 4 – Pression : Modèle Demande-contrôle

Variable	Coefficient	(Std. Err.)
ActPhysique	0.035*	(0.020)
Age	0.060	(0.039)
Age2	0.000	(0.000)
AlcoolE	-0.140	(0.105)
AlcoolM	-0.154	(0.096)
FruitE	0.024	(0.115)
FruitM	0.062	(0.121)
ViandeE	0.258	(0.189)
ViandeM	0.216	(0.195)
Homme	-0.068	(0.081)
Marie	0.056	(0.090)
Secondaire	0.014	(0.208)
Professionnel	-0.029	(0.203)
Université	-0.217	(0.233)
RevenuNet	0.000**	(0.000)
Urbain	-0.141*	(0.083)
PresLag	3.107***	(0.101)
Charge	-0.140	(0.114)
Liberte	-0.322**	(0.141)
CharLib	0.133***	(0.051)
Intercept	-4.028***	(0.993)
Log-likelihood		-575.164
$\chi^2_{(20)}$		1121.919
Niveau de significativité : * : 10% ** : 5% *** : 1%		

TABLE 5 – Pression : Modèle Iso-contraint

Variable	Coefficient	(Std. Err.)
ActPhysique	0.034*	(0.020)
Age	0.060	(0.040)
Age2	0.000	(0.000)
AlcoolE	-0.141	(0.105)
AlcoolM	-0.155	(0.097)
FruitE	0.025	(0.115)
FruitM	0.062	(0.121)
ViandeE	0.260	(0.190)
ViandeM	0.216	(0.195)
Homme	-0.071	(0.081)
Marie	0.058	(0.090)
Secondaire	0.016	(0.208)
Professionnel	-0.026	(0.202)
Université	-0.216	(0.233)
RevenuNet	0.000**	(0.000)
Urbain	-0.142*	(0.083)
PresLag	3.107***	(0.101)
Charge	-0.140	(0.114)
Liberte	-0.322**	(0.141)
CharLib	0.132***	(0.051)
Support	-0.021	(0.063)
Intercept	-3.970***	(1.012)
Log-likelihood		-575.108
$\chi^2_{(21)}$		1119.997
Niveau de significativité : * : 10% ** : 5% *** : 1%		